Experiments in predicting early stage Alzheimer's disease.

Paul Moore[1] (joint with Terry Lyons[1], John Gallacher[2])

[1]Mathematical Institute, Oxford University

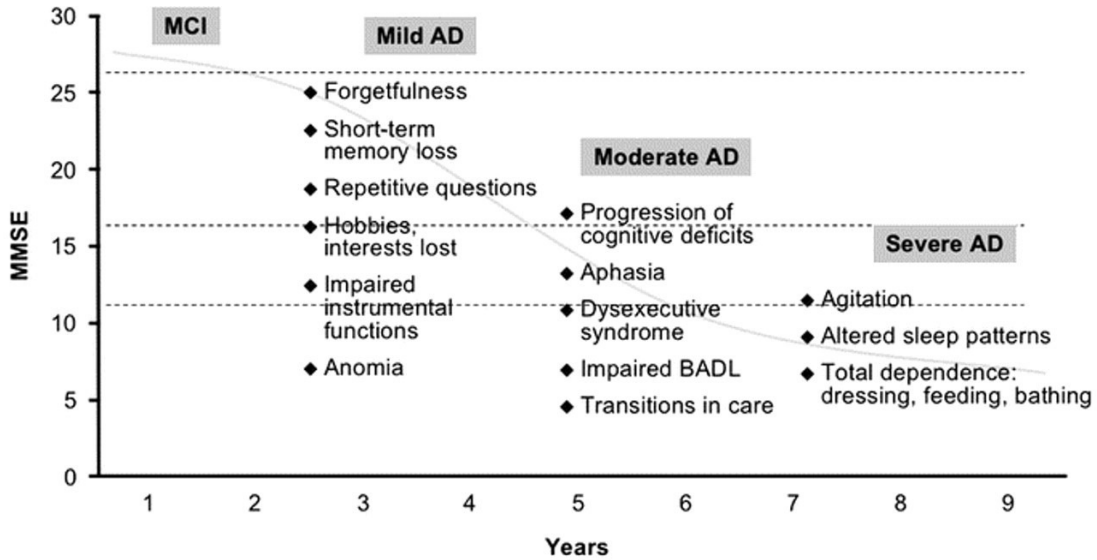
[2]Department of Psychiatry and DPUK, Oxford University

Isaac Newton Institute, 27 November 2018

Introduction

- This talk is about predicting Alzheimer's disease using brain imaging data.
- It's an application of machine learning, with feature choice as the novel aspect.
- Sections (see top of slide)
  - Alzheimer's disease (description, tests, brain regions)
  - Experiment (path signatures, machine learning)
  - Conclusion (signatures give interpretable features)

Alzheimer's disease



Symptom progression in Alzheimer disease from *The staging and assessment of moderate to severe Alzheimer disease H. H. Feldman and M. Woodward, Neurology 2005*. MMSE = Mini-Mental State Examination; MCI = mild cognitive impairment; BADL = basic activities of daily living.

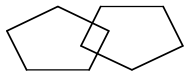
- Alzheimer's disease (AD) is an irreversible brain disorder which progressively affects cognition and behaviour, and results in an impairment in the ability to perform daily activities. The first symptom is usually memory loss.
- It is the most common form of dementia in older people and accounts for 60-80% of all cases of dementia. Dementia is a clinical syndrome manifested by difficulties in cognition and impairments in activities of daily living.
- It affects about 6% of the population aged over 65, and increases in incidence with age.

Cognitive tests

Mini-Mental State Examination (MMSE)

Patient's Name: _____ Date: _____

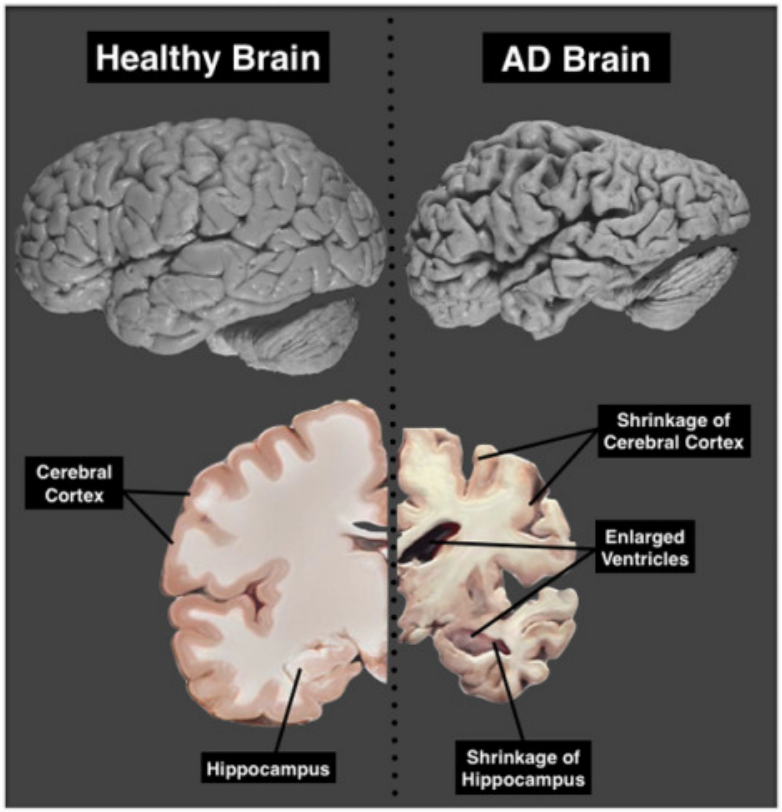
Instructions: Score one point for each correct response within each question or activity.

Maximum Score	Patient's Score	Questions
5		"What is the year? Season? Date? Day? Month?"
5		"Where are we now? State? County? Town/city? Hospital? Floor?"
3		The examiner names three unrelated objects clearly and slowly, then the instructor asks the patient to name all three of them. The patient's response is used for scoring. The examiner repeats them until patient learns all of them, if possible.
5		"I would like you to count backward from 100 by sevens." (93, 86, 79, 72, 65, …) Alternative: "Spell WORLD backwards." (D-L-R-O-W)
3		"Earlier I told you the names of three things. Can you tell me what those were?"
2		Show the patient two simple objects, such as a wristwatch and a pencil, and ask the patient to name them.
1		"Repeat the phrase: 'No ifs, ands, or buts.'"
3		"Take the paper in your right hand, fold it in half, and put it on the floor." (The examiner gives the patient a piece of blank paper.)
1		"Please read this and do what it says." (Written instruction is "Close your eyes.")
1		"Make up and write a sentence about anything." (This sentence must contain a noun and a verb.)
1		"Please copy this picture." (The examiner gives the patient a blank piece of paper and asks him/her to draw the symbol below. All 10 angles must be present and two must intersect.) 
30		TOTAL

Mild Cognitive Impairment (MCI)

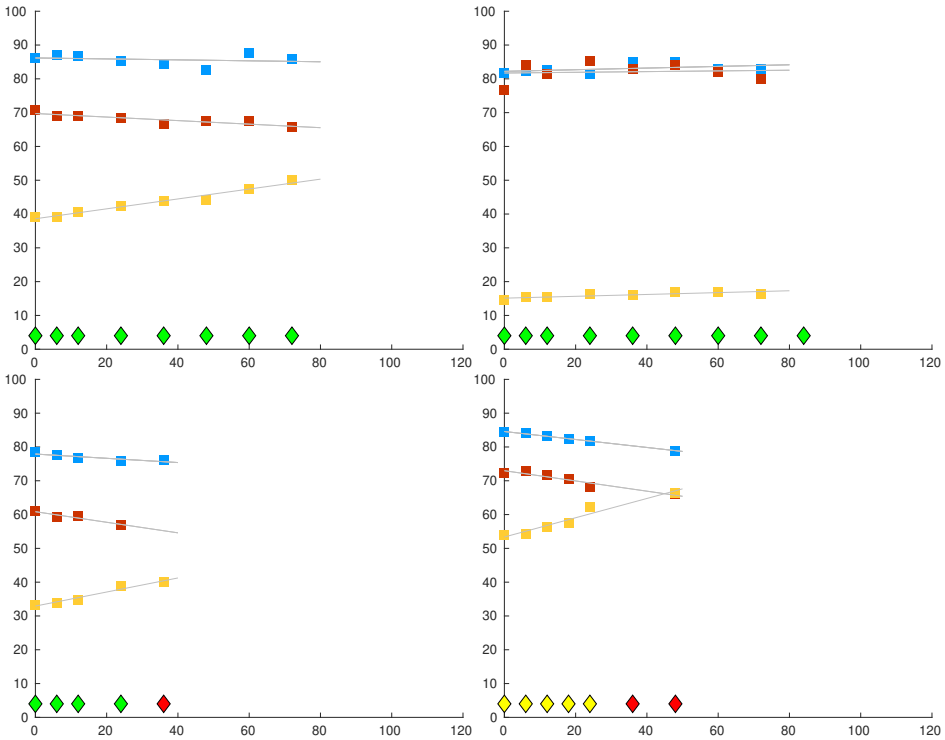
- Mild cognitive impairment (MCI), is a clinical concept, in which the individual has measurable cognitive deficits but without notable impairment in everyday living.
- A proportion of people with MCI (about 5-15%) go on to develop dementia, usually Alzheimer's disease.
- Criteria for diagnosing MCI,
  - Impaired memory function for age and education
  - Preserved general cognitive function
  - Intact activities of daily living
  - No evidence of dementia

Brain regions



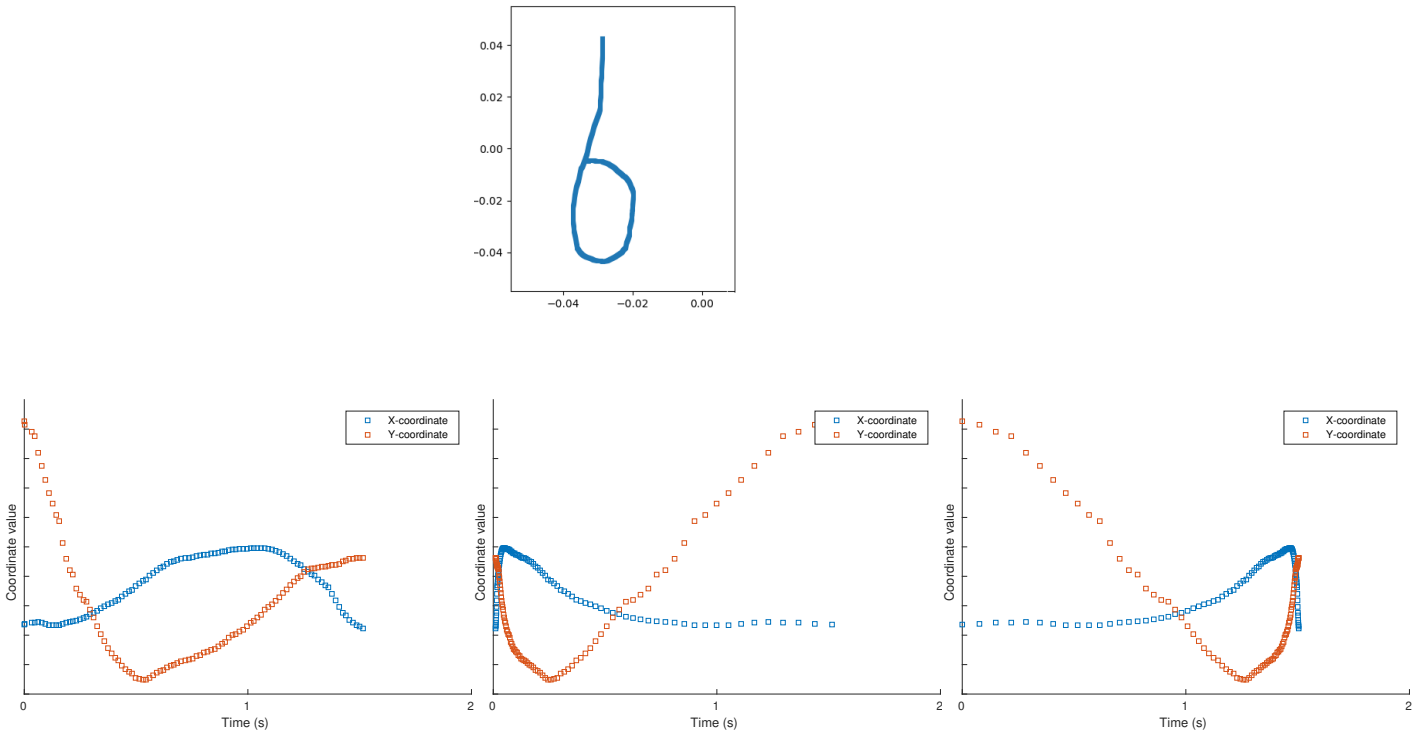
- Some brain regions of interest (ROI) are known to atrophy as a function of age and diagnosis.
- Notable among them is the hippocampus, but other regions are also affected.

Time series of brain ROI data



- Scaled brain volumes by time (months) for 4 patients: top row healthy and bottom row Alzheimer's disease.
- In each graph: Whole brain (blue), Hippocampus (red), Entorhinus (yellow).
- Diagnosis points are diamonds: Healthy (green), MCI (yellow), Alzheimer's disease (red).

Time reparameterisation



- Top: Example of a path in x and y. The letter b is drawn from top to bottom by hand. Measurement points are sampled along the path.
- Left: Original time series from the X- and Y- coordinates of the path. Middle: speed increased at start of drawing. Right: speed increased at the end of drawing.
- In contrast to features derived from the time series, the path signature uniquely characterises the original path. The signature is 1.0000,-0.0013,-0.0464,0.0000,0.0007,-0.0006,0.0011.

Path signature

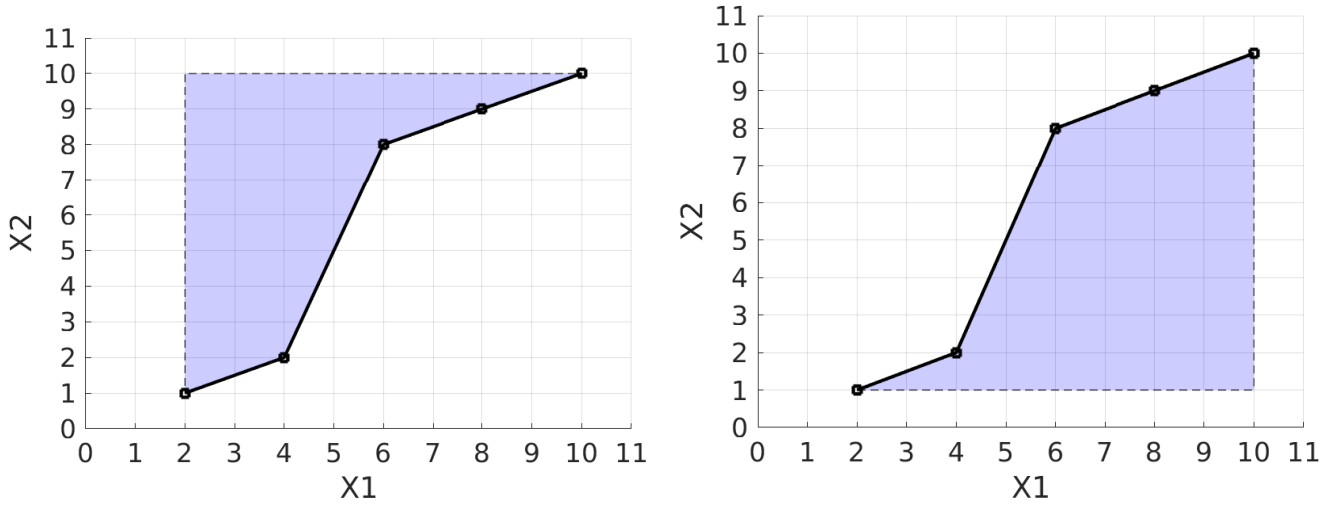
The path signature is a collection of all the iterated integrals $S(X)^{I}_{0,T}$ of X,

$$S(X)_{0,T} = (1, S(X)^{1}_{0,T}, S(X)^{2}_{0,T}, \ldots, S(X)^{d}_{0,T}, S(X)^{1,1}_{0,T}, S(X)^{1,2}_{0,T}, S(X)^{2,2}_{0,T}, \ldots) \tag{1}$$

For four dimensions with order two, the signature terms are,

$$S(X)_{0,T} = 1, (1), (2), (3), (4), (1,1), (1,2), (1,3), (1,4), (2,1), (2,2) \ldots \tag{2}$$

Signature interpretation



- The line in both left and right hand figures is the path in 2D with variables $X1$ and $X2$.
- The shaded area shows cross terms in the signature. The left hand figure shows $S^{(X1,X2)}$ with value 31. The right hand figure on the right shows $S^{(X2,X1)}$ with a value of 41. These terms both measure the deviation from a straight line and they distinguish cases where $X1$ increases quickly relative to $X2$ and *vice versa*.

Method

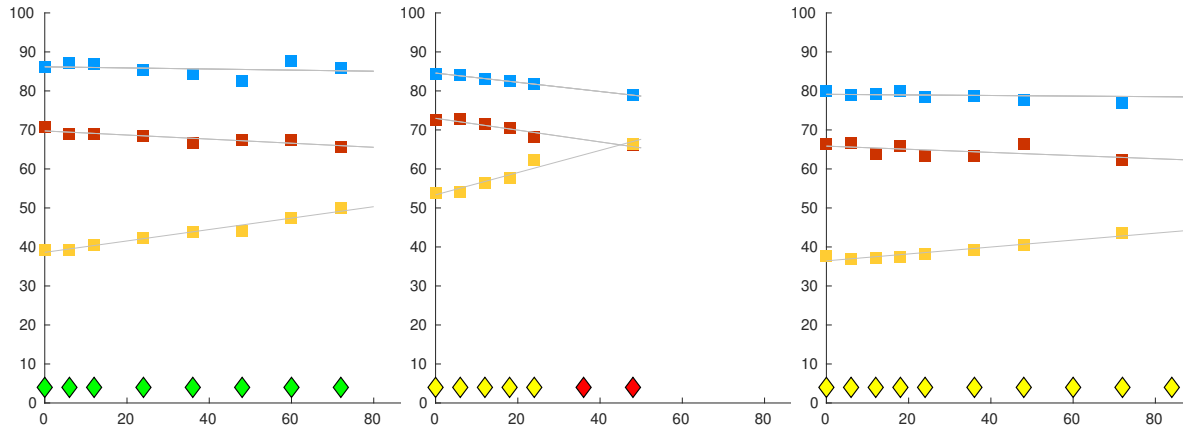
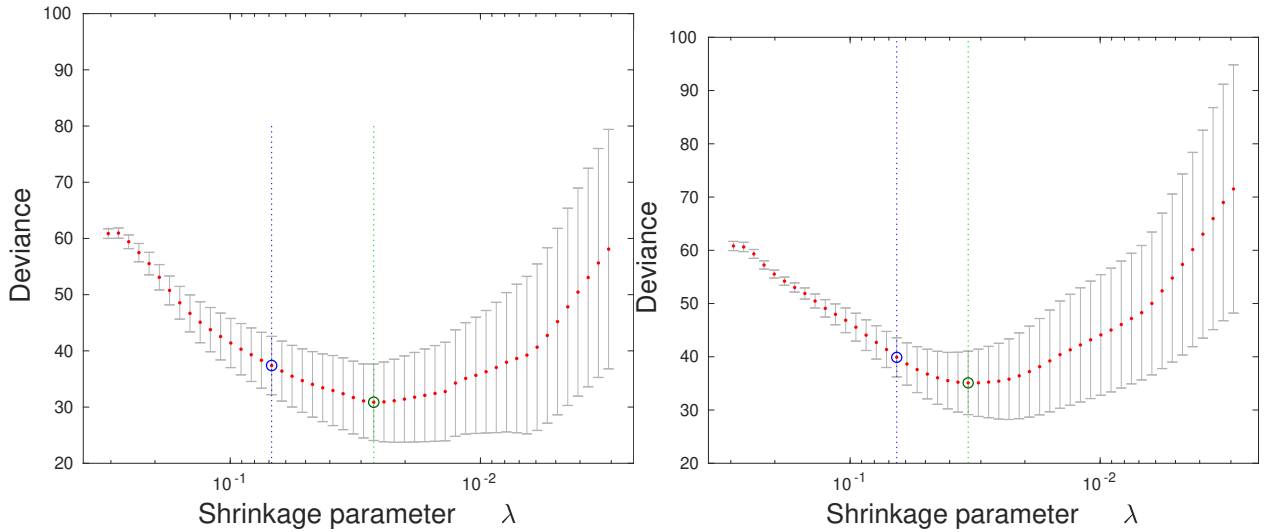


Figure: Time series from participants. Left: Healthy, Middle: Alzheimer's, Right: MCI.

- The task is to identify the future diagnosis of an individual using the MRI time series of length 24 months.
- We classify Alzheimer's vs Healthy and Alzheimer's vs MCI using binary logistic regression.
- The feature vector is formed from three variables: WholeBrain, Hippocampus and Ventricles, and the path signature derived from these variables and their time points. The signature of a path in 4 dimensions and of degree 2 has a length of 20. Lasso regularisation is used for feature selection.

Training



- Training curves for classification. Left: Alzheimer's vs Healthy. Right: Alzheimer's vs MCI.
- In each case the graph shows the optimisation of the Lasso shrinkage parameter λ using 10-fold cross validation.
- The green circle and line show the minimum cross validation error, and the blue circle and line locate the value of λ with minimum cross-validation error plus one standard deviation.

Results

	Alzheimer's - Healthy	*Alzheimer's - MCI*
Variables	Hippocampus - baseline	Hippocampus - baseline Ventricles - baseline
Increments	(Time) (Ventricles)	
Cross-terms	(Time,Ventricles) (Hippocampus,Time) (Whole brain,Hippocampus) (Whole brain,Time)	(Time,Ventricles) (Hippocampus,Time) (Whole brain,Hippocampus) (Hippocampus,Hippocampus)

- The set of variables selected by Lasso. The complete set is formed of the baseline values of variables WholeBrain, Hippocampus and Ventricles, and path signature terms. The path signature is derived from these variables with time added as an additional variable.

Conclusion

- We can distinguish MRI time series from patients with a subsequent diagnosis of Alzheimer's disease, MRI or healthy.
- The identification of an interaction between the whole brain volume and the hippocampus volume is interesting and deserves further exploration.
- While nonlinear classifiers (neural nets, random forests) can be accurate, their function can be difficult to interpret. By encoding nonlinearity into the features, we can use classifiers that give more interpretable results.