Dementia classification based on brain age estimation

Katja Franke[1], Christian Gaser[1]

[1] Structural Brain Mapping Group, Departments of Neurology & Psychiatry, Jena University Hospital, Jena, Germany
{katja.franke, christian.gaser}@uni-jena.de

Abstract. Early identification of neuroanatomical changes deviating from the normal age-related atrophy pattern has the potential to improve clinical outcomes through early treatment or prophylaxis. Especially the pathological cascade of Alzheimer's disease (AD), the most common form of dementia, is widely linked to precocious and/or accelerated (brain) aging.

This work presents a novel magnetic resonance imaging (MRI)-based biomarker that indicates discrepancies in individual brain aging. By employing automatic preprocessing of structural MR images as well as high-dimensional pattern recognition methods, this approach uses the distribution of normal brain-aging patterns to estimate the individual brain age of a given new subject. The difference between the estimated brain age and the chronological age gives an individual deviation score, with positive values indicating the degree of acceleration in cerebral atrophy, which is considered a risk factor for AD. Here, this deviation score is used to classify the subjects as NO, MCI, and AD.

Keywords: MRI, relevance vector machines (RVM), support vector machines (SVM), regression, aging, brain disease

1 Introduction

The global prevalence of dementia is projected to rise sharply over the coming decades. By 2050, 1 in 85 persons worldwide will be affected by Alzheimer's disease (AD), the most common form of dementia (Brookmeyer et al. 2007). Manifold pathological changes begin to develop years or decades before the onset of cognitive decline (Jack et al. 2010), including premature changes in gene expression (Cao et al. 2010; Saetre et al. 2011), accelerated age-associated changes of the default mode network (Jones et al. 2011), and most obviously, abnormal changes in brain structures already at the mild cognitive impairment (MCI) stage (Driscoll et al. 2009; Spulber et al. 2010). Additionally, atrophic regions detected in AD patients were recently found to largely overlap with those regions showing a normal age-related decline in healthy control subjects (Dukart et al. 2011).

Early detection and quantification of abnormal brain changes is important for the prospective identification and subsequent treatment of individuals at risk for cognitive decline and dementia. At some point in the AD disease course accelerated neurodegeneration takes place, preceding accelerated cognitive decline (Jack et al. 2010). Although brain atrophy in general is not specific for AD, MRI-detected atrophy was found to retain the closest relationship with cognitive decline (Jack et al. 2010; Vemuri et al. 2009a, b) suggesting a crucial role for structural MRI in predicting future conversion to AD (Frisoni et al. 2010; Jack et al. 2010).

Assuming AD to be preceded by precocious and/or accelerated brain aging (Bartzokis 2011; Driscoll et al. 2009; Spulber et al. 2010), a straightforward and efficient solution is to model healthy aging on the one hand, and to identify accelerated (thus pathological) brain atrophy on the other. Consequently, in order to recognize faster brain atrophy, a model of healthy and normal brain aging is needed. A straightforward and efficient solution is to model age regression based on normal brain anatomy, such that an individual's brain age can be accurately estimated from his/her brain scan alone.

The approach presented here takes into account the widespread but sequential age-related brain tissue loss. Based on single time-point structural MRI the complex, multidimensional aging patterns across the whole brain are aggregated to one single value, i.e. the estimated brain age. Consequently, although using only a standard MRI scan, the deviation in brain atrophy from normal brain aging can be directly quantified.

2 Methods

2.1 Subjects/Database

To train the age estimation framework, MRI data from the publicly accessible IXI cohort (http://www.brain-development.org) were used. In September 2011, the IXI database contained T_1-weighted images from 561 healthy subjects [250 male] aged 20-86 years [mean (SD) = 48.6 (16.5) years], which were collected on three different scanners (Philips 1.5T, General Electric 1.5T, Philips 3T). The distributions of age and gender within the training sample are shown in Figure 1. Table 1 shows the characteristics of the *CAD Dementia* training and test samples.

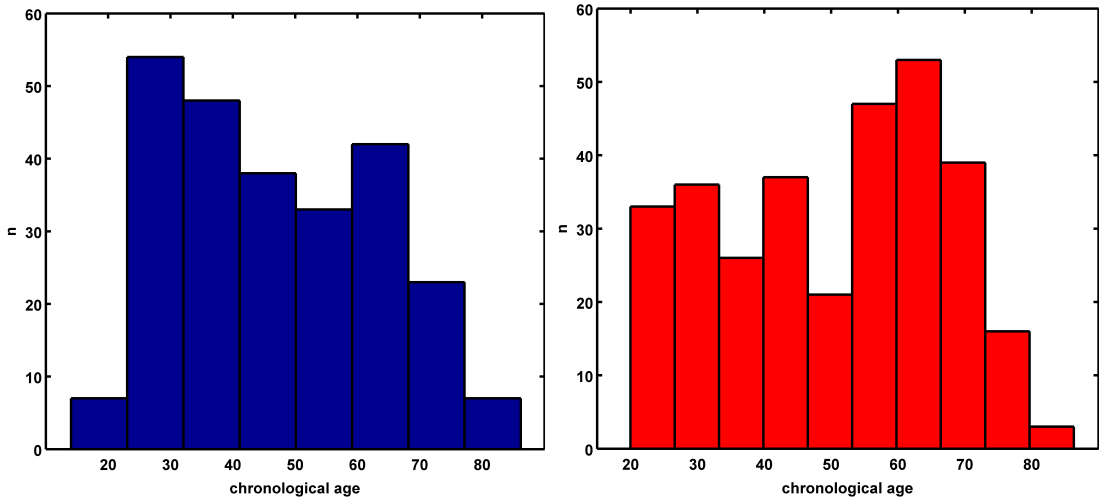

Figure 1: Age distribution in the male (left) and female (right) IXI sample used for modeling normal brain aging.

Table 1: Characteristics of the subjects in the CAD Dementia training and test samples.

CAD Dementia	*train sample*		*test sample*	
	male	female	male	female
No. subjects	17	13	213	141
Age mean (SD)	64.2 (6.7)	66.5 (7.4)	65.7 (7.4)	64.2 (8.3)
Age range	54 – 79	57 – 80	49 – 88	46-88
No. NO subjects	9	3	-	-
No. MCI subjects	5	4	-	-
No. AD subjects	3	6	-	-

2.2 Preprocessing of MRI data and data reduction

Preprocessing of the T_1-weighted images was done using the SPM8 package (http://www.fil.ion.ucl.ac.uk/spm) and the VBM8 toolbox (http://dbm.neuro.uni-jena.de), running under MATLAB. All T1-weighted images were corrected for bias-field inhomogeneities, then spatially normalized and segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) within the same generative model (Ashburner and Friston 2005). The segmentation procedure was extended by accounting for partial volume effects (Tohka et al. 2004), by applying adaptive maximum a posteriori estimations (Rajapakse et al. 1997), and by using a hidden Markov random field model (Cuadra et al. 2005; Gaser 2009). The images were processed with affine registration and smoothed with *(i)* 4-mm full-width-at-half-maximum (FWHW) or *(ii)* 8-mm FWHW smoothing kernels. Spatial resolution was set to *(i)* 4 mm or *(ii)* 8 mm, respectively. For further data reduction, principal component analysis (PCA) was performed on the training sample with subsequently applying the estimated transformation parameters to the test sample. PCA was done using the 'Matlab Toolbox for Dimensionality Reduction' (http://ict.ewi.tudelft.nl/~lvandermaaten/Home.html).

2.3 Relevance vector regression (RVR)

To capture the complex, multidimensional aging pattern across the whole brain, the brain age estimation framework utilizes relevance vector regression (RVR; Tipping, 2001). Relevance vector machines (RVM) were introduced as a Bayesian alternative to support vector machines (SVM) for obtaining sparse solutions to pattern recognition tasks. The main idea behind SVMs is the transformation of training data from input space into high-dimensional space – the feature space – via a mapping function Φ (Bennett and Campbell 2003; Schölkopf and Smola 2002). For the purpose of classification, the hyperplane that best separates the groups is computed within this feature space, resulting in a nonlinear decision boundary within the input space. The best separating hyperplane is found by maximizing the margin between the two groups. The data points lying on the margin boundaries are called support vectors since only these are used to specify the optimal separating hyperplane. For the case of real-valued output functions (rather than just binary outputs as used in classification), the SV algorithm was generalized to regression estimation (Bennett and Campbell 2003; Schölkopf and Smola 2002). In support vector regression (SVR), a function has to be found that fits as many data points as possible. Analogous to the margin in classification, the regression line is surrounded by a tube. Data points lying within that tube do not influence the course of the regression line. Data points lying on the edge or outside that tube are called support vectors.

In contrast to the support vectors in SVM, the relevance vectors in RVM represent the prototypical examples within the specified classification or regression task, instead of solely representing separating attributes. Furthermore, severe overfitting associated with the maximum likelihood estimation of the model parameters was avoided by imposing an explicit zero-mean Gaussian prior (Ghosh and Mujumdar 2008; Zheng et al. 2008). This prior is a characteristic feature of the RVM, and its use results in a vector of independent hyperparameters that reduces the data set (Faul and Tipping 2002; Tipping 2000; Tipping 2001; Tipping and Faul 2003). Therefore, in most cases the number of relevance vectors is much smaller than the number of support vectors. Furthermore, in SVR additional parameters have to be determined or statistically optimized (e.g. with cross-validation loops) in order to control for model complexity and model fit. To control the behavior of the RVR, only the type of kernel has to be chosen, whereas all other parameters are automatically estimated by the learning procedure itself. More details can be found in (Bishop 2006; Schölkopf and Smola 2002; Tipping 2000).

2.4 Brain age estimation framework

The brain age estimation framework utilizes RVR to capture the complex, multidimensional aging pattern across the whole brain and to subsequently estimate individual brain ages based on T1-weighted images. As suggested in Franke et al. (2010), the kernel was chosen to be a polynomial of degree 1, since age estimation accuracy was shown to not improve when choosing non-linear kernels. Thus, parameter optimization during the training procedure was not necessary.

In general, the model is trained with preprocessed whole brain structural MRI data (as described in 2.2) of the training sample, resulting in a complex model of healthy brain aging (Figure 2A, left panel). Put in other words, the algorithm uses those whole-brain MRI data from the training sample that represent the prototypical examples within the specified regression task (i.e., healthy brain aging). Additionally, voxel-specific weights are calculated that represent the importance of each voxel within the specified regression task (i.e., healthy brain aging). For an illustration of the most important features (i.e., the importance of voxel locations for regression with age) that were used by the RVR to model normal brain aging and more detailed information please refer to Franke et al. (2010).

Subsequently, the brain age of a test subject can be estimated using the individual tissue-classified MRI data, aggregating the complex, multidimensional aging pattern across the whole brain into one single value (Figure 2A, right panel). In other words, all the voxels of the test subject's MRI data are weighted by applying the voxel-specific weighting matrix. Then, the brain age is calculated by applying the regression pattern of healthy brain aging and aggregating all voxel-wise information across the whole brain. The difference between estimated brain age and the true chronological age will reveal an individual deviation score. Consequently, this deviation score directly quantifies the amount of acceleration or deceleration in individual brain aging (Figure 2B). For example, if a 70 years old individual has a deviation score of +5 years, this means that this individual shows the typical atrophy pattern of a 75 years old individual. To compute the final age regression model as well as to predict the

individual brain ages, we used the freely available toolbox "The Spider" (http://www.kyb.mpg.de/bs/people/spider/main.html).

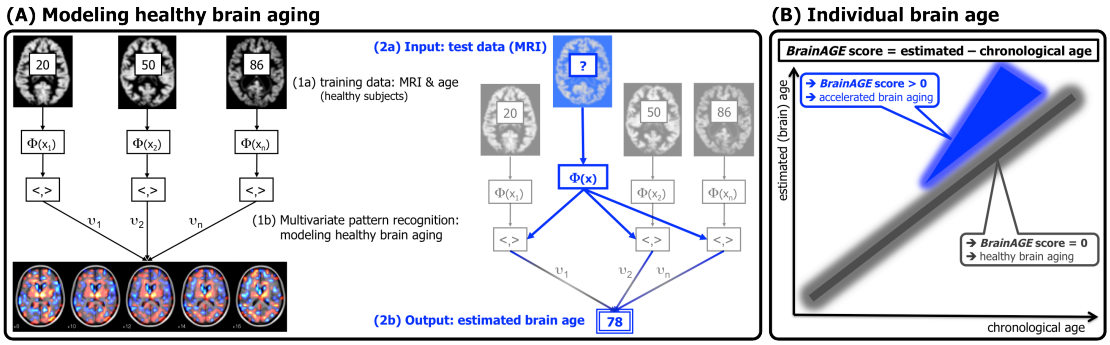

Figure 2. (A) The model of healthy brain aging is trained with the chronological age and preprocessed structural MRI data of a training sample (left; with an exemplary illustration of the most important voxel locations that were used by the age regression model). Subsequently, the individual brain ages of previously unseen test subjects are estimated, based on their MRI data (blue; picture modified from Schölkopf and Smola (2002)). **(B)** The difference between the estimated and chronological age results in the deviation (i.e., BrainAGE) score. Consequently, positive deviation scores indicate accelerated brain aging. (Image reproduced from Franke et al. (2012), with permission from Hogrefe Publishing, Bern).

The age estimation model was separately trained on male and female subjects in the IXI training sample. Furthermore it was trained using *(i)* preprocessed GM images, *(ii)* preprocessed WM images, and *(iii)* the linear combination of preprocessed GM and WM images. Subsequently, the brain ages of the subjects in the *CAD Dementia* training and test samples were estimated based on their non-uniformity corrected (nuc) MRI data. The difference between the estimated and the true age resulted in *(i)* GM, *(ii)* WM, and *(iii)* GMWM deviation scores, respectively. Sample-specific set-offs and slopes were corrected for the whole *CAD Dementia* sample using linear regression. The whole brain age estimation framework works automatically. All data preprocessing, model training, and brain age estimation was done using MATLAB.

2.5 Classification based on brain age estimation

Classification of the *CAD Dementia* subjects is based on the individual deviation scores. To explore the best classification accuracies in the *CAD Dementia* training sample, the brain age estimation model was run with all combinations of parameter variation during preprocessing (i.e. 4 mm & 8 mm FWHM smoothing kernel) and segmented brain tissue (i.e., GM, WM & GMWM). Subsequently, in each of those six brain age models two linear thresholds (i.e., NO vs. MCI and MCI vs. AD) were searched for to classify the subjects as NO, MCI, or AD based on individual deviation scores. Accuracy rates and receiver operating characteristics (ROC) for two-class classifications (i.e., NO vs. MCI, NO vs. AD, MCI vs. AD) were computed in the *CAD Dementia* training sample, resulting in the area under the ROC curve (AUC). Based on these results, the final brain age estimation models for classification were chosen. To differentiate between NO vs. MCI and MCI vs. AD the model using the linear combination of preprocessed GM and WM images with a 4 mm FWHM smoothing kernel showed best performances. To further refine the differentiation between NO vs. MCI we additionally used WM deviation scores (8 mm FWHM smoothing kernel). Then, overall classification accuracies were calculated following (Hand and Till 2001), as recommended in the *CAD Dementia challenge rules*. Subsequently, the final thresholds that reveal the highest overall classification accuracies in the *CAD Dementia* training sample were applied to classify the *CAD Dementia* test subjects.

3 Results

3.1 Performance of the age estimation model

To analyze the performance of the specific models used for training of the typical age patterns in healthy subjects, leave-one-out was used in the IXI training sample. For the GMWM age estimation

model (with 4 mm FWHM smoothing kernel) the correlation between estimated brain age and chronological age was r = 0.92 in the male [n= 250; mean (SD) age: 46.8 (16.4) years] as well as in the female subsample [n= 311; mean (SD) age: 48.6 (16.5) years]. The mean absolute error (MAE) was 5.1 years in both subsamples. For the WM age estimation model (with 8 mm FWHM smoothing kernel) the correlation between estimated brain age and chronological age was r = 0.88 in the male and r = 0.90 in the female subsample. The MAE in the male subsample was 6.4 years and 5.8 years in the female subsample.

3.2 Classification based on brain age estimation in the *CAD Dementia* training sample

Overall classification accuracies as calculated following (Hand and Till 2001), was 90% in the *CAD Dementia* training sample. Regarding the GMWM deviation score the thresholds are set at 2.6 to differentiate between NO vs. MCI and 4.5 to differentiate between MCI vs. AD. To further refine the differentiation between NO vs. MCI an additional threshold was set at -2.9 in the WM deviation score (Figure 3).

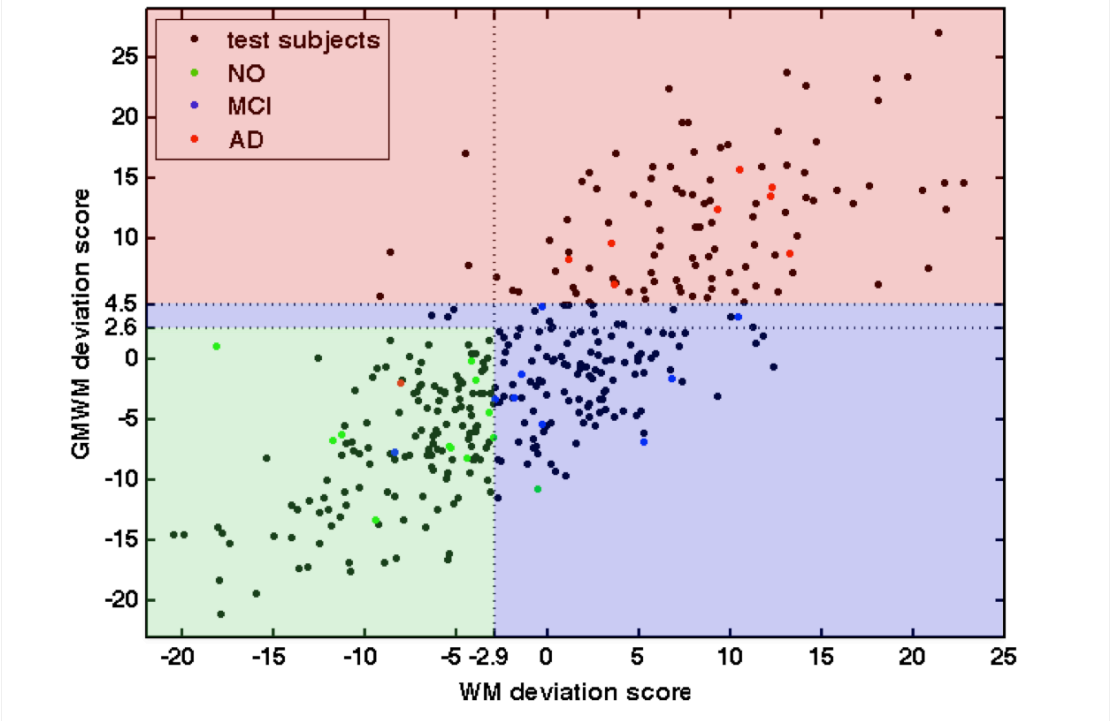

Figure 3: Classification scheme for the *CAD Dementia* training (colored dots) and test subjects (black dots) depending on the thresholds in WM and GMWM deviation scores. Subjects in the green area are classified as "NO", subjects in the blue area are classified as "MCI", and subjects in the red area are classified as "AD".

The whole approach is fully automatic. Preprocessing the MRI data (as described in 2.2) takes about 8 minutes per brain image. The whole process of training the brain age estimation models with about 560 subjects from the IXI sample, subsequent estimation of the brain ages for all 384 *CAD Dementia* subjects, and final classification of the *CAD Dementia* training subjects takes about 110 seconds in total on MAC OS X, Version 10.5.8, 2.4 GHz Intel Core 2 Duo.

4 Discussion

For estimating the brain age from T_1-weighted MRI scans, we propose a fast and fully automatic framework that includes preprocessing of the images, dimension reduction via PCA, training of a RVM for regression with a polynomial kernel of degree 1, and finally estimating the brain age of the *CAD Dementia* subjects. This age estimating framework turns out to be a straightforward method to accurately and reliably estimate brain age with as little preprocessing and parameter optimization as possible. Using MRI data of about 560 healthy subjects aged between 20 and 86 and scanned on three

different scanners, the age estimation with RVR showed excellent performance, with an overall MAE of only 5 years and a correlation of r = 0.92 between the chronological and the estimated brain age.

Using structural MRI data, our fully automated brain age estimation model aggregates the complex, multidimensional aging patterns across the whole brain to one single value (i.e. the deviation score) and finally identifies subtle pathological brain aging in the *CAD Dementia* training MCI and AD subjects, with increasing deviation scores at indicating an increased risk for AD.

In conclusion, our brain age estimation framework could potentially help to recognize and indicate advanced brain atrophy on an individual level, thus contributing to an early diagnosis of neurodegenerative diseases like AD and facilitate early treatment or a preventative intervention.

References

Ashburner J, Friston KJ (2005) Unified segmentation. NeuroImage 26 (3):839-851

Bartzokis G (2011) Alzheimer's disease as homeostatic responses to age-related myelin breakdown. Neurobiology of Aging 32 (8):1341-1371

Bennett KP, Campbell C (2003) Support vector machines: hype or hallelujah? SIGKDD Explorations 2:1-13

Bishop CM (2006) Pattern Recognition and Machine Learning. Springer, New York, NY

Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM (2007) Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 3 (3):186-191

Cao K, Chen-Plotkin AS, Plotkin JB, Wang LS (2010) Age-correlated gene expression in normal and neurodegenerative human brain tissues. PLoS One 5 (9):e13098

Cuadra MB, Cammoun L, Butz T, Cuisenaire O, Thiran JP (2005) Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. IEEE Transactions on Medical Imaging 24 (12):1548-1565

Driscoll I, Davatzikos C, An Y, Wu X, Shen D, Kraut M, Resnick SM (2009) Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. Neurology 72 (22):1906-1913

Dukart J, Schroeter ML, Mueller K (2011) Age correction in dementia--matching to a healthy brain. PLoS One 6 (7):e22193

Faul A, Tipping ME (2002) Analysis of sparse Bayesian learning. In: Dietterich TG, Becker S, Ghahramani Z (eds) Advances in Neural Information Processing Systems 14. MIT Press, pp 383–389

Franke K, Gaser C, for the Alzheimer's Disease Neuroimaging Initiative (2012) Longitudinal changes in individual *BrainAGE* in healthy aging, mild cognitive impairment, and Alzheimer’s disease. GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry 25 (4):235 - 245

Franke K, Ziegler G, Klöppel S, Gaser C, the Alzheimer's Disease Neuroimaging Initiative (2010) Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. NeuroImage 50 (3):883-892

Frisoni GB, Fox NC, Jack CR, Jr., Scheltens P, Thompson PM (2010) The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 6 (2):67-77

Gaser C (2009) Partial volume segmentation with Adaptive Maximum a Posteriori (MAP) approach. NeuroImage 47 (S1):S121

Ghosh S, Mujumdar PP (2008) Statistical downscaling of GCM simulations to streamflow using relevance vector machine. Advances in Water Resources 31 (1):132-146

Hand D, Till R (2001) A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning 45 (2):171-186

Jack CR, Jr., Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 9 (1):119-128

Jones DT, Machulda MM, Vemuri P, McDade EM, Zeng G, Senjem ML, Gunter JL, Przybelski SA, Avula RT, Knopman DS, Boeve BF, Petersen RC, Jack CR, Jr. (2011) Age-related changes in the default mode network are more advanced in Alzheimer disease. Neurology 77 (16):1524-1531

Rajapakse JC, Giedd JN, Rapoport JL (1997) Statistical approach to segmentation of single-channel cerebral MR images. IEEE Transactions on Medical Imaging 16 (2):176-186

Saetre P, Jazin E, Emilsson L (2011) Age-related changes in gene expression are accelerated in Alzheimer's disease. Synapse 65 (9):971-974

Schölkopf B, Smola A (2002) Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA

Spulber G, Niskanen E, MacDonald S, Smilovici O, Chen K, Reiman EM, Jauhiainen AM, Hallikainen M, Tervo S, Wahlund LO, Vanninen R, Kivipelto M, Soininen H (2010) Whole brain atrophy rate predicts progression from MCI to Alzheimer's disease. Neurobiology of Aging 31 (9):1601-1605

Tipping ME (2000) The Relevance Vector Machine. In: Solla SA, Leen TK, Müller K-R (eds) Advances in Neural Information Processing Systems 12. MIT Press, Cambridge, MA, pp 652-658

Tipping ME (2001) Sparse bayesian learning and the relevance vector machine. Journal of Machine Learning Research 1 (3):211-244

Tipping ME, Faul AC Fast marginal likelihood maximisation for sparse Bayesian models. In: Bishop CM, Frey BJ (eds) Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL, Jan 3-6 2003.

Tohka J, Zijdenbos A, Evans A (2004) Fast and robust parameter estimation for statistical partial volume models in brain MRI. NeuroImage 23 (1):84-97

Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, Knopman DS, Petersen RC, Jack CR, Jr. (2009a) MRI and CSF biomarkers in normal, MCI, and AD subjects: diagnostic discrimination and cognitive correlations. Neurology 73 (4):287-293

Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, Knopman DS, Petersen RC, Jack CR, Jr. (2009b) MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. Neurology 73 (4):294-301

Zheng Y-T, Neo S-Y, Chua T-S, Tian Q (2008) Probabilistic optimized ranking for multimedia semantic concept detection via RVM. Paper presented at the Proceedings of the 2008 international conference on Content-based image and video retrieval, Niagara Falls, Canada,