

# Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting state fMRI

Edward Challis<sup>1,2</sup>, Barbara Spano<sup>3</sup>, Laura Serra<sup>3</sup>, Marco Bozzali<sup>3</sup>, Seb Oliver<sup>1</sup>, and Mara Cercignani<sup>2</sup>

<sup>1</sup>Physics and Astronomy, University of Sussex, Brighton, Sussex, United Kingdom, <sup>2</sup>Clinical Imaging Sciences Centre, Brighton and Sussex Medical School, Brighton, Sussex, United Kingdom, <sup>3</sup>Neuroimaging Laboratory, IRCCS Santa Lucia, Rome, Italy

## Target audience

Clinical and non-clinical MRI scientists with an interest in the application of machine-learning algorithms to MRI data for the purposes of patient stratification.

## Purpose

To investigate the efficacy of multivariate machine learning (ML) techniques to perform patient stratification from functional connectivity patterns of brains at rest. Whilst such analyses are not unique, the majority of previous studies have employed support vector machines (SVMs)<sup>1,2,3,4,5</sup>. In this study we investigate the performance of Bayesian Gaussian process logistic regression (GP-LR) models with non-linear kernels. The GP-LR model can be interpreted as a Bayesian probabilistic analogue to SVM kernel classifiers<sup>6,7</sup>. GP-LR confers a number of benefits over kernel SVMs: since GP-LR is probabilistic it provides a principled estimate of the probability of class membership, such estimates can be useful if we wish to ascertain the confidence the model places in its predictions, additionally non-symmetric classification thresholds can be set to achieve either strong specificity or sensitivity scores; since GP-LR models are Bayesian, computationally expensive cross-validation hyper-parameter grid-search methods can be avoided.

## Materials and methods

**Participants:** 86 subjects, 37 with a diagnosis of probable Alzheimer's disease (AD) (M/F= 29%, mean age = 71, s.d. = 7 years) and 49 with a diagnosis of amnesic mild cognitive impairment (a-MCI) (M/F= 53%, mean age = 70, s.d. = 9 years) were enrolled for this study. 32 age and gender matched healthy controls (NC) were also recruited (M/F= 56%, mean age = 62, s.d. = 18 years). A global cognitive assessment was obtained for all participants using the Mini-Mental State Examination (MMSE).

**MRI:** All subjects underwent an MRI examination at 3T, including: 3D modified driven equilibrium Fourier transform (MDEFT) scan ( $TR = 1338\text{ ms}$ ,  $TE = 2.4\text{ ms}$ ); and  $T2^*$  weighted echo planar image (EPI) sensitised to Blood Oxygenation Level Dependent (BOLD) imaging contrast ( $TR = 2080\text{ ms}$ ,  $TE = 30\text{ ms}$ , 32 axial slices, 220 volumes) for Resting State fMRI (RS-fMRI).

**Image analysis:** RS-fMRI data were processed using Matlab7 and SPM8. The preprocessing steps included realignment, slice-timing correction, and normalisation. In-house software was used to remove the other sources of bias (global temporal drift, realignment parameters and the global signal). Data were then filtered by a phase-insensitive band-pass filter (pass band 0.01 to 0.08 Hz) and smoothed with a 3D Gaussian Kernel with  $8\text{ mm}^3$  FWHM.

**Feature construction:** Each subject's RS-fMRI scan was then converted to a brain region connectivity feature vector. Connectivity feature vectors were defined as the covariance in BOLD signal between each of the 82 anatomically distinct Regions of Interest (ROIs) as defined by the AAL brain atlas<sup>8</sup> (cerebellum and sensory cortex ROIs were excluded). Each subject's age and MMSE score was appended to their covariance feature vector.

**Machine learning classifier:** GP-LR binary classification was performed between: healthy control (NC) and a-MCI, NC and AD and a-MCI and AD subjects. The GP-LR covariance function was defined as the sum of the squared exponential and the additive white noise covariance functions. Hyper-parameters were estimated by numerically optimising the Bayesian evidence (approximated using the expectation propagation algorithm<sup>9</sup> and implemented using the gpml package<sup>10</sup>). To reduce the risk of over-fitting, two feature selection strategies were used: First we supplied the GP-LR classifier with only 10 features that obtained the highest Kendall tau correlation coefficient versus the class label, second we used Automatic Relevance Determination (ARD) GP covariance functions that automatically down-weight less relevant features.

## Results

Model evaluation metrics were obtained by averaging over  $N$  leave-one-out train and test set partitions of the data. Results are presented in the table below. The Bayesian evidence column is the averaged approximated log Bayesian evidence score – a larger value corresponds to a better model fit. The A.U.C. column reports the area under the receiver operated characteristic curve<sup>11</sup> (a perfect classifier achieves an A.U.C. of 1 a random classifier an A.U.C. of 0.5). The predicted probability score is the average probability the model assigned to the true class label of the test points. The accuracy, sensitivity and specificity scores are obtained by selecting the classification threshold to minimise the misclassification cost under a symmetric cost function. The results are weakest for the NC versus a-MCI classification task suggesting that this is a harder problem than discerning AD subjects. The NC versus a-MCI task results are stronger than those that can be obtained using a linear covariance function (data not shown) suggesting that differences between NC and a-MCI connectivity do not follow a simple linear relation.

Binary Task	Bayes. Evidence	A.U.C	Predicted prob.	Accuracy (%)	Specificity (%)	Sensitivity (%)
NC vs. a-MCI	-30.36	0.92	0.82	90.1	87.5	91.8
NC vs. AD	-19.62	0.96	0.84	91.3	84.4	97.3
a-MCI vs. AD	-16.66	0.98	0.90	90.7	98.0	81.1

The ROI covariances included in the feature vector for the NC versus a-MCI classification task are between: the caudate nucleus and the frontal superior gyrus, the mid cingulum and the insula gyrus, the mid cingulum and the Heshl right gyrus and the Rolandic operculum and the mid cingulum.

## Discussion

One of the primary strengths of the GP-LR approach presented here is that the Bayesian evidence provides a simple and principled means by which to optimise hyper-parameters. The model above has 13 hyper-parameters to tune – tuning this many parameters would be infeasible for a grid search nested cross validation approach that would be needed using a kernel SVM. A further benefit is that the model supplies a probability estimate of predicted class membership whose value could be thresholded to achieve either high specificity or sensitivity scores. Future work will investigate the suitability of such a confidence score to predict, for example, the conversion to AD in a short time for a-MCI patients.

## References

[1] R C Craddock et al., Magnetic Resonance in Medicine, 62(6):1619-1628, 2009; [2] L Zeng et al., Brain, 135:1498-1507, 2012; [3] T B Meier et al., NeuroImage, 60(1):601-613, 2012; [4] Y Fan et al., NeuroImage, 56(4):2058-2067, 2011; [5] D A Fair et al., Frontiers in Systems Neuroscience, 6:80, 2012; [6] F Pereira et al., NeuroImage, 45(1):S199-S209, 2009; [7] C E Rasmussen and C K I Williams, Gaussian processes for machine learning, 2006; [8] N Tzourio-Mazoyer et al., NeuroImage, 15(1):273-289, 2002; [9] H Nickisch and C E Rasmussen. The Journal of Machine Learning Research, 9:2035-2078, 2008; [10] C E Rasmussen and H Nickisch. The Journal of Machine Learning Research, 11:3011-3015, 2010; [11] A P Bradley, Pattern Recognition 30.7:1145-1159, 1997.