

Landmark-Based Alzheimer’s Disease Diagnosis Using Longitudinal Structural MR Images

Jun Zhang¹, Mingxia Liu¹, Le An¹, Yaozong Gao^{1,2}, and Dinggang Shen^{1,3}(✉)

¹ Department of Radiology and BRIC, UNC at Chapel Hill, Chapel Hill, NC, USA
dgshen@med.unc.edu

² Department of Computer Science, UNC at Chapel Hill, Chapel Hill, NC, USA

³ Department of Brain and Cognitive Engineering,
Korea University, Seoul, Republic of Korea

Abstract. In this paper, we propose a landmark-based feature extraction method for AD diagnosis using longitudinal structural MR images, which requires no nonlinear registration or tissue segmentation in the application stage and is robust to the inconsistency among longitudinal scans. Specifically, (1) the discriminative landmarks are first automatically discovered from the whole brain, which can be efficiently localized using a fast landmark detection method for the testing images; (2) High-level statistical spatial features and contextual longitudinal features are then extracted based on those detected landmarks. Using the spatial and longitudinal features, a linear support vector machine (SVM) is adopted for distinguishing AD subjects from healthy controls (HCs) and also mild cognitive impairment (MCI) subjects from HCs, respectively. Experimental results demonstrate the competitive classification accuracies, as well as a promising computational efficiency.

1 Introduction

Structural MRI has been proven to be an effective tool for Alzheimer’s disease (AD) diagnosis [1]. Compared with cross-sectional study at a single time point, longitudinal study is more sensitive to early pathological changes by focusing on both the spatial structural abnormalities and the longitudinal variations of tissues.

So far, researches that focus on cross-sectional study have obtained several achievements on AD or mild cognitive impairment (MCI) diagnosis [2, 3]. For example, Liu *et al.* investigated the AD diagnosis using multi-template representation [4–6]. Hinrichs *et al.* proposed to use spatially augmented LPboosting for AD classification [7]. Zhu *et al.* focused on selecting informative features from redundant region-based features [8–10]. Gerardin *et al.* extracted features based on hippocampal shape for the purpose of classifying AD and MCI [11]. Gao *et al.* proposed to use hypergraph learning for MCI classification and indexing [12, 13]. Kloppe *et al.* proposed to use voxel-based gray matter features for AD classification [14].

On the other hand, existing longitudinal studies largely focus on the degeneration of well-known representative biomarkers including hippocampal volume,

ventricular volume, whole brain volume and cortical thickness. For example, Chincarini *et al.* proposed four image analysis strategies based on hippocampal volume by integrating longitudinal atrophy rate as a measurement for AD diagnosis [15]. Jack *et al.* investigated the changing rates of four structures (*i.e.*, hippocampus, entorhinal cortex, whole brain and ventricle), and supported the idea of using changing rates as biomarkers for AD diagnosis [16]. Aguilar *et al.* analyzed the longitudinal atrophy changes in cortical thickness measures and subcortical volumes, and pointed out that the use of two time points data yielded better index result compared with using the cross-sectional data only [17]. Kim *et al.* adopted 93 ROI features for longitudinal analysis [18]. However, there are still several challenges in existing longitudinal analysis: (1) Limited measurements may be incapable of capturing the full pattern of morphological abnormalities from the whole brain; (2) Time-consuming nonlinear registration or tissue segmentation step is required, and the longitudinal study exacerbates the computational time since more scans are involved; (3) Longitudinal scans across subjects are usually inconsistent, since some time points might be missing during the data collection.

In this study, a landmark-based feature extraction framework is proposed for AD diagnosis using longitudinal structural MR images. Different from traditional longitudinal studies, our method (1) does not require the time-consuming nonlinear registration or tissue segmentation, (2) can cover the representative morphological abnormalities from the whole brain, and (3) is able to handle the inconsistency among longitudinal scans. Specifically, the discriminative landmarks which have significant morphological group differences are automatically discovered from the whole brain. By using a regression forest-based landmark detection method, these landmarks can be efficiently detected in the application stage. Based on these detected landmarks, high-level spatial features and contextual longitudinal features are further extracted respectively, as below. (a) A bag-of-words strategy is used to extract high-level spatial features, by calculating the frequency of low-level landmark-based morphological features from different scanning time points. In this way, the significant spatial abnormalities from all scanning time points are aggregated together, which are also invariant to the number of longitudinal scans. (b) To extract contextual longitudinal features, an interpolation step is used to generate a Jacobian map from longitudinal landmark displacements. Then, contextual features can be extracted around the landmarks from the Jacobian map. Finally, a linear support vector machine (SVM) is adopted to perform AD/MCI classification using these spatial and longitudinal features.

2 Materials and Image Processing

2.1 Dataset

The Alzheimer’s Disease Neuroimaging Initiative (ADNI)¹ is a 5-year public-private partnership to test whether serial MRI, positron emission tomography

¹ www.adni-info.org.

Table 1. Demographic information of selected subjects in the ADNI database

	Male/Female	Age (years) (Mean \pm SD)	Edu. (years) (Mean \pm SD)
AD	81/73	75.10 \pm 7.50	14.82 \pm 3.08
MCI	219/127	74.33 \pm 9.91	15.53 \pm 3.29
HC	111/96	75.83 \pm 4.98	16.10 \pm 2.86

Table 2. Number of scans for the selected subjects in the ADNI database

	3 scans	4 scans	5 scans	6 scans
AD	63	91	-	-
MCI	57	97	170	22
HC	46	145	16	-

(PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer’s disease. One goal of ADNI is to develop improved methods that will lead to uniform standards for acquiring longitudinal, multi-site MRI and PET data on subjects with AD, MCI, and elderly healthy controls (HCs).

Subjects used in this study are from the ADNI-1 database. In this paper, we selected the subjects with at least three scanning time points of structural MRI, thus resulting in 207 age-matched HCs, 154 AD, and 346 MCI subjects. The demographic information (*i.e.*, gender, age, and education) of the studied subjects used in this study are summarized in Table 1. The statistics of scans for the studied subjects is summarized in Table 2.

2.2 Image Processing

The image processing includes two major steps: linear alignment and landmark discovery.

Linear Alignment: All images are linearly aligned to a common template, namely Colin27, which was created by averaging 27 registered scans of a single subject [19]. In order to achieve high efficiency, we adopt a landmark-based affine registration method. Specifically, five pre-defined landmarks (*i.e.*, anterior commissure (AC) and posterior commissure (PC) landmarks, and the other three representative landmarks in mid-sagittal plane) are automatically detected by a pre-trained regression forest-based landmark detection model. A *global similarity transformation* matrix, which encodes 7 degree of freedom (DOF), can be estimated between the landmarks from the moving image to the template. Since each landmark has 3 coordinate values, 5 landmarks are enough to estimate the transformation matrix.

Landmark Discovery: Our target is to identify the regions with group differences in local structures between patients and HCs. To this end, we intend to

perform a voxel-wise group comparison between those two groups. However, the linearly aligned images are not voxel-wisely comparable. In order to build the correspondence among voxels from different images, all images are nonlinearly aligned to the Colin27 template after linear alignment. In general, the warped images are very similar to each other so that the subject-specific structural information in different images may not be significant. Therefore, we extract patch-based morphological features (*i.e.*, 3D histogram of orientation (HOG) features [20]) from the linearly aligned images to describe the local structures. By using the deformation field from nonlinear registration, we can build the correspondence between voxels in the template and all linearly aligned images. Therefore, for each voxel in the template, we can extract two groups of HOG features from its corresponding voxels in all training patients and HCs, respectively. We then perform the multivariate test, namely Hotelling’s T^2 statistic [21], on the two groups, through which a p -value can be calculated for each voxel in the template. Accordingly, a p -value map can be obtained according to the template. Finally, the local minima from the p -value map are identified as locations of discriminative landmarks in the template space. More details on landmark discovery can be found in our previous work [22].

Then, these landmarks, which are located in the template space, can be directly projected to all training images using their deformation fields. For testing images, in order to avoid the time-consuming nonlinear registration, we train a regression forest-based landmark detector [23, 24] to detect these landmarks. In this way, both training images and testing images would have same landmarks, and particularly, the landmarks for the testing images can be obtained efficiently, thanks to the fast landmark detector.

3 Feature Extraction

Based on the identified landmarks, we propose a landmark-based framework for extracting features from longitudinal MR images. Specifically, two types of landmark-based features, *i.e.*, spatial features and longitudinal features, are extracted to describe the spatial structural abnormalities and longitudinal landmark variations, respectively. In the following, we explain the details about the extraction process for each feature type.

3.1 Landmark-Based Spatial Feature Extraction

Intuitively, in cross-sectional study, the morphological features (*e.g.*, 3D HOG) for all landmarks can be extracted and concatenated as strong features for classification. However, there are two challenges in longitudinal study: (1) The numbers of scanning time points across subjects are inconsistent due to missing time points, and thus, it is difficult to extract a unified feature representation from different number of scans. (2) It is difficult to identify the corresponding baseline images across subjects, which means a baseline scanning time point of one subject may not correspond to that of another subject. How to extract a unified

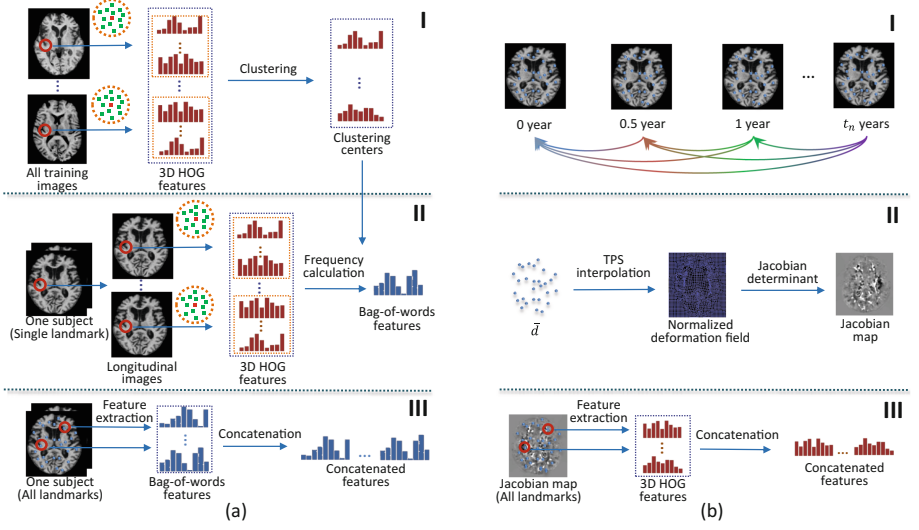


Fig. 1. Landmark-based feature extraction steps. (a) Spatial features. (b) Longitudinal features.

spatial feature representation from those inconsistent longitudinal scans is a very challenging task.

To address these two problems, we propose to use a bag-of-words strategy to extract statistical high-level spatial features. The bag-of-words strategy has demonstrated impressive performance on text, language, and image classification [25–28]. Specifically, Fig. 1(a) shows the procedure of our spatial feature extraction method, where each landmark is treated independently. As shown in Fig. 1(a) I, we first extract the 3D HOG feature vector for each landmark, as well as 3D HOG feature vectors for the supplementary voxels (*i.e.*, the neighboring voxels within a small spherical patch of the landmark). After extracting features from all training images and aggregating them together, we have a set of 3D HOG feature vectors. Then, we perform K-means clustering [29] on this set of feature vectors, and build a dictionary (*i.e.*, \mathcal{D}) with its *words* (*i.e.*, $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M$) being the clustering centers. Then, for each individual subject, we can first extract the 3D HOG feature vectors (denoted by a feature set \mathcal{F}) for each landmark and its supplementary voxels in all longitudinal scans. The statistical histogram representation is then calculated by counting the occurrence frequencies of the clustering centers in these HOG features (*i.e.*, \mathcal{F}), as shown in Fig. 1(a) II. Mathematically, the histogram representation (*i.e.*, R) for one landmark can be defined as

$$R(j) = \sum_{\mathbf{f} \in \mathcal{F}} \delta(\argmin_{i \in \{1, \dots, M\}} \|\mathbf{f} - \mathbf{w}_i\|_2^2 = j), \quad j = 1, \dots, M, \quad (1)$$

where $\delta(\cdot)$ is the Kronecker delta function defined as

$$\delta(A) = \begin{cases} 1 & \text{if } A \text{ is true;} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

In order to achieve the invariance to the number of longitudinal scans, the histogram representation is ℓ_1 normalized. Finally, we extract the statistical features for all landmarks, regardless of differences in the number of scanning time points, as shown in Fig. 1(a) III. Here, the reasons for using supplementary voxels in neighborhood of landmarks are two-fold: (1) The HOG feature set can be expanded to get statistical features by using the bag-of-word strategy; (2) It is also helpful to relieve potential errors in localizing landmark positions.

3.2 Landmark-Based Longitudinal Feature Extraction

In order to solve the problem of inconsistent longitudinal scans, we generate the normalized 3D longitudinal displacement at the beginning of feature extraction. Specifically, we first define the longitudinal displacement between two scans for one specific landmark as follows:

$$d_{i,j} = L_{t_i} - L_{t_j}, \quad (3)$$

where L_{t_i} is the landmark location of the i -th scan from all longitudinal scans and t_i is the corresponding relative scanning time point with respect to the first scan. Then, the normalized 3D displacement \bar{d} (mean displacement per year) is calculated from all possible combinations between two scans in different scanning time points, as shown in Fig. 1(b) I. Mathematically, \bar{d} is defined as follows:

$$\bar{d} = \frac{1}{\sum_{1 \leq j < i \leq n} 1} \sum_{1 \leq j < i \leq n} \frac{d_{i,j}}{t_i - t_j}, \quad (4)$$

where n is the number of existing scans. As shown in Fig. 1(b) II, a normalized deformation field can be built by applying thin plate splines (TPS) interpolation to the normalized 3D longitudinal displacement \bar{d} of all landmarks. Based on this normalized deformation field, a Jacobian map is further calculated to describe the longitudinal volume variations. Finally, as shown in Fig. 1(b) III, we can extract morphological features (*i.e.*, 3D HOG) for the landmarks in the Jacobian map. Therefore, longitudinal volume variations on these discriminative landmarks can be captured by these morphological features. It is worth noting that, instead of treating each landmark individually, the neighboring landmarks are jointed together with interpolation during the generation of the normalized deformation field. In this way, although the morphological features from Jacobian map are extracted for each landmark individually, the contextual information about the neighboring landmarks is automatically embedded into the calculated features.

4 Experiments

4.1 Parameter Setup

Using a 10-fold cross validation strategy, we conducted experiments for two classification tasks, *i.e.*, AD vs. HC and MCI vs. HC. The parameters in our approach were defined as follows: For 3D HOG feature extraction, we used 9 orientations, $2 \times 2 \times 2$ cells, and a size of $8 \times 8 \times 8$ for each cell. Therefore, the dimensionality of 3D HOG features was 72. In the bag-of-words strategy, the number of clustering centers was set to 50, and thus, the dimensionality of spatial features for each landmark was 50. The radius of spherical patch for sampling supplementary voxels was 5. For SVM classification, we fixed the margin parameter $C = 1$. Due to the data-driven property of our method, the number of landmarks was determined by the training images. In our method, we searched the local minima within a $7 \times 7 \times 7$ cubic patch, and obtained roughly 1500 identified landmarks for each fold in the cross validation.

4.2 Experimental Results

Five classification performance measures were used, namely (1) accuracy (ACC): the number of correctly classified samples divided by the total number of samples; (2) sensitivity (SEN): the number of correctly classified positive samples (patients) divided by the total number of positive samples; (3) specificity (SPE): the number of correctly classified negative samples (controls) divided by the total number of negative samples; and (4) balanced accuracy (BAC): the mean value of sensitivity and specificity; (5) area under receiver operating characteristic (ROC) curve (AUC).

For comparison, we also report the classification results of two baseline strategies based on our landmarks. The baseline spatial features are the HOG features that are directly extracted according to the landmarks only from the baseline

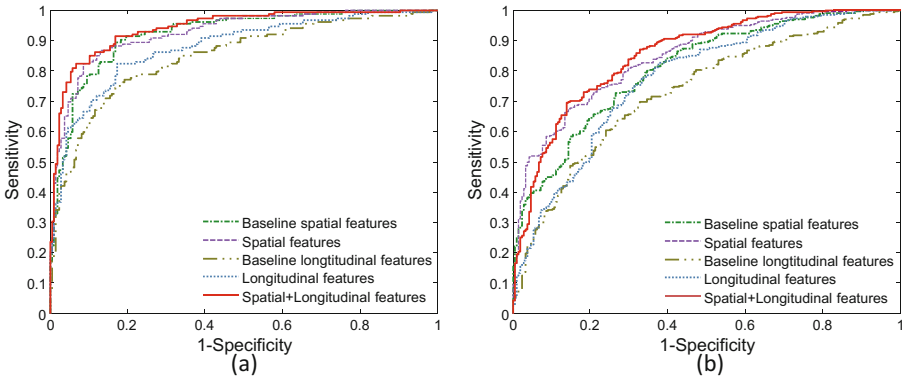


Fig. 2. ROC curves for classification. (a) AD vs. HC. (b) MCI vs. HC.

Table 3. Classification results achieved by methods using different features.

	AD vs. HC (%)					MCI vs. HC (%)				
	ACC	SEN	SPE	BAC	AUC	ACC	SEN	SPE	BAC	AUC
Baseline spatial features	84.40	74.34	91.79	83.06	91.98	74.86	84.68	58.45	71.5	80.78
Spatial features	86.35	77.63	92.75	85.19	92.52	76.49	86.13	60.39	73.26	84.57
Baseline longitudinal features	78.83	72.37	83.57	77.97	84.64	69.08	78.32	53.62	65.97	73.04
Longitudinal features	80.78	77.63	83.09	80.36	87.69	72.88	85.55	51.69	68.62	77.26
Spatial+Longitudinal features	88.30	79.61	94.69	87.15	94.01	79.02	90.46	59.90	75.18	85.19

MR image (first scan). The baseline longitudinal features refer to the features obtained by directly using normalized displacements (*i.e.*, \bar{d}) of the landmarks.

Table 3 reports the classification results, and Fig. 2 shows their corresponding ROC curves. These results demonstrate that, in both classification tasks, the proposed spatial features consistently outperform the baseline spatial features, and our longitudinal features generally achieve better performance than the baseline longitudinal features. Moreover, the combination of the proposed spatial and longitudinal features can further improve the classification performance.

In a related work, Chincarini et al. [15] used hippocampal volume and hippocampal volume atrophy rate as measurements for longitudinal AD classification. The reported AUC for AD vs. HC on ADNI-1 is 93.00%, which is slightly lower than ours (94.01%). Moreover, they used multi-atlas based method to obtain the hippocampal segmentations which is time-consuming. As we know, it usually takes hours to get accurate hippocampal segmentation. For our landmark-based method (*e.g.*, using four longitudinal scans), it takes less than 3 min to complete all feature extraction steps, including linear registration, landmark detection, and spatial and longitudinal feature extraction.

5 Discussions and Conclusions

Landmark-based Framework: The major advantages of using landmark-based framework are two folds: (1) The identified discriminative landmarks can cover all possible abnormalities from the whole brain without using several pre-defined biomarkers; (2) The use of landmarks makes it possible to integrate a fast landmark detection model to the diagnosis framework such that both time-consuming nonlinear registration and tissue segmentation are avoided. It is worth noting that, although each landmark is a weak descriptor that only covers the information from a small local patch, thousands of landmarks can well describe the brain structure and thus leading to a stable classification performance.

Spatial Features: In the bag-of-words representation, *words* in the dictionary can be regarded as representative local spatial structures. Thus, the calculation of their occurrence frequency can be regarded as labeling the spatial structure of each landmark with its similarities to all *words*. This high-level statistic ignores both the numbers and the orders of scanning time points and only focuses on the spatial abnormalities, which is suitable for extracting unified spatial features

from inconsistent longitudinal scans. As can be seen in Table 3 and Fig. 2, the method of using bag-of-words based spatial features achieves better classification performance, compared with that using baseline spatial features.

Longitudinal Features: Intuitively, one type of longitudinal information is the trajectory of landmarks along time. However, the coherence among neighboring landmarks is ignored if we just simply use the mean longitudinal displacements (\bar{d}) as features. In our method, we generate a normalized deformation field by interpolation, through which the contextual information can be employed by jointly using the neighboring landmarks. Moreover, it is also well known that the Jacobian determinant can indicate the volume variation. Therefore, the morphological features from the Jacobian map comprehensively capture the longitudinal volume variation around landmarks. The experimental results show that using the longitudinal features from Jacobian map achieves 2% to 4% improvement in terms of accuracy as compared with the baseline longitudinal features.

Limitations and Future Work: Since each landmark has 72 spatial features and 50 longitudinal features, the concatenation of the features from all landmarks would be high dimensional, with respect to the number of training subjects. Also, there may be some redundant or noisy features that can adversely affect the classification model learning. Therefore, selecting most discriminative landmarks and features is important and will provide a reasonable solution for further performance improvement, which is our future work.

References

1. Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M.: The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* **6**(2), 67–77 (2010)
2. Thung, K.H., Wee, C.Y., Yap, P.T., Shen, D., Initiative, A.D.N., et al.: Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion. *NeuroImage* **91**, 386–400 (2014)
3. Thung, K.-H., Yap, P.-T., Adeli-M, E., Shen, D.: Joint diagnosis and conversion time prediction of progressive mild cognitive impairment (PMCI) using low-rank subspace clustering and matrix completion. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 527–534. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4_63](https://doi.org/10.1007/978-3-319-24574-4_63)
4. Liu, M., Zhang, D., Shen, D.: View-centralized multi-atlas classification for Alzheimer’s disease diagnosis. *Hum. Brain Mapp.* **36**(5), 1847–1865 (2015)
5. Liu, M., Zhang, D., Adeli-Mosabbe, E., Shen, D.: Inherent structure based multi-view learning with multi-template feature representation for Alzheimer’s disease diagnosis. *IEEE Trans. Biomed. Eng.* **63**(7), 1473–1482 (2016)
6. Liu, M., Zhang, D., Shen, D.: Relationship induced multi-template learning for diagnosis of Alzheimer’s disease and mild cognitive impairment. *IEEE Trans. Med. Imaging* **35**(6), 1463–1474 (2016)
7. Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M.K., Johnson, S.C., Initiative, A.D.N., et al.: Spatially augmented lppboosting for ad classification with evaluations on the adni dataset. *Neuroimage* **48**(1), 138–149 (2009)

8. Zhu, X., Suk, H.I., Shen, D.: A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *NeuroImage* **100**, 91–105 (2014)
9. Zhu, X., Suk, H.I., Lee, S.W., Shen, D.: Canonical feature selection for joint regression and multi-class identification in Alzheimer's disease diagnosis. *Brain Imaging Behav.* **10**(3), 1–11 (2015)
10. Zhu, X., Suk, H.I., Lee, S.W., Shen, D.: Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification. *IEEE Trans. Biomed. Eng.* **63**(3), 607–618 (2016)
11. Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.S., Niethammer, M., Dubois, B., Lehericy, S., Garnero, L., et al.: Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging. *Neuroimage* **47**(4), 1476–1486 (2009)
12. Gao, Y., Adeli-M., E., Kim, M., Giannakopoulos, P., Haller, S., Shen, D.: Medical image retrieval using multi-graph learning for MCI diagnostic assistance. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9350, pp. 86–93. Springer, Cham (2015). doi:[10.1007/978-3-319-24571-3_11](https://doi.org/10.1007/978-3-319-24571-3_11)
13. Gao, Y., Wee, C.-Y., Kim, M., Giannakopoulos, P., Montandon, M.-L., Haller, S., Shen, D.: MCI identification by joint learning on multiple MRI data. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9350, pp. 78–85. Springer, Cham (2015). doi:[10.1007/978-3-319-24571-3_10](https://doi.org/10.1007/978-3-319-24571-3_10)
14. Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Schill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S.: Automatic classification of MR scans in Alzheimer's disease. *Brain* **131**(3), 681–689 (2008)
15. Chincarini, A., Sensi, F., Rei, L., Gemme, G., Squarcia, S., Longo, R., Brun, F., Tangaro, S., Bellotti, R., Amoroso, N., et al.: Integrating longitudinal information in hippocampal volume measurements for the early detection of Alzheimer's disease. *NeuroImage* **125**, 834–847 (2016)
16. Jack, C., Shiung, M., Gunter, J., Obrien, P., Weigand, S., Knopman, D.S., Boeve, B.F., Ivnik, R.J., Smith, G.E., Cha, R., et al.: Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology* **62**(4), 591–600 (2004)
17. Aguilar, C., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Wahlund, L.O., Simmons, A., et al.: Application of a MRI based index to longitudinal atrophy change in Alzheimer disease, mild cognitive impairment and healthy older individuals in the addneuromed cohort. *Front. Aging Neurosci.* **6**, 145 (2014)
18. Thung, K.H., Wee, C.Y., Yap, P.T., Shen, D.: Identification of progressive mild cognitive impairment patients using incomplete longitudinal MRI scans. *Brain Struct. Funct.* 1–17 (2015)
19. Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C.: Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr.* **22**(2), 324–333 (1998)
20. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005*. CVPR 2005, pp. 886–893. IEEE (2005)
21. Mardia, K.: Assessment of multinormality and the robustness of Hotelling's T^2 test. *Appl. Stat.* **24**, 163–171 (1975)
22. Zhang, J., Gao, Y., Gao, Y., Brent, M., Shen, D.: Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. *IEEE Trans. Med. Imaging* **35**(12), 2524–2533 (2016)

23. Gao, Y., Shen, D.: Context-aware anatomical landmark detection: application to deformable model initialization in prostate CT images. In: Wu, G., Zhang, D., Zhou, L. (eds.) MLMI 2014. LNCS, vol. 8679, pp. 165–173. Springer, Cham (2014). doi:[10.1007/978-3-319-10581-9_21](https://doi.org/10.1007/978-3-319-10581-9_21)
24. Zhang, J., Gao, Y., Wang, L., Tang, Z., Xia, J.J., Shen, D.: Automatic craniomaxillofacial landmark digitization via segmentation-guided partially-joint regression forest model and multi-scale statistical features. *IEEE Trans. Biomed. Eng.* **63**(9), 1820–1829 (2016)
25. Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int. J. Comput. Vision* **43**(1), 29–44 (2001)
26. Nowak, E., Jurie, F., Triggs, B.: Sampling strategies for bag-of-features image classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 490–503. Springer, Heidelberg (2006). doi:[10.1007/11744085_38](https://doi.org/10.1007/11744085_38)
27. Yang, J., Jiang, Y.G., Hauptmann, A.G., Ngo, C.W.: Evaluating bag-of-visual-words representations in scene classification. In: Proceedings of the International Workshop on Multimedia Information Retrieval, pp. 197–206. ACM (2007)
28. Jiang, Y.G., Ngo, C.W., Yang, J.: Towards optimal bag-of-features for object categorization and semantic video retrieval. In: Proceedings of the 6th ACM International Conference on Image and Video Retrieval, ACM 494–501(2007)
29. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: A k-means clustering algorithm. *J. Roy. Stat. Soc.: Ser. C (Appl. Stat.)* **28**(1), 100–108 (1979)