Alzheimer's disease state classification using structural volumetry, cortical thickness and intensity features

Christian Ledig[1], Ricardo Guerrero[1], Tong Tong[1], Katherine Gray[1], Alexander Schmidt-Richberg[1], Antonios Makropoulos[1,2], Rolf A. Heckemann[3,4], Daniel Rueckert[1]

[1]Department of Computing, Imperial College London, UK
[2]Division of Imaging Sciences and Biomedical Engineering, King's College London, UK
[3]MedTech West, Institute of Neuroscience and Physiology, University of Gothenburg, Sweden
[4]Division of Brain Sciences, Faculty of Medicine, Imperial College London, UK

Abstract. Evaluating algorithms for the image-based classification of dementia on the basis of common data and using common metrics is essential for an objective comparison of different approaches. It is the aim of the *Computer-Aided Diagnosis of Dementia based on structural MRI data* (CADDementia) challenge to address this need by providing the opportunity of objectively evaluating individual approaches. In this paper, a classification framework is presented, in which four different sets of features extracted from structural MR images are examined with respect to their discriminative abilities. These features are based on volumetric and morphologic parameters, image intensities and patch similarities. Moreover, a combined feature set is employed to analyse the amount of complementary information contained in the features. For the three-class classification problem – Alzheimer's Disease (AD) vs. Mild Cognitive Impairment (MCI) vs. Control (CN) – a classification rate on a subset of the ADNI1-2 databases between 51% and 59% is achieved with all five feature sets. The combined feature set leveraging the potential of all four methods leads to only a minor improvement over the individual sets.

1 Introduction

The *Computer-Aided Diagnosis of Dementia based on structural MRI data* (CADDementia) challenge [1] is aimed at evaluating different methods for the image-based diagnosis of dementia. Held in the course of the MICCAI 2014 conference, the challenge provides a standardised platform for objectively testing different approaches on a common set of image data, which allows an objective comparison.
A crucial aspect of the computer-aided diagnosis based on MRI data is the extraction of features from the images. These features have to be meaningful for the diagnostic purpose, that means they are designed to contain valuable information about the state and the progression of the disease. In the case of Alzheimer's disease, for example, features proposed in the literature range from concrete and well-known clinical biomarkers like the hippocampal volume [2] or the cortical thickness [3] to abstract parameters derived directly from the image intensities by manifold learning approaches [4].
In this contribution, we aim at comparing different feature extraction methods that have

all been shown to be of high potential for the differential diagnosis of Alzheimer's disease and related forms of dementia. These features comprise:

- **Volumetric features (VOL):** A total of 134 distinct brain volumes are automatically segmented (see Sec. 3.1).
- **Morphologic features (CORT):** Based on the brain segmentations, morphological features such as cortical thickness and cortical surface measurements are computed (see Sec. 3.2).
- **Manifold-based learning features (MBL):** Manifold-based learning is used to map intensity texture descriptors of all subjects into a d-dimensional space, such that similarities between the images are maintained. The d coordinates of this space (or a subset of them) are then considered as features (see Sec. 3.3).
- **Patch-based grading features (GRAD):** A patch-based approach is employed to find similar intensity patterns in scans of other subjects. Features are then computed as the weighted average of the labels associated with similar patches (see Sec. 3.4).

The special focus of this work is on A) comparing the performance of the features on a set of unseen image data provided by the CADDementia challenge, and B) assessing the degree of complementary information contained in the feature sets. Therefore, as well as analysing the four groups of features independently, a joint feature set (ALL) is tested that combines all available features.
For classification, Random Forests were employed consistently in all experiments as they have been shown to be powerful in particular for multi-class classification [5] (see Sec. 4). They were trained on the Alzheimer's Disease Neuroimaging Initiative (ADNI) [6] database using the five different feature sets and the classification results were submitted to the CADDementia challenge.
In Sec. 5, classification results for a 10-fold cross validation on the ADNI database and for the classification of the 30 training cases hosted by the CADDementia challenge are additionally provided. These experiments showed that a high performance of all feature sets for the classification of the Alzheimer's disease state, with three-class classification accuracies ranging between 51% and 59% for the cross validation. The intensity based features MBL and GRAD slightly outperformed the other features. Only a small improvement was reached with the combined feature set, which suggests that the amount of complementary information contained in the features is limited.

2 CADDementia data and common preprocessing

The challenge data is comprised of 30 plus 354 T1-weighted images for training and testing data respectively. The images were acquired at three different sites at 3T magnetic field strength. More details can be found at [1]. Both training and testing data were preprocessed using the same pipeline. All T1-weighted MR scans with in-place resolution below 0.5 mm were resampled and their in-plane resolution doubled. The images were further corrected for potential intensity inhomogeneities employing the N4 algorithm [7]. Brain masks were then calculated for both training and testing images using pincram [8]. As atlas database for pincram, 64 subjects of the ADNI1 cohort were

chosen. After visual inspection $\approx 10\%$ of the brain masks for the CADDementia testing dataset did not meet our quality criterion. An updated pincram version was rerun on these subjects adding 3T scans from the ADNI2 study as well as successfully extracted subjects from the CADDementia training set to the atlas database. The extended database lead to an improved segmentation quality sufficient for a further analysis, such that no manual editing was required. It was observed that scans acquired at EMC Rotterdam were particularly challenging to extract.

3 Feature Extraction

In the following we present four approaches to extract biomarkers that have been shown to have potential for Alzheimer's disease state classification. The focus is on providing a brief description of the individual approaches and how they are applied. For further details we refer to the papers describing the whole brain segmentation approach (VOL, [9]), the cortical thickness measurement (CORT, [10]), the extraction of intensity features (MBL, [11]) and the calculation of patch-based grading values (GRAD, [12]).

3.1 Volumetric features from multi-structure whole brain segmentation (VOL)

Training data To automatically parcellate the provided MR scans into anatomical regions, the 30 brain atlases (excluding repeat scans) provided through the "MICCAI Grand Challenge on Multi-Atlas Labeling 2012" [13] were employed. This atlas database consists of 30 T1-weighted MR scans from the OASIS database that were annotated by expert raters[1] into 134 distinct brain regions.

Method description We employed the multi-atlas label propagation method described in [9]. In this approach, all 30 brain atlases are aligned to an unsegmented subject MRI using a robust nonrigid registration approach based on multi-level free form deformations [14–16]. The individual atlas label maps are then transformed to the image space of the unsegmented image using the calculated transformation and a nearest neighbour interpolation scheme. The 30 propagated atlas label maps are fused into a consensus probabilistic segmentation using a local weighting approach. The obtained probabilistic segmentation is further refined using a method that exploits image intensities in an expectation maximisation framework [9]. For each subject we finally extract 135 volumetric features, including background, based on the segmentations.

3.2 Cortical morphology features (CORT)

Training data Measurements of cortical morphology were obtained based on the segmented regions of the cortex. Cortical surface area, curvature and thickness features were calculated for the whole cortex and its 98 regional subdivisions (in total 591 features).

[1] provided by Neuromorphometrics, Inc. (http://Neuromorphometrics.com/) under academic subscription.

Cortical thickness estimation Cortical thickness was estimated as described in [10]. In this approach, a potential field from the GM-WM to the GM-CSF interface is determined by solving the Laplace's equation. Voxel-wise thickness was then calculated as the sum of the voxel's distance to the WM and to the CSF, following the direction perpendicular to the potential field. Cortical sulci correction was performed similarly to [17]. We calculated cortical thickness (mean and standard deviation) for each cortical region (98+98 features).

Cortical surface measurements Cortical surface meshes were obtained by triangulation of the CGM-WM isosurface of each hemisphere with the marching cubes algorithm [18]. The surfaces were smoothed with Laplacian smoothing [19] for an even distribution of the mesh vertices. An example cortical surface is presented in Figure 1.

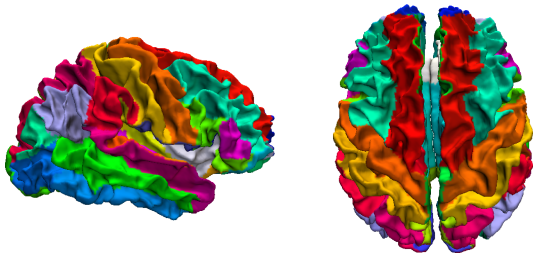

Fig. 1: Cortical surface of the CGM-WM interface with overlaid segmentation.

Surface area and curvature measures of the cortex were computed from the meshes. We adopted a number of area-independent curvature measures from [20] with T-normalisation that are not sensitive to the surface area. The surface measures included in this study were: surface area in the whole cortex and each region (1+98), relative surface area (98), mean curvature L^2 norm (1+98) and Gaussian curvature L^2 norm (1+98).

3.3 Manifold-based learning for multi-level variable selection (MBL)

Training data Data used was obtained from the ADNI database. In this work, a subset of 292 ADNI-1 subjects with baseline 1.5T MR images and that did not have 1.5T MR images available at 12 or 24 month follow-up, were used for training a multilevel variable selection scheme based on sparsity. The remaining 1.5T and 3T ADNI-1, -GO and -2 baseline images (as of November 2013) were used to evaluate and tune the proposed framework. In total 1701 images were used, from which 292 were used for multilevel variable selection and 1409 for evaluation and tuning.

Multilevel relevant variable selection The goal of variable selection is to reduce the amount of input variables to those that are relevant for a specific task. In this work a sparse regression was used in a similar fashion as the method described in [11] to select

relevant variables. Here, independent variables are the MR image intensities, while the mini mental state examination (MMSE) score acts as dependent variable [11]. Furthermore, as disease specific imaging characteristics might manifest at different alignment levels, each of the N subjects has associated R images that have been created by aligning the original scan to the MNI152 template at R different levels using [15]. A matrix $\mathbf{X}$ is built, where each column represents a location in MNI space at a certain alignment level. Each row in matrix $\mathbf{X}$ represents a subject $n \in N$ via concatenating its vectorised MR images at multiple levels $r \in R$. The algorithm then selects a subset of D variables from $\mathbf{X}$, that correspond to column indices of $\mathbf{X}$. This yields a 4D mask, where the first three dimensions are coordinates in MNI space and the fourth is the alignment level of the image to the template.

Local binary patterns MR images acquired using different acquisition protocols have different intensity appearance and thus cannot be easily combined into a single framework. In this work we extract local binary patterns (LBP) [21] around the 26-connected neighbourhood of each selected voxel and encode them as binary vectors. This transforms MR intensity features to an augmented binary space where the images lie in the same space and thus can be combined, assuming that the original acquisitions protocols are reasonably similar (e.g. both are T1-weighted).

Dimensionality reduction The aim of this work is to produce a three-class classifier. For this purpose, it was found in [11] that the learned 4D mask is still relatively high-dimensional and that reducing the dimensionality generally improves classification results. In this work principal component analysis (PCA) [22] was used to reduce the dimensionality of the data.

3.4 Patch-based disease grading features (GRAD)

Training data ADNI baseline scans of 629 subjects (233 CN, 231 MCI, 165 AD) acquired at 3T and 30 training images from the challenge were utilised as training dataset. For each image, 150 grading features were calculated for classification. Then, an optimised classifier was trained on the training dataset and used to predict the class labels of the testing images from the challenge.

Method description After preprocessing, non-rigid registration, based on B-spline free-form deformation [15] with a final control point spacing of 5mm, was performed to align all images to the MNI152 template space. The intensity was normalised between each image and the template using the approach proposed in [23]. Then, the sparse regression method proposed in [24] was used to calculate a probability map for selecting patches. Finally, 150 patches with the highest probabilities are extracted from each image for calculating grading features.
A patch-based approach was proposed in [25] to calculate grading features for classification of AD. Our method is an extension of this method by introducing sparse representation techniques. To calculate a grading value for each target patch p_t, corresponding patches in the training images are extracted to form a training patch library

P_L. The patch library P_L typically contains thousands of training patches. The relationship between the target patch p_t and the patches in the training library P_L is modelled by a weighting function in [25]. In our work, the Elastic Net sparse regression model [12] is used to model the relationship between the target patch and the training patches. Since the pathological status of the training patches are known, we can propagate the pathological status to the target patch by using the patch relationship. The grading value of the target patch p_t can then be estimated as:

$$\begin{cases} \hat{a}_t = \min\limits_{a_t} \frac{1}{2} \|p_t - P_L a_t\|_2^2 + \lambda_1 \|a_t\|_1 + \frac{\lambda_2}{2} \|a_t\|_2^2 \\ g_t = \frac{\sum_{j=1}^{N} \hat{a}_t(j) s_j}{\sum_{j=1}^{N} \hat{a}_t(j)} \end{cases} \tag{1}$$

where $\hat{a}_t$ are the coding coefficients for the target patch p_t. Most of the coefficients in $\hat{a}_t$ are zero due to the sparsity constraint. If the coefficient in $\hat{a}_t$ is not zero, it indicates that the corresponding training patch has been selected to propagate their pathological information to the target patch. N is the number of the training patches in the patch library P_L. The CN and AD training groups were selected from the training dataset to propagate pathological status to the target image as suggested in [25]. The pathological status of the training patch $P_L(j)$ is denoted as s_j. If the training patch is extracted from CN subjects, s_j is set to 1. $s_j = -1$ is used for patches extracted from AD subjects. Finally, a grading value g_t can be calculated for each target patch p_t. Since 150 patches are extracted from each image, 150 grading features are calculated for classification.

4 Classification based on Random Forests

For all classification experiments, we applied the random forests algorithm [26] implemented in scitkit-learn ([27]; http://scikit-learn.org/). This method is able to predict both binary labels and class probabilities, and can be directly applied for three-class classification. As in the originally proposed algorithm, no maximum tree depth was specified, and a bootstrap sample of the training data was passed to each tree. Tree nodes were split based on an entropy criterion, and 100 trees were included in each forest. The number of features randomly sampled at each tree node was set to $\sqrt{n}$ for all experiments, following the recommendation of Liaw and Wiener [28].

5 Results

5.1 Training data

Features generated by different methods seek to model differences in subgroups in a distinct way and thus may hold complementary information between each other. In order to asses this, a subset of subjects that overlap for all of the proposed methods was used for training. Features generated by those methods were combined into a single classification framework, where if complementary information exists, overall classification results would be expected to improve. In total, 734 subjects from the ADNI1-2 datasets where included in this analysis. See Table 1 for the demographics of the data.

Table 1: Subject groups mean age, sample size, MMSE scores, gender, CDR scores and Magnetic strength from ADNI1-2.

	N	Age	MMSE	Men (#)	CDR	1.5T (3T)
AD	170	74.77±7.62	23.12±2.06	46% (78)	0.79±0.27	110 (80)
MCI	288	73.79±7.47	27.28±1.79	55% (158)	0.50±0.00	185 (103)
CN	276	74.75±5.82	29.07±1.16	47% (131)	0.00±0.00	156 (120)

We performed classification experiments in two variations. Firstly, we performed a 10-fold cross validation approach within this subset of the ADNI database. Secondly, we trained the random forest classifier using this ADNI subset, and classified the provided CADDementia training set. The second experiment is the same setup that was used for calculating the submitted results based on the CADDementia testing data.
For each classification task we employed a random forest classifier that was trained on five subsets of the available features. We used features provided by each of the four individual methods (VOL, CORT, MBL, GRAD) as presented in 3.1-3.4. We furthermore combined all available features (ALL=$\{$VOL, CORT, MBL, GRAD$\}$) provided by the presented methods to investigate whether they provide complementary information. We did not use available meta information such as age, field strength and gender.
The mean accuracies obtained in the ADNI cross-validation experiment are provided in Table 2. The mean accuracies obtained by 10 classification runs for the CADDementia training data are summarised in Table 3.
When using ALL features, we further investigated which features were most informative by extracting feature importances from the random forest classifier trained for the three-class classification task. We found that the first manifold coordinate of the MBL features was most important. Otherwise the top 50 features were, except left and right Amygdala volume (#18, #24), constituted by exclusively grading features provided by GRAD.

Table 2: Overview of the classification results for the 10-fold cross validation on the subset of the ADNI1-2 cohort. Mean classification accuracy (± SD) based on 10-fold cross validation.

Type	# Feat.	AD vs. CN	AD vs. MCI	MCI vs. HC	AD vs. MCI vs. HC
VOL	135	0.83±0.05	0.68±0.04	0.67±0.05	0.54±0.04
CORT	591	0.80±0.05	0.65±0.06	0.63±0.04	0.51±0.05
MBL	20	0.89±0.05	0.67±0.07	0.70±0.05	0.58±0.03
GRAD	150	0.86±0.04	0.67±0.04	0.69±0.04	0.56±0.04
ALL	896	0.87±0.03	0.68±0.04	0.72±0.05	0.59±0.04

Table 3: Overview of the classification results obtained on CADDementia training data. Mean classification accuracy (± SD) based on 10 classification runs.

Type	# Feat.	AD vs. CN	AD vs. MCI	MCI vs. HC	AD vs. MCI vs. HC
VOL	135	0.86±0.03	0.73±0.05	0.68±0.05	0.56±0.08
CORT	591	0.91±0.05	0.67±0.09	0.65±0.05	0.58±0.07
MBL	20	0.94±0.02	0.62±0.04	0.75±0.04	0.66±0.01
GRAD	150	0.88±0.03	0.75±0.06	0.76±0.03	0.67±0.05
ALL	896	0.92±0.02	0.78±0.05	0.75±0.04	0.68±0.05

5.2 Computation times

The approximate computation times per subject are summarised in Table 4. None of the presented methods requires manual interaction.

Table 4: Overview of the approximate computation times per subject.

Task	Runtime	Implementation	Automatic
N4 bias correction	$<$ 30 minutes	single core	yes
pincram brain extraction	$<$ 1 hour	parallel	yes*
registration of the 30 atlases (VOL)	$<$ 2 hours	parallel	yes
atlas fusion (VOL)	$<$ 20 minutes	single core	yes
cortical thickness (CORT)	$<$ 15 minutes	single core	yes
variable selection, 292 images (MBL)	$<$ 2.5 hours	single core	yes
local binary patterns (MBL)	$<$ 1 second	single core	yes
dimensionality reduction, $\sim$1800 subjects (MBL)	$<$ 10 seconds	parallel	yes
Grading feature extraction (GRAD)	$<$ 5 minutes	single core	yes
classification	$<$ 1 second	single core	yes

*(manual quality control)

6 Discussion

We have presented four independent approaches for extracting both structural and intensity based biomarkers from MR images with the goal of Alzheimer's disease state classification. Cross-validation experiments on the ADNI database suggest that the volumetric features (VOL) and in particular the intensity (MBL) and grading features (GRAD) are competitive to published state-of-the-art classification results. We have found that combining these features by training a single random forest for all features jointly seems to have no great impact on the classification performance. This finding suggests that the presented biomarkers provide little complementary information, which was unexpected but indicates the difficulty of an accurate three-way classification on

the ADNI database. Classification results on the training data provided through the CADDementia challenge were substantially higher than those obtained on the ADNI database. While this finding is, due to the small sample size of the training dataset, not definitive it is highly interesting as it suggests that the subjects in the challenge data show a clearer group separation than the ADNI cohort.

A Author's contributions

CL preprocessed and brain extracted all available scans. CL visually inspected all brain masks and calculated the multi-class whole brain segmentations. AM and CL calculated the cortical morphology features based on these segmentations. RG performed the multi-level variable selection using intensity features. TT calculated the gradient based grading features. KG combined all features and performed the final classification experiments. ASR supported the work sharing experience on this type of challenge and image registration in particular. RAH developed and assisted in the application of pincram. DR continuously supervised the project and enabled this work by providing the necessary resources. The individuals drafted the corresponding paragraphs of the manuscript.

References

1. "Computer-Aided Diagnosis of Dementia based on structural MRI data (CADDementia)." http://caddementia.grand-challenge.org/.
2. R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, D. Rueckert, and Alzheimer's Disease Neuroimaging Initiative, "LEAP: learning embeddings for atlas propagation," *NeuroImage*, vol. 49, pp. 1316–1325, Jan. 2010.
3. R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, O. Colliot, and Alzheimer's Disease Neuroimaging Initiative, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, pp. 766–781, May 2011.
4. S. Gerber, T. Tasdizen, P. Thomas Fletcher, S. Joshi, R. Whitaker, and Alzheimers Disease Neuroimaging Initiative (ADNI), "Manifold modeling for brain population analysis.," *Med Image Anal*, vol. 14, pp. 643–653, Oct. 2010.
5. K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, and Alzheimer's Disease Neuroimaging Initiative, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, Jan. 2013.
6. S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "The Alzheimer's Disease Neuroimaging Initiative," *Neuroimaging Clinics of North America*, vol. 15, no. 4, pp. 869–877, 2005. Alzheimer's disease: 100 years of progress.
7. N. Tustison, B. Avants, P. Cook, Y. Zheng, A. Egan, P. Yushkevich, and J. Gee, "N4ITK: Improved N3 bias correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.
8. R. A. Heckemann, C. Ledig, P. Aljabar, K. R. Gray, D. Rueckert, J. V. Hajnal, and A. Hammers, "Label propagation using group agreement DISPATCH," pp. 75–78, MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling, 2012.
9. C. Ledig, R. Wolz, P. Aljabar, J. Lötjönen, R. A. Heckemann, A. Hammers, and D. Rueckert, "Multi-class brain segmentation using atlas propagation and EM-based refinement," *Proceedings of ISBI 2012*, pp. 896–899, 2012.

10. S. E. Jones, B. R. Buchbinder, and I. Aharon, "Three-dimensional mapping of cortical thickness using laplace's equation," *Human Brain Mapping*, vol. 11, no. 1, pp. 12–32, 2000.
11. R. Guerrero, R. Wolz, A. W. Rao, and D. Rueckert, "Manifold population modeling as a neuro-imaging biomarker: Application to ADNI and ADNI-GO.," *NeuroImage*, vol. 94C, pp. 275–286, 2014.
12. T. Tong, R. Wolz, P. Coupé, J. V. Hajnal, and D. Rueckert, "Segmentation of mr images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling," *NeuroImage*, vol. 76, pp. 11–23, 2013.
13. B. Landman and S. K. Warfield, "MICCAI 2012 Workshop on Multi-Atlas Labeling. https://masi.vuse.vanderbilt.edu/workshop2012/images/c/c8/miccai_2012_workshop_v2.pdf," 2012.
14. R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal, and A. Hammers, "Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation," *NeuroImage*, vol. 51, no. 1, pp. 221–227, 2010.
15. D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, and D. J. Leach, M. O.and Hawkes, "Non-rigid registration using free-form deformations: Application to breast MR images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.
16. M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units," *Computer Methods and Programs in Biomedicine*, vol. 98, no. 3, pp. 278–284, 2010.
17. X. Han, D. L. Pham, D. Tosun, M. E. Rettmann, C. Xu, and J. L. Prince, "CRUISE: cortical reconstruction using implicit surface evolution," *NeuroImage*, vol. 23, pp. 997–1012, Nov. 2004.
18. W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '87, (New York, NY, USA), pp. 163–169, ACM, 1987.
19. L. R. Herrmann, "Laplacian-isoparametric grid generation scheme," *Journal of the Engineering Mechanics Division*, vol. 102, pp. 749–907, Oct. 1976.
20. C. E. Rodriguez-Carranza, P. Mukherjee, D. Vigneron, J. Barkovich, and C. Studholme, "A framework for in vivo quantification of regional brain folding in premature neonates," *NeuroImage*, vol. 41, pp. 462–478, June 2008.
21. T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
22. H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417–441, 1933.
23. L. G. Nyu and J. K. Udupa, "On standardizing the MR image intensity scale," *Magnetic Resonance in Medicine*, vol. 42, no. 6, p. 1072, 1999.
24. E. Janousova, M. Vounou, R. Wolz, K. Gray, D. Rueckert, and G. Montana, "Biomarker discovery for sparse classification of brain images in Alzheimer's disease," *Annals of the British Machine Vision Association (BMVA)*, vol. 2012, no. 2, pp. 1–11, 2012.
25. P. Coupé, S. F. Eskildsen, J. V. Manjón, V. S. Fonov, J. C. Pruessner, M. Allard, and D. L. Collins, "Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease," *NeuroImage: Clinical*, vol. 1, no. 1, pp. 141–152, 2012.
26. L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
27. F. Pedregosa, G. Varoquaux, A. Gramford, *et al.*, "Scikit-learn: Machine Learning in Python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
28. A. Liaw and M. Wiener, "Classification and regression by randomForest.," *R news*, vol. 2, pp. 18–22, 2002.