Machine learning for prediction and visualisation of brain diseases. Demonstration on Alzheimer's disease

Răzvan V. Marinescu

Medical Vision Group, Massachusetts Institute of Technology
Centre for Medical Image Computing, University College London, UK

Slides available online: https://people.csail.mit.edu/razvan

What to expect from my presentation

- Why is prediction of Alzheimer's disease important? Why do drug trials fail?

What to expect from my presentation

- Why is prediction of Alzheimer's disease important? Why do drug trials fail?
- Which biomarkers can we predict, and which we cannot?

What to expect from my presentation

- Why is prediction of Alzheimer's disease important? Why do drug trials fail?
- Which biomarkers can we predict, and which we cannot?
- What is the state-of-the-art in Alzheimer's prediction?

What to expect from my presentation

- Why is prediction of Alzheimer's disease important? Why do drug trials fail?
- Which biomarkers can we predict, and which we cannot?
- What is the state-of-the-art in Alzheimer's prediction?
- What are the best algorithms? Should I use deep learning or not?

What to expect from my presentation

- Why is prediction of Alzheimer's disease important? Why do drug trials fail?
- Which biomarkers can we predict, and which we cannot?
- What is the state-of-the-art in Alzheimer's prediction?
- What are the best algorithms? Should I use deep learning or not?
- Features: which ones are most informative? Do I need to pre-process those DTI scans, are MRIs not enough?

What to expect from my presentation

- Why is prediction of Alzheimer's disease important? Why do drug trials fail?
- Which biomarkers can we predict, and which we cannot?
- What is the state-of-the-art in Alzheimer's prediction?
- What are the best algorithms? Should I use deep learning or not?
- Features: which ones are most informative? Do I need to pre-process those DTI scans, are MRIs not enough?
- How well do algorithms work on "real data", i.e. mimicking clinical trials?

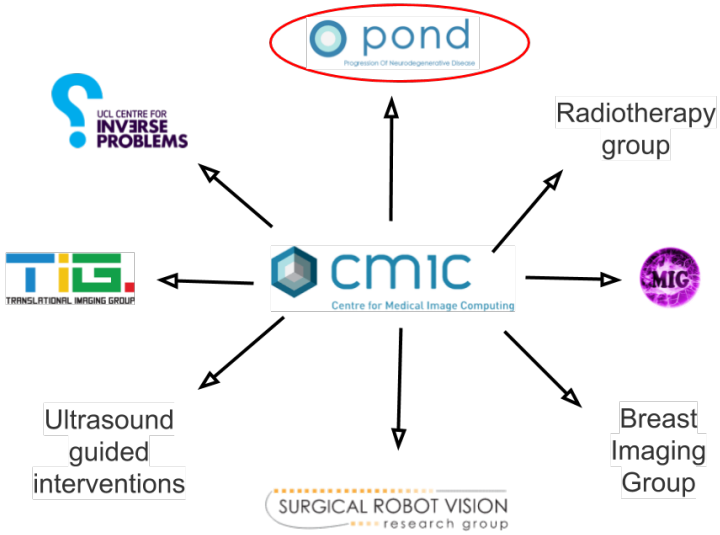
What to expect from my presentation

- Why is prediction of Alzheimer's disease important? Why do drug trials fail?
- Which biomarkers can we predict, and which we cannot?
- What is the state-of-the-art in Alzheimer's prediction?
- What are the best algorithms? Should I use deep learning or not?
- Features: which ones are most informative? Do I need to pre-process those DTI scans, are MRIs not enough?
- How well do algorithms work on "real data", i.e. mimicking clinical trials?
- How can we visualise the progression of Alzheimer's disease?

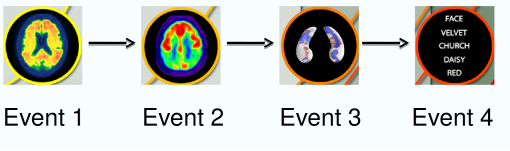
About me

- Grew up in Pitesti, Romania
- 2010-2014: Studied a 4-year MEng in Computer Science at Imperial College London
- 2014-2019: PhD in Medical Imaging at UCL (with Daniel Alexander)
- 2019-present: Postdoc in CSAIL at MIT (with Polina Golland)

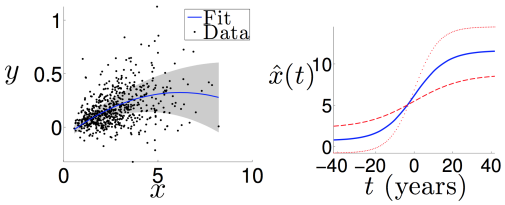



pond
Progression Of Neurodegenerative Disease
UCL CENTRE FOR INVERSE PROBLEMS
Radiotherapy group
TIG
TRANSLATIONAL IMAGING GROUP
cmic
Centre for Medical Image Computing
MIC
Ultrasound guided interventions
Breast Imaging Group
SURGICAL ROBOT VISION
research group

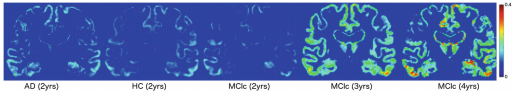
Event-Based Model
(Fontejin et al., Neuroimage, 2012)



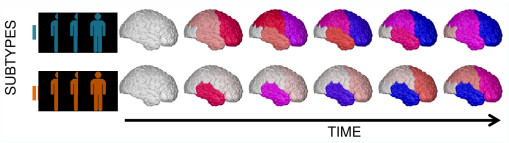
Differential Equation Model
(Oxtoby et al., Brain, 2018)



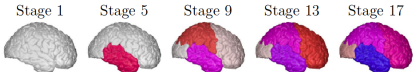
Gaussian-Process Regression
(Lorenzi et al., IPMI, 2015)



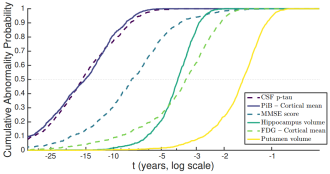
Subtype and Stage Inference
(Young et al., Nature Comms., 2018)



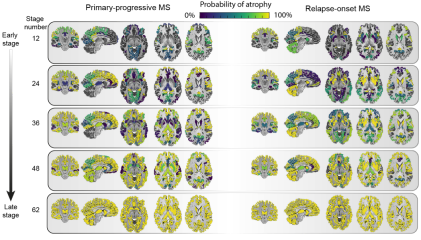
typical Alzheimer's Disease
(Young et al., Nature Comms., 2018)



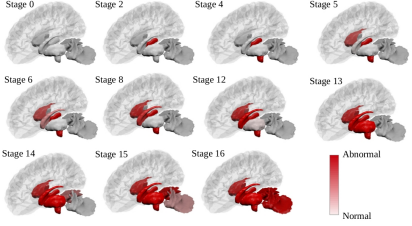
Familial Alzheimer's Disease
(Oxtoby et al., Brain, 2018)



Multiple sclerosis
(Eshaghi et al., Brain, 2017)

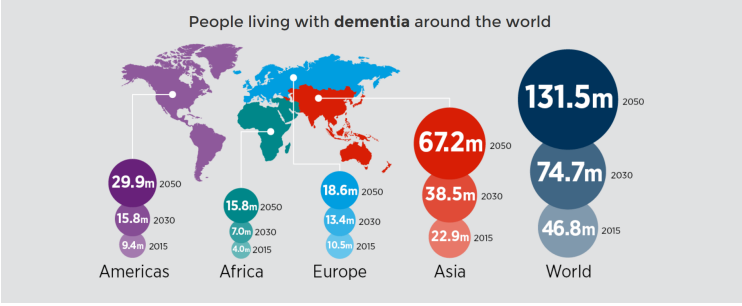


Huntington's disease
(Wijeratne et al., Ann. Clin. Neurol;, 2018)



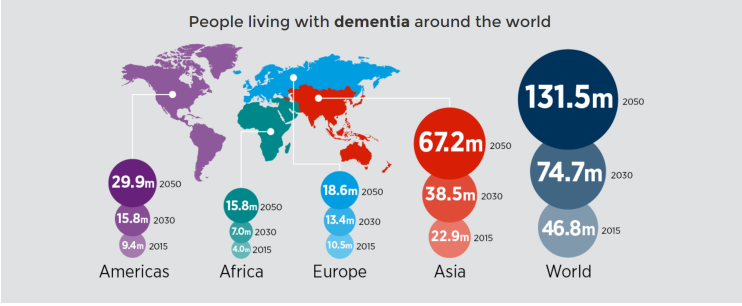
Alzheimer's Disease is a Devastating Disease

- 46 million people affected worldwide



Alzheimer's Disease is a Devastating Disease

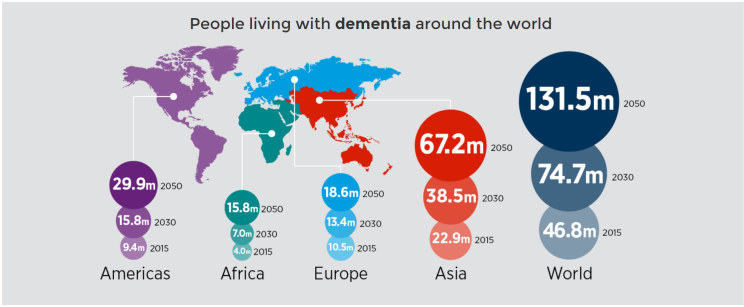
- 46 million people affected worldwide



- No treatments available that stop or slow down cognitive decline
- Q: Why did clinical trials fail? A: Treatments were not administered early enough

Alzheimer's Disease is a Devastating Disease

- 46 million people affected worldwide

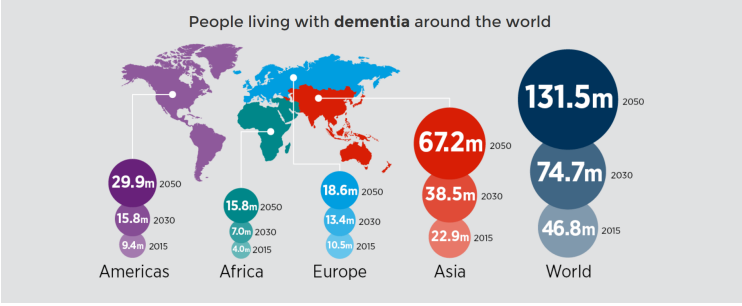


- No treatments available that stop or slow down cognitive decline
- Q: Why did clinical trials fail? A: Treatments were not administered early enough

- Q: How can we then identify subjects **early** in order to administer treatments?
- A: Build models that predict evolution of Alzheimer's biomarkers (i.e. biological markers) for at-risk subjects

Alzheimer's Disease is a Devastating Disease

- 46 million people affected worldwide

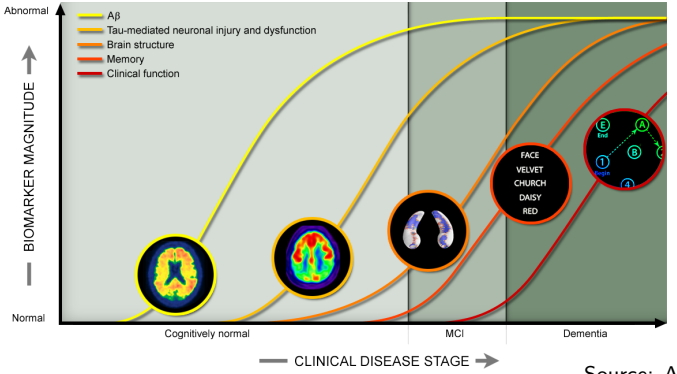


- No treatments available that stop or slow down cognitive decline
- Q: Why did clinical trials fail? A: Treatments were not administered early enough

- Q: How can we then identify subjects **early** in order to administer treatments?
- A: Build models that predict evolution of Alzheimer's biomarkers (i.e. biological markers) for at-risk subjects

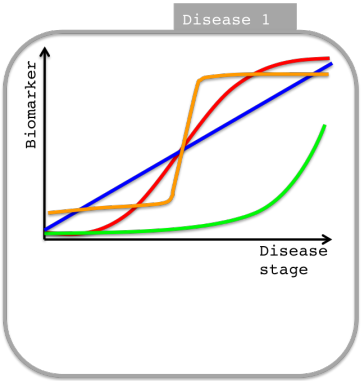
- These models can help stage and refine cohorts in Alzheimer's clinical trials

Biomarker Evolution creates a Unique Disease Signature that can be used for Staging Individuals in Clinical Trials



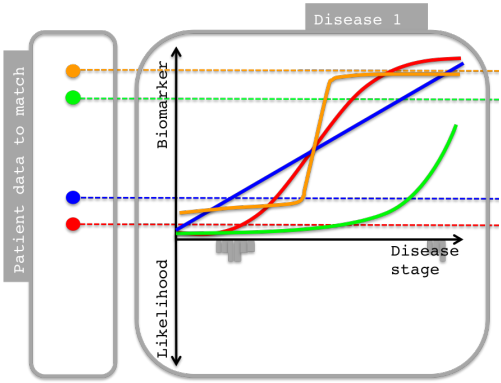
Source: ADNI website

- Accurate disease staging → better patient stratification
- Problem: This is a "hypothetical" (i.e. qualitative) disease progression model
- Why construct a quantitative model?



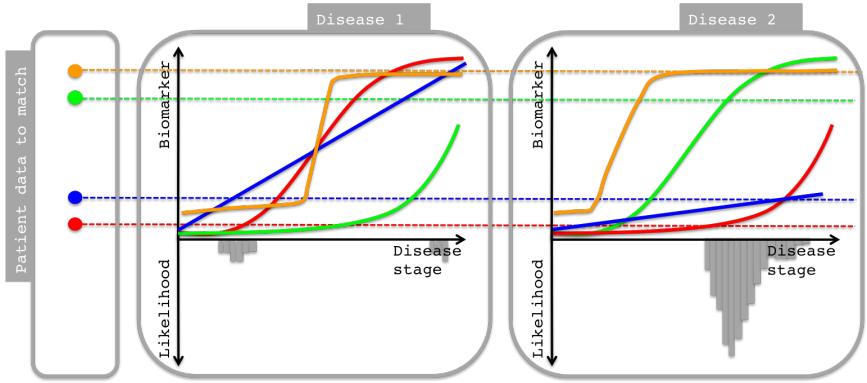
- Basic biological insight

Benefits of Quantitative Disease Progression Models



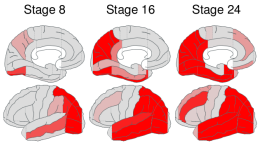
- Basic biological insight
- Staging can help stratification in clinical trials

Benefits of Quantitative Disease Progression Models

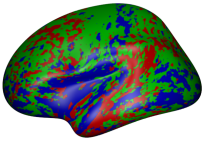


- Basic biological insight
- Staging can help stratification in clinical trials
- Differential diagnosis and prognosis

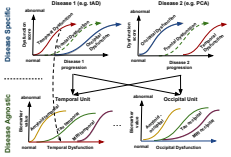
1. Modelled progression of PCA and tAD



2. Developed Novel Spatio-temporal Model



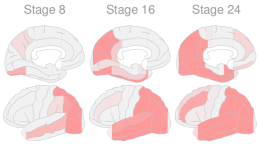
3. Developed Transfer Learning Model



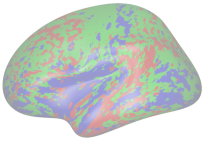
4. Meta-analysis of AD prediction algorithms

5. Created BrainPainter software

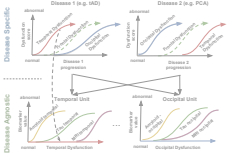
1. Modelled progression of PCA and tAD



2. Developed Novel Spatio-temporal Model



3. Developed Transfer Learning Model

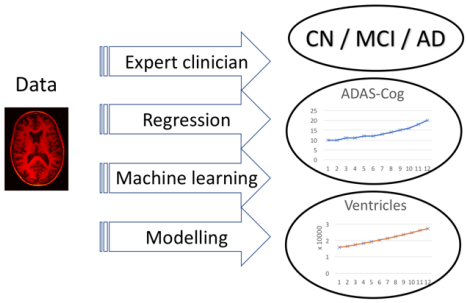


4. Meta-analysis of AD prediction algorithms

5. Created BrainPainter software

TADPOLE is a Challenge to Predict the Progression of Individuals at Risk of AD

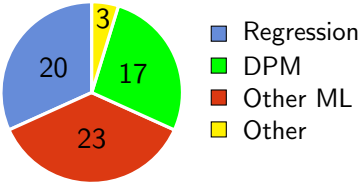
- Identify people that will develop Alzheimer's disease (AD) over the next 1-5 years.
  - Predict three target domains: clinical diagnosis, MRI (Ventricle Volume) and cognition (ADAS-Cog 13)
- Evaluation data on 219 subjects acquired by ADNI
- TADPOLE was entirely **prospective** – evaluation data acquired after submission deadline: Nov 2017



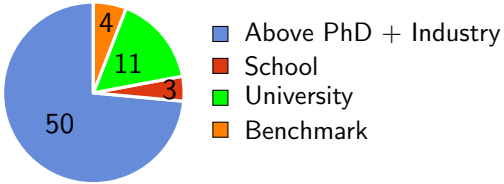
Submission statistics

33 teams from 12 countries

Algorithms



Teams



Submission methods were very diverse

Submission	Feature Selection	Nr. of features	Missing data imputation	Diagnosis prediction	ADAS/Vent. prediction
AlgosForGood	manual	16+5*	forward-filling	Aalen model	linear regression
Apocalypse	manual	16	population average	SVM	linear regression
ARAMIS-Pascal	manual	20	population average	Aalen model	-
ATRI-Biostat-JMM	automatic	15	random forest	random forest	linear mixed effects model
ATRI-Biostat-LTJMM	automatic	15	random forest	random forest	DPM
ATRI-Biostat-MA	automatic	15	random forest	random forest	DPM + linear mixed effects model
BGU-LSTM	automatic	67	none	feed-forward NN	LSTM
BGU-RF/ BGU-RFFIX	automatic	67+1340*	none	semi-temporal RF	semi-temporal RF
BIGS2	automatic	all	Iterative Soft-Thresholded SVD	RF	linear regression
Billabong (all)	manual	15-16	linear regression	linear scale	non-parametric SM
BORREGOSTECMTY	automatic	100 + 400*	nearest-neighbour	regression ensemble	ensemble of regression + hazard models
BravoLab	automatic	25	hot deck	LSTM	LSTM
CBIL	manual	21	linear interpolation	LSTM	LSTM
Chen-MCW	manual	9	none	linear regression	DPM
CN2L-NeuralNetwork	automatic	all	forward-filling	RNN	RNN
CN2L-RandomForest	manual	$>$200	forward-filling	RF	RF
CN2L-Average	automatic	all	forward-filling	RNN/RF	RNN/RF
CyberBrains	manual	5	population average	linear regression	linear regression
DIKU (all)	semi-automatic	18	none	Bayesian classifier/LDA + DPM	DPM
DIVE	manual	13	none	KDE+DPM	DPM
EMC1	automatic	250	nearest neighbour	DPM + 2D spline + SVM	DPM + 2D spline
EMC-EB	automatic	200-338	nearest-neighbour	SVM classifier	SVM regressor
FortuneTellerFish-Control	manual	19	nearest neighbour	multiclass ECOC SVM	linear mixed effects model
...	...	...	...	...	...
BenchmaskLastVisit	None	3	none	constant model	constant model
BenchmarkMixedEffect	None	3	none	Gaussian model	linear mixed effects model
BenchmarkMixedEffectAPOE	None	4	none	Gaussian model	linear mixed effects model
BenchmarkSVM	manual	6	mean of previous values	SVM	support vector regressor (SVR)

Prizes

- 30,000 GBP prize fund offered by sponsors:

- Prizes were split according into six categories:

Prize amount	Outcome measure	Eligibility
5,000	Diagnosis	all
5,000	Cognition	all
5,000	Ventricles	all
5,000	Overall best	all
5,000	Diagnosis	University teams
5,000	Diagnosis	High-school teams

Results Outline

- Prediction results:
  - Clinical diagnosis
  - Ventricle volume
  - Cognition
- Overall winners & winning strategy
- Results on limited dataset mimicking clinical trial
- Most informative features

Clinical Diagnosis prediction: Winner algorithms achieve considerable gains over best benchmarks and state-of-the-art

- MAUC error reduced by 58% compared to the best benchmark
- Winner (Frog) used a method based on gradient boosting (xgboost)
- TADPOLE algorithms pushed ahead the state-of-the-art:
  - Best/29 algos in CADDementia challenge had a diagnosis MAUC of 0.78
  - Best/15 algos (Morandi, NeuroImage, 2015) obtained AUC of 0.902
- Full results on TADPOLE website: https://tadpole.grand-challenge.org/Results

Team Name	RANK MAUC	MAUC
Frog	1	0.931
Threedays	2	0.921
EMC-EB	3	0.907
GlassFrog-SM	4-6	0.902
GlassFrog-Average	4-6	0.902
GlassFrog-LCMEM-HDR	4-6	0.902
Apocalypse	7	0.902
EMC1-Std	8	0.898
CBIL	9	0.897
CN2L-RandomForest	10	0.896
...	...	...
BenchmarkSVM	30	0.836
...	...	...

- MAUC - multiclass area under the receiver-operator curve

Ventricle prediction: Winner algorithms achieve considerable gains over best benchmarks

- MAE reduced by 58% compared to best benchmark
- Winner (EMC1) used a method based on disease progression models
- No previous state-of-the-art due to lack of studies predicting ventricles

FileName	Rank Ventricles	MAE Ventricles
EMC1-Std	1-2	0.4116
EMC1-Custom	1-2	0.4116
ImaUCL-Covariates	3	0.4155
ImaUCL-Std	4	0.4207
BORREGOTECMTY	5	0.4299
ImaUCL-halfD1	6	0.4402
CN2L-NeuralNetwork	7	0.4409
SBIA	8	0.4410
EMC-EB	9	0.4466
Frog	10	0.4469
VikingAI-Logistic	11-12	0.4534
VikingAI-Sigmoid	11-12	0.4534
CBIL	13	0.4625
...	...	...
BenchmarkMixedEffectsAPOE	23	0.5664
...	...	...

- MAE - mean absolute error

Cognition prediction: TADPOLE algorithms **fail to predict** significantly better than random

- RandomisedBest - best out of 100 random guesses
- Likely too much noise in cognitive test (ADAS-Cog 13)
- Methods might be better than random over longer time-windows ($>$ 2 years)

FileName	RANK Cognition	MAE Cognition
RandomisedBest	-	4.52
FortuneTellerFish-Control	1	4.70
BenchmarkMixedEffectsAPOE	2	4.75
FortuneTellerFish-SuStaIn	3	4.81
Frog	4	4.85
Mayo-BAI-ASU	5	4.98
CyberBrains	6	5.16
VikingAI-Sigmoid	7	5.20
GlassFrog-Average	8	5.26
CN2L-Average	9	5.31
CN2L-NeuralNetwork	10	5.36
DIKU-GeneralisedLog-Std	11-12	5.40
DIKU-GeneralisedLog-Custom	11-12	5.40
...	...	...

- MAE - mean absolute error

There was no clear winner method. Deep learning not among top entries.

- Deep Learning

Rank	Diagnosis
1	Gradient boosting
2	Random forest
3	SVM
4-6	Multi state model
4-6	Multi state model
4-6	Multi state model
7	SVM
8	DPM+SVM
9	LSTM
10	Random Forest
11	DPM+SVM
12	feed-forward NN
13-14	Bayesian classifier/LDA + DPM
13-14	Bayesian classifier/LDA + DPM
15	Aalen model
16	DPM + ordered logit model
17	Random forest
...	...

Rank	Ventricles
1-2	DPM + spline regression
1-2	DPM + spline regression
3	Multi-task learning
4	Multi-task learning
5	Ensenble of regression + hazard
6	Multi-task learning
7	RNN
8	Linear mixed effects
9	SVM regressor
10	Gradient boosting
11-12	DPM
11-12	DPM
13	LSTM
14	DPM
15	DPM
16	RNN+RF
17	RF
...	...

Consensus methods achieve top results

- Compared to the best TADPOLE submissions, consensus reduced the error by 11% for Cognition (ADAS) and 8% for Ventricles
- Most methods make systematic errors, either over- or under-estimating the future measurements

Submission	Overall	Diagnosis		Cognition		Ventricles	
	Rank	Rank	MAUC	Rank	MAE	Rank	MAE
ConsensusMedian	-	-	0.925	-	5.12	-	**0.38**
Frog	**1**	**1**	**0.931**	4	4.85	10	0.45
ConsensusMean	-	-	0.920	-	**3.75**	-	0.48
EMC1-Std	2	8	0.898	23-24	6.05	1-2	0.41
VikingAI-Sigmoid	3	16	0.875	7	5.20	11-12	0.45
EMC1-Custom	4	11	0.892	23-24	6.05	1-2	0.41
CBIL	5	9	0.897	15	5.66	13	0.46
Apocalypse	6	7	0.902	14	5.57	20	0.52
...		...		...		...	

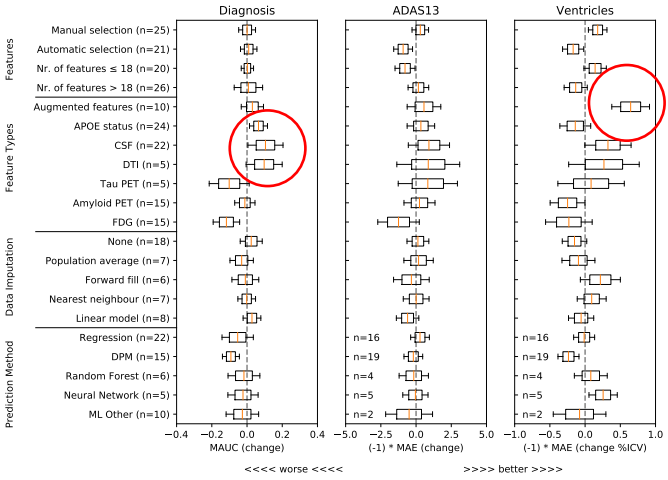
Prediction results on limited cross-sectional dataset mimicking a clinical trial are comparable to the full dataset

- Little loss of accuracy for the best methods
  - 0.48 vs 0.42 for ventricle MAE
  - 0.917 vs 0.931 for diagnosis MAUC
- Results suggest TADPOLE methods could be applied to clinical trial settings

Submission	Overall Rank	Diagnosis Rank	Diagnosis MAUC	Cognition Rank	Cognition MAE	Ventricles Rank	Ventricles MAE
ConsensusMean	-	-	**0.917**	-	4.58	-	0.73
ConsensusMedian	-	-	0.905	-	5.44	-	0.71
GlassFrog-Average	**1**	2-4	0.897	5	5.86	3	0.68
GlassFrog-LCMEM-HDR	2	2-4	0.897	9	6.57	**1**	**0.48**
GlassFrog-SM	3	2-4	0.897	4	5.77	9	0.82
Tohka-Ciszek-RandomForestLin	4	11	0.865	2	4.92	10	0.83
RandomisedBest	-	-	0.811	-	4.54	-	0.92
...		...		...		...	

What matters for good predictions?

- DTI and CSF features for clinical diagnosis prediction
- Augmented features for ventricle prediction
- However, further analysis needs to be done to make clear conclusions

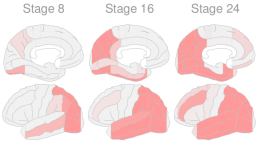


Next steps

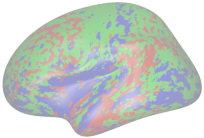
- TADPOLE SHARE: `https://tadpole-share.github.io/`
  - share methods for validation and further development
  - 11 teams already sharing
  - Lead by Esther Bron: e.bron@erasmusmc.nl
- AAIC 2020 special symposium
- Follow-on evaluations as more ADNI data becomes available
- Challenge still ongoing, D4 leaderboard now live

Overview

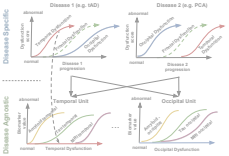
1. Modelled progression of PCA and tAD



2. Developed Novel Spatio-temporal Model



3. Developed Transfer Learning Model



4. Meta-analysis of AD prediction algorithms

5. Created BrainPainter software