

# An Oblique Approach to Prediction of Conversion to Alzheimer's Disease with Multikernel Gaussian Processes

Jonathan Young<sup>1</sup>(✉), Marc Modat<sup>1</sup>, Manuel J. Cardoso<sup>1</sup>, John Ashburner<sup>2</sup>,  
and Sebastien Ourselin<sup>1,3</sup>

<sup>1</sup> Centre for Medical Image Computing, University College London, London, UK  
jonathan.young@ucl.ac.uk

<sup>2</sup> Wellcome Trust Centre for Neuroimaging, Institute of Neurology,  
University College London, London, UK

<sup>3</sup> Dementia Research Centre, Institute of Neurology,  
University College London, London, UK

**Abstract.** Machine learning approaches have had some success in predicting conversion to Alzheimer's Disease (AD) in subjects with mild cognitive impairment (MCI), a less serious condition that nonetheless is a risk factor for AD. Predicting conversion is clinically important as because novel drugs currently being developed require administration early in the disease process to be effective. Traditionally training data are labelled with discrete disease states; which may explain the limited accuracies obtained as labels are noisy due to the difficulty in providing a definitive diagnosis of Alzheimer's without post-mortem confirmation, and ignore the existence of a continuous spectrum of disease severity. Here, we dispense with discrete training labels and instead predict the loss of brain volume over one year, a quantity that can be repeatably and objectively measured with the boundary shift integral and is strongly correlated with conversion. The method combines MRI and PET image data and cerebrospinal fluid biomarker levels in an Bayesian multi-kernel learning framework. The resulting predicted atrophy separates converting and non-converting MCI subjects with 74.6% accuracy, which compares well to state of the art methods despite a small training set size.

**Keywords:** Gaussian processes · Regression · Atrophy · BSI · Multi-kernel learning · MRI · PET Alzheimer's disease · Mild cognitive impairment

## 1 Introduction

In the study of AD, in recent years an increasing emphasis has been placed on the importance of early diagnosis. This is because while currently available treatments are only able to mitigate the downstream effects of the disease process, pending ones are focused on actually disrupting the disease process itself, by interfering with the amyloid cascade that is thought to be one of the underlying causes of AD [1]. To be effective, such treatments would have to begin

before the patient is showing the full symptoms of AD. This involves studying patients who have mild cognitive impairment (MCI). Clinically, MCI is defined as having isolated memory deficits that are not severe enough to affect normal living [2]. MCI patients convert to AD at an annual rate of 10–15% per year [3], although some develop other diseases or remain stable. As stable and converting MCI (MCI-s and MCI-c) patients by definition have similar symptoms, standard cognitive tests used to diagnose AD are by themselves of little help for this problem; instead, imaging and other biomarkers can be used with machine learning methods to detect subtle differences between the groups. A classifier can be trained on labeled examples of MCI-s and MCI-c images, or alternatively on examples of AD patients and healthy controls (HC), under the assumption that MCI-s subjects are more HC like and MCI-c subjects are more AD like. Most such studies use magnetic resonance imaging (MRI), from which a variety of features can be extracted. However the results can be improved by combining MRI features with imaging data measuring metabolic activity using fluorodeoxyglucose positron emission tomography (FDG-PET) and biomarkers measured in a sample of cerebrospinal fluid (CSF) or genetic information in a multi-kernel framework [4, 5].

A limiting factor in the accuracy these studies may be mislabeling of training subjects. The gold standard for diagnosis of AD is autopsy, but most studies use subjects whose diagnosis has been determined by standard clinical testing, which has been shown to have an error rate of at least 10% [6]. This is an issue that has not been widely addressed; Aksu et al. [7] point out that training labels for MCI-s and MCI-c are uncertain and go on to generate their own MCI training labels by following the classification of MCI subjects by an HC versus AD classifier across multiple timepoint. However even this neglects the uncertainty in the HC and AD labels this scheme ultimately depends on.

Our proposed method follows [8] in abandoning discrete disease state labels for training altogether. We also perform a regression to predict a continuous proxy for disease status, but instead of age we use atrophy over a period of one year as measured by the boundary shift integral (BSI) [9]. This then provides a predicted atrophy rate for each test subject. We use Gaussian process (GP) regression [10], with a multiple kernel framework to optimally combine MRI, FDG-PET and CSF data. This results in a measure that can predict MCI conversion within 3 years with a balanced accuracy of 74.6%, as good as state of the art techniques having a much larger training set, including our own previous work using multikernel GPs for classification [5].

## 2 Materials and Methods

### 2.1 Image and Biomarker Data

All data were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database<sup>1</sup>. The MRI images were T1 weighted structural scans from

<sup>1</sup> <http://adni.loni.ucla.edu/>.

1.5T scanners taken at baseline and 12 month follow-up. All were subjected to quality control and automatically corrected for spatial distortion caused by gradient nonlinearity and B1 field inhomogeneity.

FDG-PET images were acquired according to the ADNI protocol: acquired 30–60 min post-injection, averaged, spatially aligned, interpolated to a standard voxel size, intensity normalized, and smoothed to a common resolution of 8-mm full width at half maximum.

CSF samples were obtained from subjects by a lumbar puncture around the time of their baseline scan. Levels of the proteins amyloid- $\beta_{42}$  ( $a\beta_{42}$ ), tau, and phosphorylated tau were measured and recorded.

The original ADNI project collected baseline structural MRI scans for all subjects. However FDG-PET scanning and collection of CSF data were only done on subsets of these subjects. Furthermore, calculation of BSI requires a 12-month follow-up structural MRI, which were also missing for some subjects. As our method requires FDG-PET and CSF and a 12-month BSI as well as structural MRI data, only 129 subjects could be included in the study. The details of these are shown in Table 1. Subjects were classified as HC, AD or MCI by neuropsychological and clinical testing at the time of the baseline scan, with MCI conversion status decided by whether subjects were subsequently diagnosed as AD at any stage during the 36 month follow-up period.

**Table 1.** Subject groups and demographics

Disease status	Number	Female	Mean age (sd)
HC	28	19	74.1 (4.5)
MCI-s	38	22	75.3 (7.3)
MCI-c	29	18	75.1 (7.4)
AD	34	23	75.1 (6.8)

## 2.2 Image Processing

Probabilistic grey matter (GM) maps were produced from the native space baseline scans using the NiftySeg tool [11]. The native space images were also anatomically parcellated into 83 regions with a novel label fusion algorithm [12] in a multi-atlas label propagation scheme. The resulting parcelations were used to mask out the brainstem and cerebellum from the native space GM segmentations.

Also, a custom template was produced, performing all registrations using the NiftyReg toolkit [13]. The native GM space segmentations were then warped into the groupwise space. Finally, the groupwise space, masked GM segmentations were modulated by the Jacobian determinants of this final deformation. This step ensures the total mass of tissue remains constant. Hence the MRI features used were voxel level GM probabilities.

The native space anatomical parcelations were also rigidly transferred to the space of the FDG-PET images for the corresponding subjects. The parcelation was used to normalise each FDG-PET image by its mean cerebellar activity, and then to calculate the mean activity within each anatomical region, generating a set of 83 features for each FDG-PET image.

### 2.3 Boundary Shift Integral

The BSI is a method for robustly assessing volume loss of whole brains or brain regions from structural MRI. It calculates a change in volume by integrating across the longitudinal change in position of the boundary between CSF and GM surrounding the region of interest. Preprocessing is needed to extract the region of interest (which in our case is the whole brain) from each image, linearly align the baseline and follow-up images, and correct for intensity inhomogeneity between scans. We use the latest version of BSI [9] which uses a symmetric registration scheme to minimise bias and maximise desirable qualities for an atrophy measurement such as inverse consistency and transitivity between multiple time-points.

We normalise the resulting volume changes by the baseline brain volumes and by the exact interval between baseline and follow-up scans, and multiply by 100. This produces a normalised brain atrophy rate (BAR) in percentage of original brain volume per year for each subject. These are then used as targets in the following regression analysis.

### 2.4 Gaussian Process Regression

Gaussian processes (GPs) provide a Bayesian, kernelised framework for solving both regression and classification problems. As an in depth explanation of GPs is beyond the scope of this paper, we refer the reader to [10] for a theoretical treatment and our previous work [5] for an application of multi-kernel GPs to predicting conversion to AD.

Briefly, however, a GP is a multivariate Gaussian forming the prior on the value of a latent function, on top of which is put a Gaussian noise model. The covariance of the prior is a function of the covariance between instances of training data  $\mathbf{X}$ , and a set of hyperparameters  $\theta$  that control the overall form of the prior and the noise variance. During the training phase, the hyperparameters are learned from the training data  $\mathbf{X}$  and targets  $\mathbf{y}$  by type-II maximum likelihood. Once the hyperparameters have been set, predictions on unseen data are made by integrating across this prior, which can be calculated analytically for the regression case.

### 2.5 Gaussian Processes as Multimodal Kernel Methods

GP regression is based on a covariance kernel  $K$ , a symmetric positive definite matrix where entry  $K_{ij}$  is given by a covariance kernel function  $k$  of the feature

vectors for the  $i$ th and  $j$ th subject  $\mathbf{x}_i$  and  $\mathbf{x}_j$  and a hyperparameter or hyperparameters  $\theta$ . We use a linear kernel function, which is simply the scalar product of  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . As GPs belongs to the family of kernel methods, a positive sum of valid kernels is a valid kernel, and a valid kernel multiplied by a positive scalar is also a valid kernel. This implies that to do multimodal classification, we can define our kernel function as the weighted sum of a number of subkernels, each of which has been calculated from a the feature vectors representing a particular type of data or modality for each subject. Each subkernel has a scaling hyperparameter  $\alpha$  representing the modality’s weight in the overall kernel. A bias term  $\beta$  is also included in the sum. So in the case of multimodal classification using information derived from the MRI, PET and CSF data for each subject the overall kernel is

$$K_{ij} = \alpha_{\text{MRI}}(\mathbf{x}_{i,\text{MRI}} \cdot \mathbf{x}_{j,\text{MRI}}) + \alpha_{\text{PET}}(\mathbf{x}_{i,\text{PET}} \cdot \mathbf{x}_{j,\text{PET}}) + \alpha_{\text{CSF}}(\mathbf{x}_{i,\text{CSF}} \cdot \mathbf{x}_{j,\text{CSF}}) + \beta \quad (1)$$

giving a total of four covariance hyperparameters to set.

### 3 Results

To generate predicted BARs for all 129 subjects, we perform a leave-one-out cross validation (LOOCV) across the entire set. The correlation coefficient between predicted and measured BARs for the subjects is 0.38 ( $p < 0.0001$ ) and the root mean squared error is 0.61. However our primary focus is not on the predicted brain atrophy rates themselves, but on whether they can be used to predict conversion in MCI subjects. Figures 1 and 2 show the spread of both measured and predicted BAR values for all four disease groups (HC, MCI-s, MCI-c, AD).

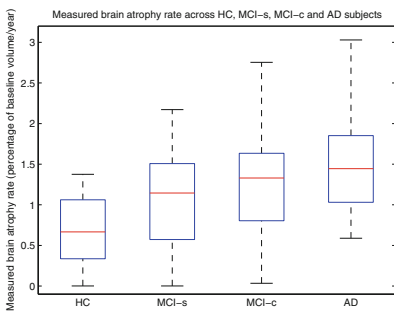


Fig. 1. Measured BAR across groups

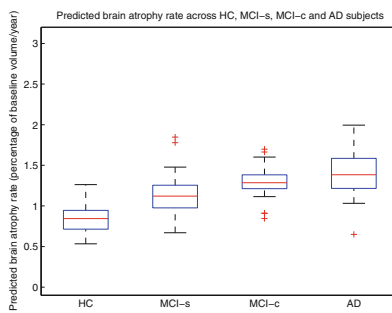


Fig. 2. Predicted BAR across groups

As shown in Figs. 1 and 2, while the mean predicted BARs for each group are similar to the corresponding means for measured BARs, each clinical group occupies a much tighter cluster of values, even allowing for a few outliers (marked as

**Table 2.** Accuracy of discrimination between MCI-s and MCI-c with predicted BAR

Modalities	Accuracy (%)	AUC
MRI	59.7	0.595
PET	73.1	0.777
CSF	52.2	0.545
MRI, PET, CSF	74.6	0.725

**Table 3.** Accuracy of discrimination between MCI-s and MCI-c with training on binary diagnostic class labels

Training	Accuracy (%)	AUC
MCI (CV)	40.3	0.401
HC, MCI, AD (CV)	52.2	0.569
HC, AD	55.2	0.661

a +). This results in reduced overlap between the clinical groups, which is especially noticeable between the MCI-s and MCI-c groups. The resulting accuracy is 74.6%, which is similar to the best previously reported results. The balanced accuracy and area under the ROC curve (AUC) are shown in Table 2. This also shows results for single modalities, demonstrating the benefit of combining sources of data with multikernel learning.

We also compare our method to performing direct binary classification on the conversion status again using GPs. This can be done by training on the MCI subjects only in an LOOCV loop, by training on all subjects, again with an LOOCV loop and grouping HC subjects with MCI-s and MCI-c subjects with AD, and finally by training on the HC and AD subjects, and testing on the MCI subjects. The results are given in Table 3.

## 4 Discussion

These results show a clear advantage for our method of training on a well-characterised proxy for MCI conversion, rather than the diagnostic status itself. Training on BAR enables us to reach accuracies of up to 74.6%, whereas training on diagnostic labels struggles to perform better than chance. It therefore appears that the use of BAR bypasses the problems caused by binary diagnostic labels. Data is made better use of as subjects can be used for training regardless of diagnostic label, and as parameters are learned automatically there is no need to set subjects aside for tuning. We also show an advantage for multimodal regression. Although direct comparisons between methods are difficult [5], the resulting accuracy in forecasting MCI conversion is among the best yet achieved. The main drawback of our the proposed method is that all three types of data are all required for the best results (although FDG-PET alone does almost as well) which limits the number of subjects that can be included. However we intend to further evaluate this method as much greater numbers of subjects with all modalities become available in ADNI 2. Finally, while 12-month follow-up scans are also required to calculate BSI values for training data, it should be emphasised they are *not* needed for testing data.

**Acknowledgments.** We would like to thank Dr Kelvin Leung of the Dementia Research Centre, University College London for his assistance and provision of BSI data.

## References

1. Robert, R., Wark, K.L.: Engineered antibody approaches for Alzheimer's disease immunotherapy. *Arch. Biochem. Biophys.* **526**(2), 132–138 (2012)
2. Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E.: Mild cognitive impairment: clinical characterization and outcome. *Arch. Neurol.* **56**(3), 303–308 (1999)
3. Braak, H., Braak, E.: Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol. Aging* **16**(3), 271–278 (1995)
4. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage* **55**(3), 856–867 (2011)
5. Young, J., Modat, M., Cardoso, M.J., Mendelson, A., Cash, D., Ourselin, S.: Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *NeuroImage: Clin.* **2**, 735–745 (2013)
6. Beach, T.G., Monsell, S.E., Phillips, L.E., Kukull, W.: Accuracy of the clinical diagnosis of Alzheimer disease at national institute on aging Alzheimer disease centers, 2005–2010. *J. Neuropathol. Exp. Neurol.* **71**(4), 266–273 (2012)
7. Aksu, Y., Miller, D.J., Kesidis, G., Bigler, D.C., Yang, Q.X.: An MRI-Derived definition of MCI-to-AD conversion for long-term, automatic prognosis of MCI patients. *PLoS ONE* **6**(10), e25074 (2011)
8. Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H.: Alzheimer's disease neuroimaging initiative: BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer disease. *PLoS ONE* **8**(6), e67346 (2013)
9. Leung, K.K., Ridgway, G.R., Ourselin, S., Fox, N.C.: Consistent multi-time-point brain atrophy estimation from the boundary shift integral. *NeuroImage* **59**(4), 3995–4005 (2012)
10. Rasmussen, C.E., Williams, C.K.I.: *Gaussian Processes for Machine Learning*. MIT Press, Cambridge (2006)
11. Cardoso, M.J., Clarkson, M.J., Ridgway, G.R., Modat, M., Fox, N.C., Ourselin, S.: LoAd: a locally adaptive cortical segmentation algorithm. *NeuroImage* **56**, 1386–1397 (2011)
12. Cardoso, M., Modat, M., Ourselin, S., Keihaninejad, S., Cash, D.: Multi-STEPS: multi-label similarity and truth estimation for propagated segmentations. In: 2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), pp. 153–158, January 2012
13. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. *Comput. Methods Programs Biomed.* **98**(3), 278–284 (2010)