Random Forest-Based Manifold Learning for Classification of Imaging Data in Dementia

Katherine R. Gray[1⋆], Paul Aljabar[1], Rolf A. Heckemann[2,3], Alexander Hammers[2,3], and Daniel Rueckert[1]

[1] Department of Computing, Imperial College London, United Kingdom
[2] Fondation Neurodis, CERMEP-Imagerie du Vivant, Lyon, France
[3] Faculty of Medicine, Imperial College London, United Kingdom
katherine.gray03@imperial.ac.uk

Abstract. Neurodegenerative disorders are characterized by changes in multiple biomarkers, which may provide complementary information for diagnosis and prognosis. We present a framework in which proximities derived from random forests are used to learn a low-dimensional manifold from labelled training data and then to infer the clinical labels of test data mapped to this space. The proposed method facilitates the combination of embeddings from multiple datasets, resulting in the generation of a joint embedding that simultaneously encodes information about all the available features. It is possible to combine different types of data without additional processing, and we demonstrate this key feature by application to voxel-based FDG-PET and region-based MR imaging data from the ADNI study. Classification based on the joint embedding coordinates out-performs classification based on either modality alone. Results are impressive compared with other state-of-the-art machine learning techniques applied to multi-modality imaging data.

1 Introduction

There is much interest in the application of manifold learning techniques to medical imaging data. These techniques aim to convert high-dimensional data to a lower-dimensional representation in which further analysis, such as classification, may be more easily performed. This is relevant to neuroimaging, where a considerable amount of research focuses on the identification of imaging biomarkers which are desirable for improved diagnosis and monitoring, and drug discovery.

Typically, measures encoding pairwise similarities between images are used in manifold learning. For example, Laplacian eigenmaps [1] have been used to generate an embedding of brain MR images based on similarities derived from overlaps of their structural segmentations [2]. Laplacian eigenmaps have also

⋆ This project is partially funded under the 7th Framework Programme by the European Commission (http://cordis.europa.eu/ist). Imaging data were provided by the Alzheimer's Disease Neuroimaging Initiative. KRG received a studentship from the EPSRC. RAH was supported by a research grant from the Dunhill Medical Trust.

been applied to ultrasound data to aid detection of breathing motion for image-based respiratory gating [3]. The manifold structure of brain MR images has been estimated by applying Isomap [4] using distance measures based on non-rigid transformations between image pairs [5]. A framework for fusing manifold learning steps, based on multiple pairwise similarity measures, is proposed in [6].

Here, we derive a low-dimensional manifold from labelled training images and use it to infer the clinical labels of test images mapped to this space. We use proximity measures derived from random forests [7], coupled with multidimensional scaling (MDS) [8], to learn the manifold on which to perform classification. The use of MDS with a random forest-derived proximity matrix is commonly used for low-dimensional data visualization [9]. Random forest proximities have also been successfully applied in unsupervised clustering tasks, in particular those involving genetic data [10]. We derive a supervised similarity measure which should generate a manifold that is optimal for the task of clinical group discrimination. We evaluate the proposed method using FDG-PET and MR imaging data from 287 participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI)[4].

The proposed framework facilitates the incorporation of multi-modality data, since combining proximities derived from two modalities generates an embedding that simultaneously encodes information about both. This is a key feature of the method, since neurodegenerative disorders such as Alzheimer's disease are characterized by changes in multiple biomarkers which may prove more powerful when used in combination. Recently, improvements in classification accuracy have been reported when combining FDG-PET and MR imaging data using multi-kernel learning [11, 12]. Importantly, the proposed method can combine different types of features without additional processing. We demonstrate this by application to voxel-based FDG-PET and region-based MR imaging data.

Our contributions are (1) the application of manifold learning to a diverse multi-modality set of brain images, (2) the use of a supervised proximity matrix derived from random forests to develop a clinically relevant manifold for classification, and (3) the proposal of a framework within which to combine multi-modality data in a single joint manifold, leading to state-of-the-art classification performance superior to that achievable using data from a single modality.

2 Methods

Features were extracted from the images using voxel-based analysis for the FDG-PET and region-based analysis for the MRI. Details of the image processing and feature extraction steps are provided in Section 3.1. A random forest classifier was applied to the data from each modality independently, firstly to obtain a baseline classification rate for comparison, but also to derive the proximity matrices required for manifold learning. Classical MDS was applied to each proximity matrix to generate a low-dimensional embedding for each modality. The two proximity matrices were then combined, and MDS applied to generate a joint embedding. A schematic illustration of the approach is provided in Figure 1.

[4] www.loni.ucla.edu/ADNI

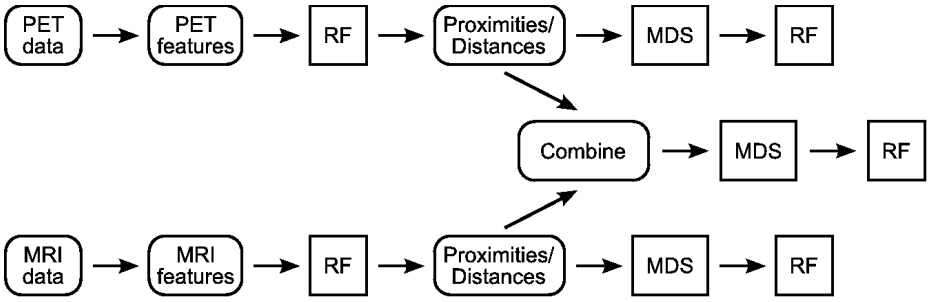

Fig. 1. An overview of the proposed analysis pipeline. Each random forest (RF) step provides a classification result whose performance is reported in Section 3.2.

2.1 Random Forests for Classification

A random forest is an ensemble classifier consisting of many decision trees, where the final predicted class for a test object is the mode of the predictions of all individual trees. For a dataset consisting of N objects, each with M features, a value $m << M$ is selected, and each tree grown as follows. To construct the training set for a tree, N objects are sampled at random with replacement. At each node in the tree, m features are randomly selected from the available M, and the node is partitioned using the best possible binary split. Each tree is fully grown without pruning. We use the R implementation of random forests[5].

2.2 Manifold Learning based on Random Forest Proximities

Random forest proximities provide pairwise measures of the similarities between objects in the dataset. All N objects are passed down each tree in the forest, and if objects i and j finish in the same terminal node, their proximity, p_{ij}, is increased by one. The proximities are normalized by the total number of trees.

The proximities form an $N \times N$ matrix, P_{ij}, for which values of the corresponding distance matrix, $D_{ij} = 1 - P_{ij}$, may be viewed as squared distances in a Euclidean space of dimension not greater than the number of objects [8]. MDS can be applied to this distance matrix to generate a lower-dimensional embedding for the data. To generate a joint embedding that simultaneously incorporates information from both modalities, the distance matrices derived from the individual modalities are additively combined, and MDS applied to the resulting joint proximity matrix.

3 Data and Results

3.1 Imaging Data and Feature Extraction

We applied the proposed methods to images of 287 ADNI participants for whom baseline 1.5T MRI and FDG-PET images are available. These include 71 patients

[5] cran.r-project.org/web/packages/randomForest

with Alzheimer's disease (AD), 147 patients with mild cognitive impairment (MCI), and 69 age-matched healthy controls (HC).

MRI: region-based feature extraction. Automatic whole-brain segmentations into 83 anatomical regions were prepared in native-space using multi-atlas propagation with enhanced registration (MAPER), an approach that has been previously described and validated for use in AD [13]. The segmentations are available to download through the ADNI website, and full details of the procedure and morphometric analysis are presented in [14]. Individual tissue probability maps for gray matter, white matter and cerebrospinal fluid were obtained using FSL FAST[6]. For feature extraction, masked segmentations were employed, in which all regions except ventricles, central structures, cerebellum and brainstem were masked with a gray matter label, and lateral ventricles with a cerebrospinal fluid label. Regional volumes were calculated and normalized by the intracranial volume, resulting in 83 volumetric region-based features per image.

FDG-PET: voxel-based feature extraction. Each FDG-PET image was motion-corrected as necessary, converted to a 30-minute static image, and affinely aligned with the corresponding MRI using the Image Registration Toolkit (IRTK)[7]. An affine transformation was preferred over a rigid one because it can account for any scaling or voxel size errors remaining after phantom correction of the MRI [15]. Using the "Segment" module of SPM5[8], each MRI was linearly and non-linearly deformed to the MNI template. The resulting transformation parameters were applied to the MR-space FDG-PET images using a trilinear interpolation. The MNI-space FDG-PET images were smoothed to a common isotropic spatial resolution of 8mm FWHM using scanner-specific kernels [16], and then by an additional 8mm FWHM isotropic Gaussian kernel. The smoothed images were normalized to account for inter-subject variability in overall radioactivity using a reference cluster derived from an independent dataset [17]. A MNI-space brainmask was applied to each normalized FDG-PET image, thresholded at 50% to exclude background, and signal intensities were extracted from each voxel, resulting in 239,304 features per image.

3.2 Results

We assess classification performance between two clinically significant pairs of groups: AD patients vs HC, and MCI patients vs HC. All experiments were assessed using ten-fold cross-validation. Since there are approximately twice as many MCI patients as HC, these data were balanced according to the class distributions during training. Before classification, the number of features randomly selected at each node, m, and the number of trees grown in each forest, t, had to

[6] www.fmrib.ox.ac.uk/fsl
[7] www.doc.ic.ac.uk/∼dr/software
[8] www.fil.ion.ucl.ac.uk/spm

be selected. We used 5,000 trees, and the recommended default value, $m = \sqrt{M}$. Table 1 shows the classification performance following the application of a random forest classifier to the original imaging data. In addition, the FDG-PET and MRI features were combined by concatenatation, and a random forest classifier applied. The performance based on this feature set is also shown, and does not significantly differ from the results based solely on the FDG-PET features.

Table 1. Classification accuracy, sensitivity and specificity following application of a random forest classifier to the original imaging data. Classification is performed using ten-fold cross-validation, and mean (standard error) values are reported.

	AD vs HC			MCI vs HC		
	Acc. (%)	Sens. (%)	Spec. (%)	Acc. (%)	Sens. (%)	Spec. (%)
Region-based MRI	84.4 (1.6)	83.2 (3.5)	85.5 (3.0)	64.4 (3.0)	59.3 (4.5)	75.7 (5.2)
Voxel-based FDG-PET	87.9 (2.6)	92.0 (3.9)	83.8 (4.7)	63.9 (2.3)	59.1 (2.8)	74.1 (4.6)
Concatenated features	87.9 (2.6)	92.0 (3.9)	83.8 (4.7)	64.3 (2.4)	59.8 (3.2)	74.0 (4.6)

After applying MDS to the proximity matrix for each modality, the eigenvectors corresponding to the 25 largest-valued eigenvalues were used in generating low-dimensional embeddings. The value of 25 was empirically determined to ensure that zero-valued eigenvalues were not included, whilst capturing the maximum possible amount of information. Table 2 shows the classification performance following application of a random forest classifier to the separate embedding coordinates for each modality. The accuracy achieved based on the embedding coordinates does not differ significantly from that achieved using the original imaging data. Results based on the joint embedding are also shown. This out-performs the corresponding application to the separate embedding coordinates. The improvement is significant ($p < 0.05$) for the MCI vs HC classification.

Table 2. Classification accuracy, sensitivity and specificity obtained following the application of a random forest classifier to the embedded data from the two individual modalities, as well as the joint embedding. Classification is performed using ten-fold cross-validation, and mean (standard error) values are reported.

	AD vs HC			MCI vs HC		
	Acc. (%)	Sens. (%)	Spec. (%)	Acc. (%)	Sens. (%)	Spec. (%)
Region-based MRI	87.2 (2.0)	87.5 (3.2)	86.9 (2.6)	64.8 (3.0)	64.8 (3.9)	65.5 (6.7)
Voxel-based FDG-PET	87.8 (2.6)	91.8 (2.9)	83.8 (5.1)	65.3 (1.9)	65.3 (2.9)	65.2 (3.8)
Combined embedding	90.0 (2.6)	88.9 (3.4)	89.8 (3.8)	75.5 (2.2)	76.9 (3.2)	72.4 (4.5)

It is possible to extract estimates of the importance of each feature for classification from the random forest. The importance measure for an individual feature

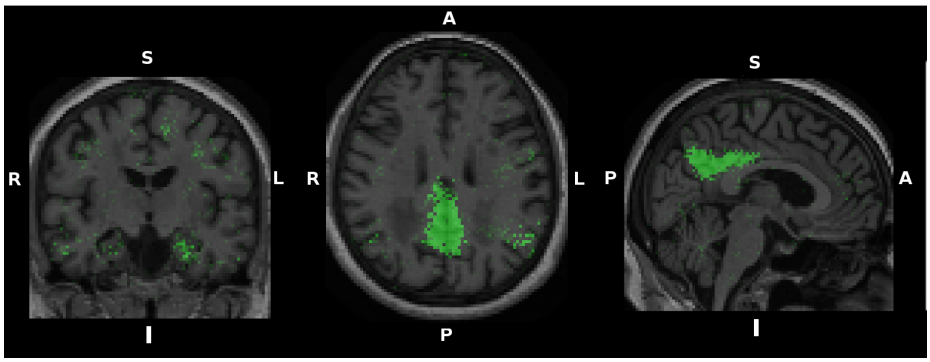

Fig. 2. Feature importances for discriminating between AD patients and HC using FDG-PET imaging data. All voxels with an importance measure greater than 5% of the maximum value are displayed in green, and overlaid onto the MNI-space MR image of a typical HC. R: right, L: left, S: superior, I: inferior, A: anterior, P: posterior.

is determined by summing the decreases in the Gini impurity criterion [18] over all nodes in the forest that are partitioned based on that feature. The Gini impurity of a node measures the likelihood that an object would be incorrectly labelled if it were randomly assigned a label according to the distribution of labels within the node. Feature importances for discriminating between clinical groups were extracted from the random forest classifiers applied to the original imaging data. For FDG-PET, Figure 2 shows the locations of the most important voxels for discriminating between AD patients and HC. Important voxels are mostly found in the posterior cingulate gyrus, parietal lobe, posterior temporal lobe, and around the hippocampus. For discriminating between MCI patients and HC, important voxels are also located in these areas, but at lower magnitude, as well as in the frontal cortex. For MRI, the most important features include the hippocampus, amygdala, and other medial temporal lobe structures. A larger number of regions are identified as relatively important in distinguishing MCI patients from HC, but with lower magnitude.

4 Discussion and Conclusions

We have presented a framework in which proximity measures derived from random forests are used to learn a low-dimensional manifold from labelled training data and then to infer the clinical labels of test data mapped to this space. The applicability of the method to a large and diverse set of brain images is demonstrated using FDG-PET and MR images from the ADNI study. A key feature of the method is its ability to incorporate multi-modality data, since it is unlikely that AD can be fully characterized by a single biomarker. The method could be extended to include any number of features, for example longitudinal imaging data, non-imaging data, and categorical and genetic features. We demonstrate

the potential to combine different types of features without additional processing, providing the freedom to perform feature extraction in the most appropriate way for each individual dataset. In other similar methods, such as that described in [6], a scaling step is required before separate embeddings may be combined.

We assess the performance of the method by application to voxel-based FDG-PET and region-based MR imaging data from the ADNI study. Classification based on the joint embedding derived from these two modalities out-performs classification based on either modality alone. This supports previous suggestions that there is some complementary information between MRI and FDG-PET which can be exploited to produce a more powerful combined biomarker for AD. The lack of significant difference between classification performance based on the original imaging data compared with the embedding coordinates for each individual modality is to be expected, since a random forest is already a nonlinear classifier. The motivation for the embedding step was the incorporation of multi-modality data. We demonstrate that a simple concatenation of FDG-PET and MRI features does not optimally combine these data, as this does not improve classification performance compared with the single modalities.

The classification accuracy for discriminating MCI patients from HC based on the joint embedding is impressive compared with other state-of-the-art multi-modality machine learning techniques. For example, [11] also report an accuracy of 76% using multi-kernel learning, but based on the combination of MRI, FDG-PET and cerebrospinal fluid biomarkers. The classification accuracy for discriminating AD patients from HC is in line with other state-of-the-art methods which use either single-modality [21] or multi-modality [11, 12] imaging data. Since the diagnostic labels used as a gold-standard themselves have an accuracy of around 90% [22], significant further improvement may not be possible using this dataset.

The ability to extract feature importances from the random forest is valuable because it allows verification that the features contributing most to the classification make biological sense. The most important features for discriminating between HC and both AD and MCI patients are in line with those known to be different between clinical groups in the respective modalities [19, 20]. The important features for discriminating AD patients from HC are more localized to affected areas, with the more challenging discrimination between MCI patients and HC also requiring features spread across a wider area of the brain.

The method is readily generalizable, in that the manifold learning step could be performed using Laplacian eigenmaps, Isomap, or any other suitable algorithm. Classification performance could be assessed by applying a clustering algorithm to the embedding coordinates, and proximities could be combined using a more sophisticated metric. Future work will explore these options, and consider other promising datasets, such as cerebrospinal fluid biomarkers.

References

1. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation 15(6), 1373–1396 (2003)

2. Aljabar, P., Rueckert, D., Crum, W.: Automated morphological analysis of magnetic resonance brain imaging using spectral analysis. Neuroimage 43(2), 225–235 (2008)
3. Wachinger, C., Yigitsoy, M., Navab, N.: Manifold learning for image-based breathing gating with application to 4D ultrasound. In: MICCAI 2010. LNCS, 6362, 26–33. Springer, Heidelberg (2010)
4. Tenenbaum, J., de Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319-2323 (2000)
5. Gerber, S., Tasdizen, T., Joshi, S., et al.: On the manifold structure of the space of brain images. In: MICCAI 2009. LNCS, 5761, 305-312. Springer, Heidelberg (2009)
6. Aljabar, P., Wolz, R., Srinivasan, L., et al.: Combining morphological information in a manifold learning framework: application to neonatal MRI. In: MICCAI 2010. LNCS, 6363, 1–8. Springer, Heidelberg (2010)
7. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
8. Cox, T.F., Cox, M.A.A.: Multidimensional scaling. Chapman and Hall (2001)
9. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning: data mining, inference, and prediction, 2nd ed. Springer, corrected 5th printing (2011)
10. Shi, T., Horvath, S.: Unsupervised learning with random forest predictors. J. Comp. Graph. Stat. 15(1), 118–138 (2006)
11. Zhang, D., Wang, Y., Zhou, L., et al.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 55(3), 856–867 (2011)
12. Hinrichs, C., Singh, V., Xu, G., et al.: Predictive markers for AD in a multimodality framework: an analysis of MCI progression in the ADNI population. Neuroimage 55(2), 574–589 (2011)
13. Heckemann, R.A., Keihaninejad, S, Aljabar, P, et al: Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. Neuroimage 51(1), 221–227 (2010)
14. Heckemann, R.A., Keihaninejad, S., Aljabar, P., et al.: Automatic morphometry in Alzheimer's disease and mild cognitive impairment. Neuroimage 56(4), 2024–2037 (2011)
15. Clarkson, M.J., Ourselin, S., Nielsen, C., et al.: Comparison of phantom and registration scaling corrections using the ADNI cohort. Neuroimage 47(4), 1506–1513 (2009)
16. Joshi, A., Koeppe, R.A., Fessler, J.A.: Reducing between scanner differences in multi-center PET studies. Neuroimage 46(1), 154–159 (2009)
17. Yakushev, I., Hammers, A., Fellgiebel, A., et al.: SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnestic mild cognitive impairment from healthy aging. Neuroimage 44(1), 43–50 (2009)
18. Breiman, L., Friedman, J.H., Olshen, R.A., et al.: Classification and regression trees. Wadsworth, Belmont, California (1984)
19. Hampel, H., Burger, K., Teipel, S.J., et al.: Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. Alzh. & Dementia 4(1), 38–48 (2008)
20. Patwardhan, M.B., McCrory, D.C., Matchar, D.B., et al.: Alzheimer disease: operating characteristics of PET – a meta-analysis. Radiology 231(1), 73-80 (2004)
21. Cuingnet, R., Gerardin, E., Tessieras, J., et al.: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage 56(2), 766–781 (2011)
22. Ranginwala, N.A., Hynan, L.S., Weiner, M.F., et al.: Clinical criteria for the diagnosis of Alzheimer disease: still good after all these years. Am. J. Geriat. Psychiatry 16(5), 384–388 (2008)