NeuroImage: Clinical 7 (2015) 367–376

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Content-based image retrieval for brain MRI: An image-searching engine and population-based analysis to utilize past clinical data for future diagnosis

Andreia V. Faria[a,*], Kenichi Oishi[a], Shoko Yoshida[a], Argye Hillis[b,c,d], Michael I. Miller[e], Susumu Mori[a,f]

[a]The Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, MD, USA
[b]Department of Neurology, Johns Hopkins University, Baltimore, MD, USA
[c]Department of Physical Medicine & Rehabilitation Medicine, Johns Hopkins University, Baltimore, MD, USA
[d]Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA
[e]Department of Biomedical Engineering, The Johns Hopkins University School of Medicine, Baltimore, MD, USA
[f]F.M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, MD, USA

ARTICLE INFO

Article history:
Received 28 January 2014
Received in revised form 5 December 2014
Accepted 13 January 2015
Available online 15 January 2015

Keywords:
Automated parcellation
Brain
MRI
Content-based image retrieval
Atlas-based analysis

ABSTRACT

Radiological diagnosis is based on subjective judgment by radiologists. The reasoning behind this process is difficult to document and share, which is a major obstacle in adopting evidence-based medicine in radiology. We report our attempt to use a comprehensive brain parcellation tool to systematically capture image features and use them to record, search, and evaluate anatomical phenotypes. Anatomical images (T1-weighted MRI) were converted to a standardized index by using a high-dimensional image transformation method followed by atlas-based parcellation of the entire brain. We investigated how the indexed anatomical data captured the anatomical features of healthy controls and a population with Primary Progressive Aphasia (PPA). PPA was chosen because patients have apparent atrophy at different degrees and locations, thus the automated quantitative results can be compared with trained clinicians' qualitative evaluations. We explored and tested the power of individual classifications and of performing a search for images with similar anatomical features in a database using partial least squares-discriminant analysis (PLS-DA) and principal component analysis (PCA). The agreement between the automated z-score and the averaged visual scores for atrophy ($r = 0.8$) was virtually the same as the inter-evaluator agreement. The PCA plot distribution correlated with the anatomical phenotypes and the PLS-DA resulted in a model with an accuracy of 88% for distinguishing PPA variants. The quantitative indices captured the main anatomical features. The indexing of image data has a potential to be an effective, comprehensive, and easily translatable tool for clinical practice, providing new opportunities to mine clinical databases for medical decision support.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Currently, radiological diagnosis is based on subjective judgment by radiologists, in which image-based findings and text-based clinical information are integrated to reach a decision. This ability is acquired by education and personal experience, and the exact reasoning leading to the final conclusion is difficult to document and share. This is a major obstacle to adopting evidence-based medicine in radiology, which calls for a systematic integration of past evidence for medical decision making (Hsu, 2012). Presently, radiological data are stored in PACS (Picture Archive and Communication System). Once a diagnosis is completed, this rich data source is rarely used to enrich future daily diagnostic capability. The heart of the problem is that the data in PACS are not fully structurized, which makes information retrieval impractical.

The PACS contains two types of information: text and images. The text is usually composed of a few structured and searchable fields, such as name, gender, and age, as well as free-text fields with image descriptions and diagnosis. There have been many ongoing attempts to structurize the PACS free-text fields (Radiological Society of North America, 2014a,b; Hussein, 2004). If successful, physicians would be able to search, for example, for "past cases with a diagnosis of Alzheimer's disease (AD)," and, in return, say, they would retrieve 100 images of AD patients. This is very useful for research purposes. However, the retrieved 100 images may not be directly useful in supporting routine clinical decisions.

* Corresponding author at: 217 B. Traylor Bldg., 720 Rutland Ave., Baltimore, MD 21205, USA. Tel.: +1 6149554215; fax: +1 410 614 1948.
E-mail address: afaria1@jhmi.edu (A.V. Faria).

http://dx.doi.org/10.1016/j.nicl.2015.01.008
2213-1582/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Table 1
Demographic and clinical information of the PPA participants.

Participant	PPA variant	Age	Gender	Years from the first symptom, approximated	Classifier group
1	Lv	67	F	3	Train
2	Lv	71	F	5.5	Train
3	Lv	70	M	1	Train
4	Lv	62	F	5	Train
5	Lv	58	M	2	Train
6	Lv	67	F	1	Train
7	Lv	72	F	5	Train
8	Lv	73	F	2	Train
9	Lv	64	F	2	Test
10	Lv	79	F	7	Test
11	Lv	71	F	3	Test
12	Lv	66	F	10	Test
13	Lv	65	F	5.2	Test
14	Lv	70	F	4	Test
15	Lv	73	F	1.25	Test
16	Lv	51	M	2	Test
17	Lv	63	F	4.7	Test
18	Lv	71	M	2.2	Test
19	Sv	62	M	2	Train
20	Sv	73	M	10	Train
21	Sv	56	F	7.5	Train
22	Sv	62	M	1	Train
23	Sv	68	M	6	Train
24	Sv	60	F	1.5	Train
25	Sv	71	M	4	Test
26	Sv	68	F	1	Test
27	Sv	58	F	1	Test
28	Sv	74	F	9	Test
29	Sv	70	M	2	Test
30	Sv	57	M	1	Test
Summary					
	Lv train	67.5 ± 5.2	2M, 6F	3.1 ± 1.8	
	Lv test	67.3 ± 7.4	2M, 8F	4.1 ± 2.7	
	Sv train	63.5 ± 6.0	4M, 2F	4.6 ± 3.7	
	Sv test	66.3 ± 7.1	3M, 3F	3 ± 3.1	

To unlock the rich information stored in PACS, we also need to structurize the image data to support query, search, and relational analysis of anatomical information in the images. For example, physicians would submit an image of a new patient and search 100 images with a similar anatomical phenotype. The clinical reports related to these images would be retrieved and a statistical report of the diagnosis and prognosis could be generated. The technology for direct image searching, called content-based image retrieval (CBIR), is similar to that involved in face recognition (for review, see Müller, 2004), and is highly anticipated advancement in medical imaging for CT and MRI (Greenspan and Pinhas, 2007; Rahman et al., 2007; Robinson, 1996; Orphanoudakis et al., 1996; El-Kwae et al., 2000; Sinha et al., 2001; Unay et al., 2010). While CBIR is certainly promising, to date, very few image retrieval systems have been used for education or for routine clinical practice. Some difficulties in applying these established CBIR systems to the human brain are the complexity of the structures and the importance of location information.

We report our initial attempt to extend CBIR to three-dimensional feature extraction and structuration of the brain anatomy. The most obvious approach would be to align all the images in a database into a common template space. With a resolution of 1 mm^3, each image would be converted into a vector with about 1 million-voxel elements and vector-to-vector correlation analysis could be performed to find past similar cases. Alternatively, subtraction between a pair of images could be used to evaluate the similarity. These approaches, however, have several challenges. First, the number of observation (voxels) is too large and the information is too noisy. Second, each voxel does not carry anatomic notion by itself. Here, we assume that the conversion of the 1 million-voxel image representation needs to be contracted to a far smaller number of anatomical representation for subsequent information storage, searching, and correlation analyses. The recent advances in the structural parcellation technique of the entire brain (Shen and Davatzikos, 2002; Shi, 2010; Wang, 2013) now provides tools to structure raw MR images into a standardized and quantitative indices.

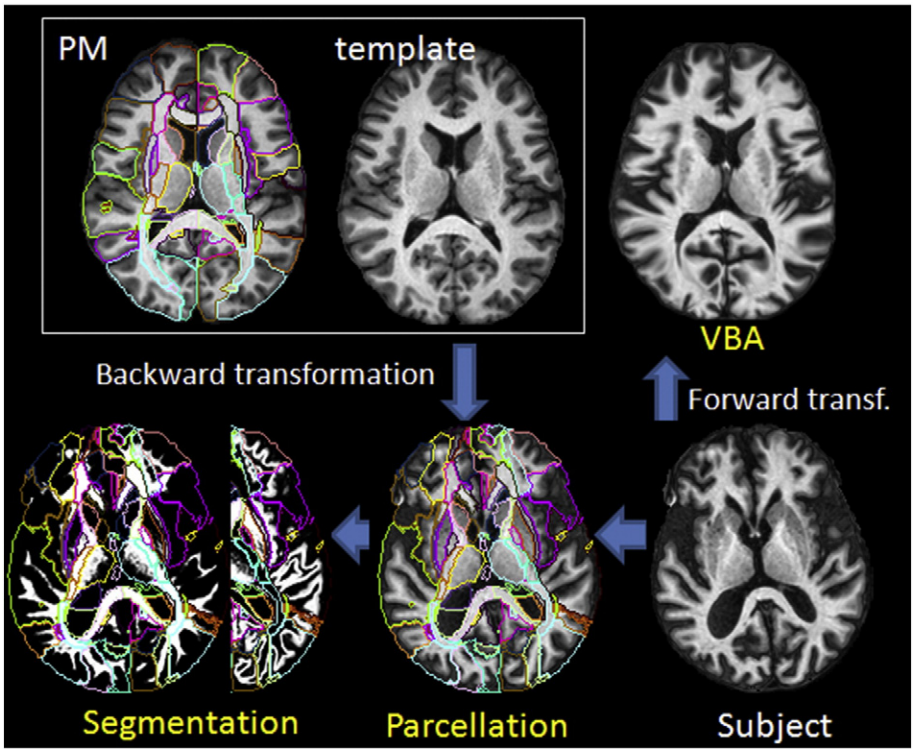

Fig. 1. Schematic representation of the normalization procedure. Each subject's brain is accurately normalized to the template, which was previously parcellated into regions that were anatomically predefined. Due to the reversible nature of the normalization algorithm, this parcellation map (PM) can be warped to each subject's brain, thus enabling automatic measurement of the volume (and contrast) of hundreds of areas in each individual. Thus, voxel-based analysis (VBA) and atlas-based analysis (ABA) can be performed on the same framework. The segmentation according to tissue type allows the segregation of CSF, white, and gray matter in the peripheral areas in the native space of each subject.

In this paper, we test the efficacy of the parcellation-based image structuration approach using T1-weighted images of a population with primary progressive aphasia (PPA) and age-matched controls. PPA is an ideal model because, while brain atrophy is a common feature, its extension and location vary. In addition, the atrophy is visible and the results can be compared with subjective human judgment. We tested whether the structured anatomical data actually captured the anatomical features that can be perceived by trained clinicians. Using principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA), we characterized the anatomical variability in PPA, tested our image search engine, and performed individual evaluations and integrative analyses of anatomical (image) and clinical (diagnosis) phenotypes.

2. Methods and materials

2.1. Participants

We enrolled 30 participants with PPA who were seen in one author's (AH's) outpatient clinic. Table 1 summarizes demographic and clinical information; the Appendix table contains the complete clinical information. These patients gave written informed consent and this study was approved by the local IRB. They were diagnosed with PPA on the basis of having a predominant and progressive deterioration in language in the absence of a major change in personality, behavior, or cognition (other than praxis) for at least 2 years. They were classified as Logopenic variant (Lv; $n = 19$, 4 male, age: 67.4 ± 6.4) or Semantic variant (Sv; 7 male, age: 65 ± 6.5), according to recent guidelines (Gorno-Tempini, 2011). Patients with nonfluent/agrammatic variant PPA were not included. Axial MPRAGE T1 − WIs (TR/TE = 8.4/3.9 ms) were acquired using a 3 T MRI scanner, with a 256 × 256 matrix. Sixteen participants were scanned with a field of view (FOV) of 230 × 230 mm and 120 slices of 1 mm thickness; fourteen were scanned with a FOV of 212 × 212 mm and 140 slices of 1.1 mm thickness. Healthy controls ($n = 24$) with same age, gender, and image protocol distribution were obtained from our normal database.

2.2. Automated parcellation

To measure the volume of each anatomical region, we performed an atlas-based analysis (ABA), in which an atlas image (Mori, 2008) with a pre-defined anatomical parcellation was warped to each participant's brain, thereby automatically parcellating each brain (Fig. 1). The process was performed by DiffeoMap and RoiEditor (Laboratory of Brain Anatomical MRI and Center for Imaging Science at Johns Hopkins University, 2014) and was made possible due to the high accuracy of the mapping algorithm, the large deformation diffeomorphic metric mapping, LDDMM (Miller, 2005; Wang, 2007). The dual-contrast LDDMM (Ceritoglu, 2009) was based on T1-WIs and cerebrospinal fluid (CSF) maps (Djamanakova, 2013). The template was the JHU-MNI "Eve," a single-subject atlas in the ICBM-152 space, extensively parcelled and labeled to 211 regions (Oishi, 2009). The 52 "coarse" peripheral regions (Oishi, 2008) were subsegmented in each native individual space to separate the cortex, the white matter beneath the cortex, and the cerebrospinal fluid (CSF), using tissue maps from SPM8, which resulted in the final 211 ROIs for each participant. This parcellation–segmentation

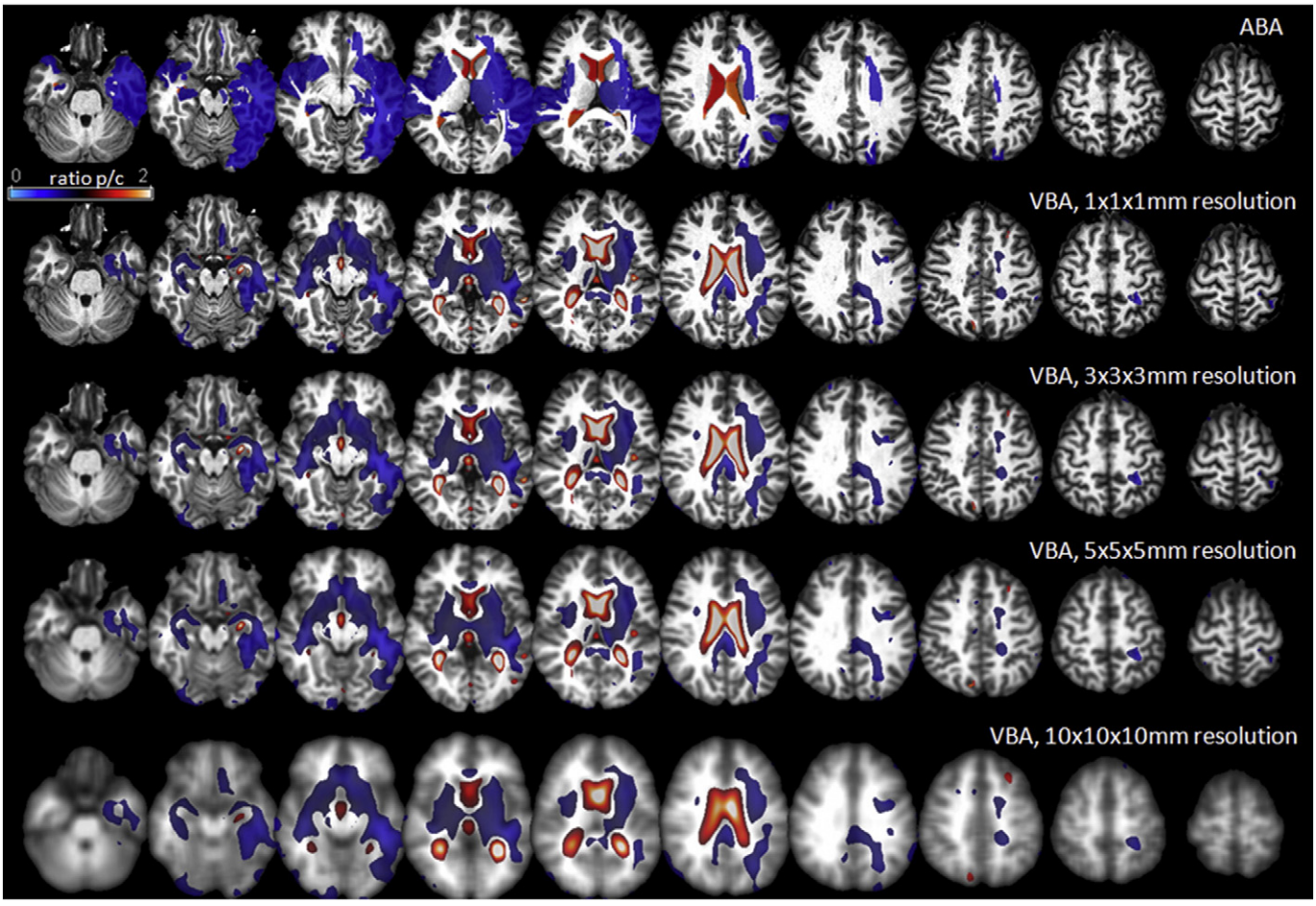

Fig. 2. Volumetric comparison between PPA patients and controls using ABA and VBA with different sizes of isotropic filters. Colored areas were significantly different between groups (t-test at p-value < 0.05 with correction for multiple comparison using FDR). The colors code the ratio of volume in patients/controls.

combined approach circumvents the mis-registration problems in the cortical areas, although mis-registration may exist between the parcels (e.g., in the boundary between adjacent gyri or between peripheral and deep white matter parcels).

In order to compare different degrees of granularity, we combined lobar parcels and deep white matter areas, and created a "low granularity" parcellation that contains 14 regions. Both the high and low granularity parcellations were tested against the visual analysis by 3

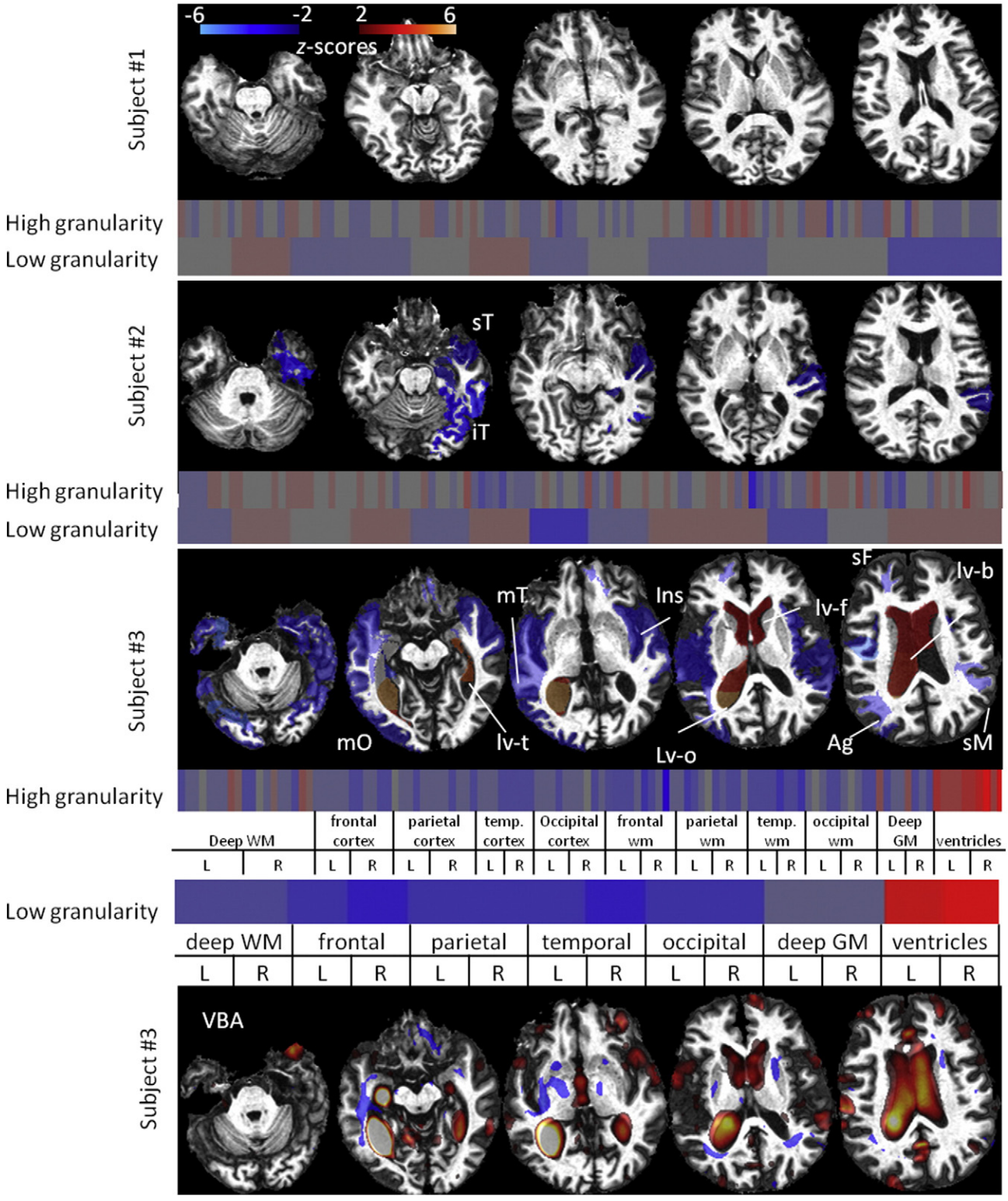

Fig. 3. Representation of degrees of regional atrophy as *z*-scores measured by ABA in patients with different anatomical phenotypes. The intensity plots of the regional degree of atrophy are represented at the bottom of the respective image, in two different degrees of granularity (high: 211 areas; low: 14 areas). The bottom row is shown for comparison with the *z*-scores obtained from VBA in one of the subjects, brought to this subject space. sT, mT, iT: superior, middle, and inferior temporal gyri, mO: middle occipital gyrus, Ag: angular gyrus, sM: supra marginal gyrus, ins: insula, sF: superior frontal gyrus, lv-b, lv-f, lv-o, lv-t: body, frontal horn, occipital horn, and temporal horn of the lateral ventricles.

experienced clinicians (two neuroradiologists — AVF and SY — and one neurologist — KO).

2.3. Data analysis

For the voxel-based analysis (VBA, Fig. 1) in the template space, we used the Jacobian determinant (i.e., the local expansion factor) of the LDDMM deformation fields as a quantitative metric of local volume changes (Chung, 2001; Riddle, 2004; Thompson, 2000). For the ABA, we used the volume of each parcel in the native space, normalized by the intracranial volume. For each PPA patient, regional *z*-scores were calculated using the average and standard deviations of the controls.

The principal component analysis (PCA) was used to explore the anatomical characterization. For image classification (Lv vs. Sv), and to exemplify how the image search engine works, we used PLS-DA. To guarantee classifier stability, we trained the model using participants scanned with one protocol ($n = 14$) and tested it with participants scanned with another protocol ($n = 16$). These groups had no other significant clinical or demographic differences other than the image protocol. We report the variables (i.e., the anatomic regions) with the highest weight to explain the population variance (in the PCA case) and characterize and segregate groups (in the PLS-DA case).

2.4. Correlation with visual assessments

To determine to what extent the visual analysis and the ABA agreed, we asked the neurologist / neuroradiologists to visit each parcel automatically defined in the native space of all the subjects and to classify the atrophy in degrees of severity (0 = no atrophy, 1 = mild, 2 = moderate, 3 = severe atrophy). The average scores of visual analysis in each region were correlated with the *z*-scores of volume obtained from the automated ABA, using a non-parametric Spearman test. In the same way, we calculated the inter-rater agreement. We also "dichotomized" the data (visual scores: 0 and 1 in class 1, and scores 2 and 3 in class 2; ABA *z*-scores: ≤2 were class 1 and >2 were class 2) and performed an accuracy analysis to measure the sensitivity and specificity of the ABA compared to the subjective visual impression.

3. Results

3.1. Population-based VBA and ABA comparison

Fig. 2 is a volumetric comparison of controls and PPA participants with VBA and ABA, by *t*-tests at a p-value < 0.05, corrected for multiple comparisons using the false discovery rate (FDR). Both methods detected atrophy in the white matter and the deep gray matter, as well as enlargement of the ventricles. However, there were several noticeable differences. First, the ABA examines the volumes structure-by-structure, while the VBA examines each voxel independently and, therefore, the report does not follow anatomical boundaries. Second, VBA results are largely silent in the cortical areas, probably due to mis-registration of voxels in the CSF, the cortex, and the peripheral white matter across patients. Isotropic filters do not ameliorate this fact. This is less of an issue for ABA because these three compartments are measured and averaged in the native space.

3.2. Individual-based analysis and comparison with visual examinations

Fig. 2 reports the results of a population-based analysis, in which data within a group are congregated and atrophic locations that are common among the patients are characterized. This analysis is, however, not compatible with clinical diagnosis, which requires information for each patient. A simple approach to report quantitative individual features is to use a *z*-score map of each patient (Fig. 3). The individual analysis based on VBA is, again, mostly silent in peripheral areas. Fig. 3 shows examples of individual ABA results for three PPA patients. ABA converts the native T1-WIs images into a 211- (high granularity) or 14- (low granularity) element *z*-score vector, which is visualized by a bar-code-type graph. The agreement of the subjective analysis (inter-evaluators) was higher using the low-granularity parcellation ($r = 0.81$) than with the high-granularity ($r = 0.54$). The agreement between the automated *z*-score and the averaged visual scores ($r = 0.8$) was virtually the same as the inter-evaluator agreement, showing that the automated method's performance can be considered as good as that of one of the raters. The agreement between visual and quantitative analysis improved at moderate and severe degrees of atrophy ($r = 0.68$ for visual scores >1, against $r = 0.22$ for visual scores <1, at high granularity). The sensitivity, specificity, and predictive values of evaluators and quantitative analysis were comparable (Table 2).

3.3. PCA-based analysis for population-based interpretation of the anatomy of individual patients

Fig. 4 shows the tridimensional PCA plot, where the principal components (PCs) were linear combinations of the volume of the parcels. The three PCs explained 47.5% of all anatomical variability in this cohort. The definition of the three groups (controls and two PPA variants) was based on clinical information, not on anatomical information. Although there is a natural segregation among the groups, it is not perfect. For example, a patient clinically diagnosed as Sv (indicated as "patient nearby controls") is "anatomically" close to the controls on the PCA plot. This is understandable if we examine the MRI, which seems to demonstrate a normal appearance. Similarly, controls "anatomically" located near the PPA patients (such as "control nearby patients") had enlarged ventricles and parenchyma atrophy. The images of six subjects numbered in Fig. 4 and shown in Fig. 5 were chosen based on their locations in the PCA space. Two neighboring cases in the PCA space (#1 and 2, #3 and 4, or #5 and 6) had visually similar anatomical features, while the cases distant from each other (#1, 3, and 5 or #2, 4, and 6) had markedly different anatomical features, indicating that the structured anatomical vectors captured the gross anatomical features.

Fig. 6 shows the areas with highest absolute weight in the first three PCs. The highest loading weights in the first component were widespread in tissue (negative), particularly at the left, and at the ventricles (positive), characterizing global tissue atrophy and ventricle enlargement. This component contributed to the segregation of controls from PPA patients. The other components were marked by high (absolute) weights at the deep gray matter and the core white matter (second), and at the frontal, parietal, and temporal lobes, particularly at right, in the latter (third). This reflected the common phenotype of PPA, as well as particular characteristics of each variant, such as the predominance of atrophy in the frontal and parietal areas in Lv, and at the left temporal in Sv (Wilson, 2009).

Differences in protocol contributed to 5% of the data variance, suggesting that the anatomical effect size in this population was large enough to be delineated beyond the protocol. Age contributed to 1% of the variance and gender, to 0.73%.

3.4. Evaluation of the searching results

The PLS-DA resulted in a model with an accuracy of 0.875 in distinguishing Lv from Sv (Fig. 7). In the first component, which had

Table 2
Inter-evaluator agreement in describing atrophy and agreement between the average evaluators' scores and automated qualitative *z*-scores.

	Inter-evaluators	Quantitative vs. evaluators
Sensitivity	0.74	0.87
Specificity	0.72	0.67
Positive predictive value	0.54	0.66
Negative predictive value	0.87	0.88

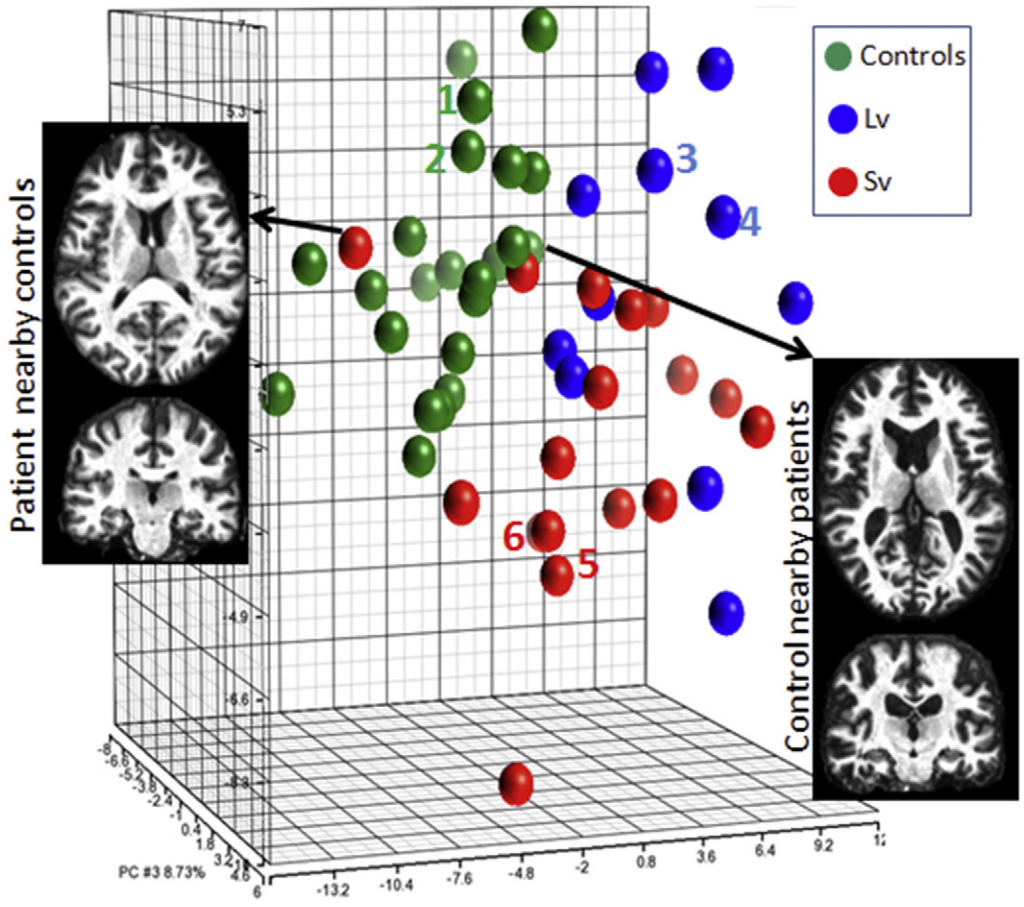

Fig. 4. PCA plot of the volume vectors. The colors represent clinical classifications. In detail, the images of two individuals that "fall," at the PCA space, far from the center of the group of his/her "clinical" phenotype, but close to individuals that share his/her "anatomical" phenotype.

the greatest power to segregate the groups, the loading (absolute) weights were higher in the frontal and parietal areas in Lv, and in the left temporal areas in Sv, which agreed with the pattern of atrophy described previously (Wilson, 2009). The second component, much less powerful in segregating the groups, revealed high loading weights in the frontal areas, the deep gray and white matter, in some temporal, parietal, and occipital left areas, and in the frontal horn of left ventricle, which is a common pattern of atrophy in this population.

4. Discussion

4.1. Advantages and disadvantages of the contraction of spatial information by ABA

We demonstrated the concept of converting images to a standardized vector to be used for the quantitative analysis of atrophy at population and individual levels. The accuracy of this conversion hinges on the

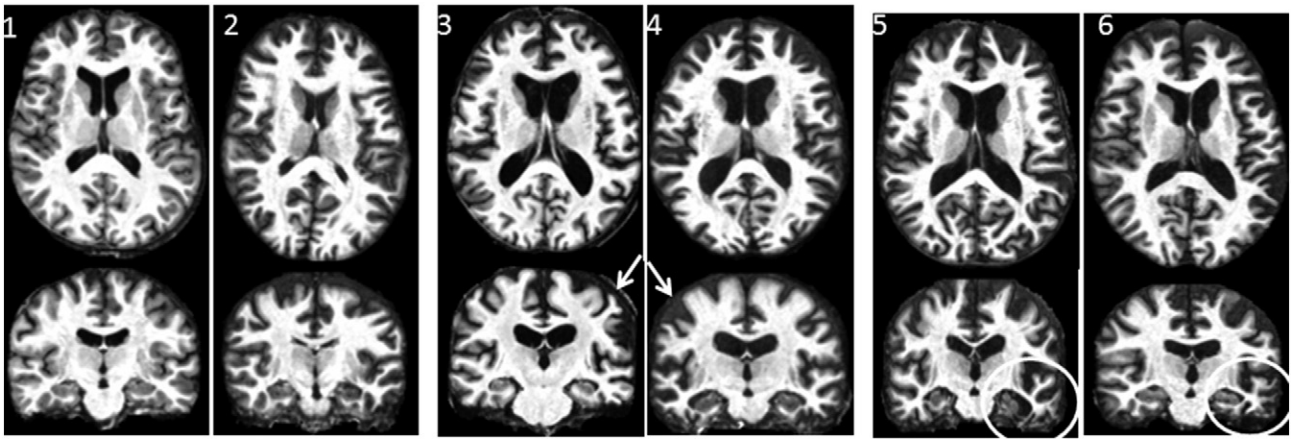

Fig. 5. These images correspond to the individuals numbered in Fig. 4. "Neighbors" in the PCA plot share a similar anatomical phenotype (#1 and #2 have normal appearance; #3 and #4 [Lv] have marked frontal and parietal atrophy (arrows); #5 and #6 [Sv] have global severe atrophy, particularly at the left temporal lobe (circles)), while individuals that are distant from each other have very distinctive anatomical characteristics.

accuracy of the image transformation and subsequent atlas warping. Our past publications reported a high level of accuracy using the LDDMM algorithm for populations with marked atrophy (Oishi, 2009; Djamanakova, 2013), although it depends on the structure in question. In addition, our healthy young adult atlas may not be the best template to achieve high transformation accuracy. More sophisticated approaches, such as multi-atlases (Aljabar, 2009; Heckemann, 2006; Iosifescu, 1997; Jia et al., 2012; Klein and Hirsch, 2005; Lao, 2004; Liu et al., 2004; Rohlfing et al., 2004; Wang, 2010; Warfield et al., 2004; Wu et al., 2007), may increase the accuracy in the future.

For the image-vector conversion, ABA plays an important role in dimensional reduction. The comparison between ABA and VBA shown in Fig. 2 has several relevant points for CBIR. Both approaches share the same image transformation scheme and, therefore, they also share the same level of transformation accuracy. The difference is the spatial filtering: in this study no filter was employed for VBA while one extra step to group voxels within pre-defined structures was employed for ABA. This makes the impact of filtering comparable. Assuming the brain volume is 1300 ml^3, there are 1.3 million pixels in an image with $1 \times 1 \times 1$ mm resolution. If these pixels are treated independently, the subsequent image structuration (Fig. 3) and PCA (Fig. 4) would become prohibitively difficult due to the sheer amount of information. Another important difference is the way peripheral areas (the CSF, the cortex, and the white matter beneath the cortex) are defined. Because of their large anatomical variability, a problem that is aggravated by the atrophy, we cannot expect accurate cross-subject registration (Diaz-de-Grenu, 2014), even with highly elastic transformation by LDDMM. The mis-registration leads to lower sensitivity for atrophy detection. In the ABA, the separation of these three compartments is performed in the native space of each subject (Fig. 1), reducing the effect of voxel-by-voxel mis-registration. Averaging of a large number of voxels in ABA may also increase the sensitivity to detect diffuse atrophy patterns, although it could be insensitive to atrophy confined in a region significantly smaller than a parcel.

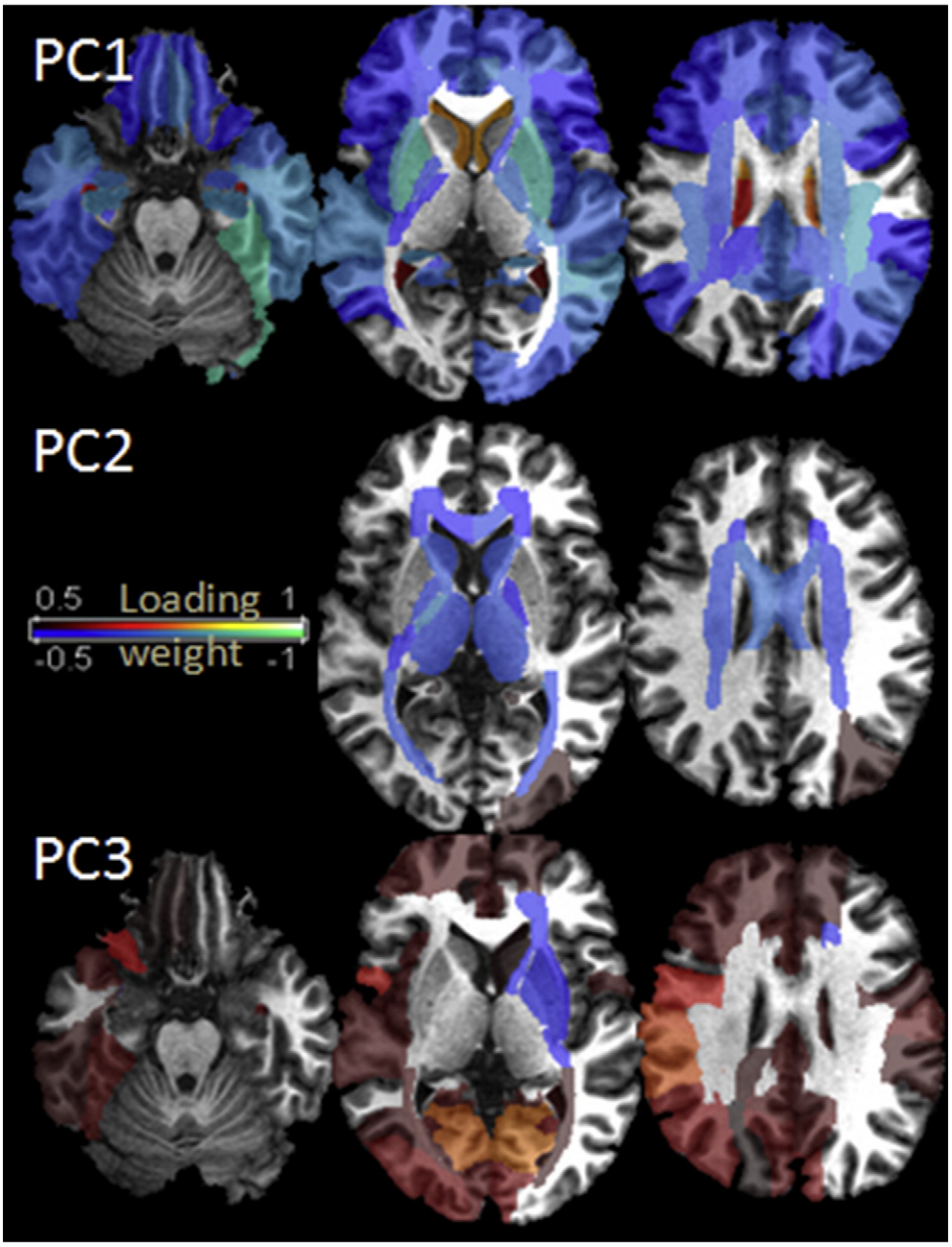

Fig. 6. Highest loading weights of the three first principal components. The first component characterizes global tissue atrophy and ventricle enlargement, and contributed most to the segregation of controls from PPA patients. The second and third components reveal, respectively, general and particular characteristics of PPA variants, such as global deep gray matter atrophy and atrophy at the fronto-parietal regions (Lv) and temporal left (Sv).

A drawback of ABA is that it requires pre-defined structural definitions. In this study, we used a parcellation based on classical anatomical definitions (Mori, 2008; Oishi, 2009). There are multiple other ways to parcellate the brain based on different criteria, for example, vascular, cytoarchitectonic, or functional maps. The employed parcellation may not optimally capture important pathological features; the best parcellation scheme depends on many factors, including the pathology in question.

Another difficulty is to define the ideal level of granularity (the number of parcels). If an abnormality exists in a confined area, the granularity must be high enough to detect it. On the other hand, if the abnormality is widespread and subtle, to use lower granularity is a better choice. It is known that the brain structures are based on hierarchical relationships and brain pathology could affect at the different levels of the hierarchy. For example, in a cohort of neurodegenerative diseases, various types of brain atrophy are known to exist depending on the patient pathology, such as left-dominant brain atrophy, focal temporal lobe atrophy, or hippocampal atrophy. As shown in Fig. 3, evaluation of the brain anatomy with multiple granularity levels could be an interesting approach to characterize atrophy patterns of neurodegenerative patients.

In this study, we were interested in the level of granularity in which trained radiologists evaluate images; it is certainly not voxel-by-voxel. Judged from the level of inter-rater agreement, the granularity level seems to be less than 211 parcels. At level of 14 lobar-based parcels, we found good agreement among raters as well as between the human and automated analysis. More comprehensive study with more granularity levels may be needed to further investigate how human examines images.

4.2. Evaluation of structured anatomical information by PCA

For conditions like PPA, which is known to contain heterogeneous pathological subtracts, the population-based analysis such as what

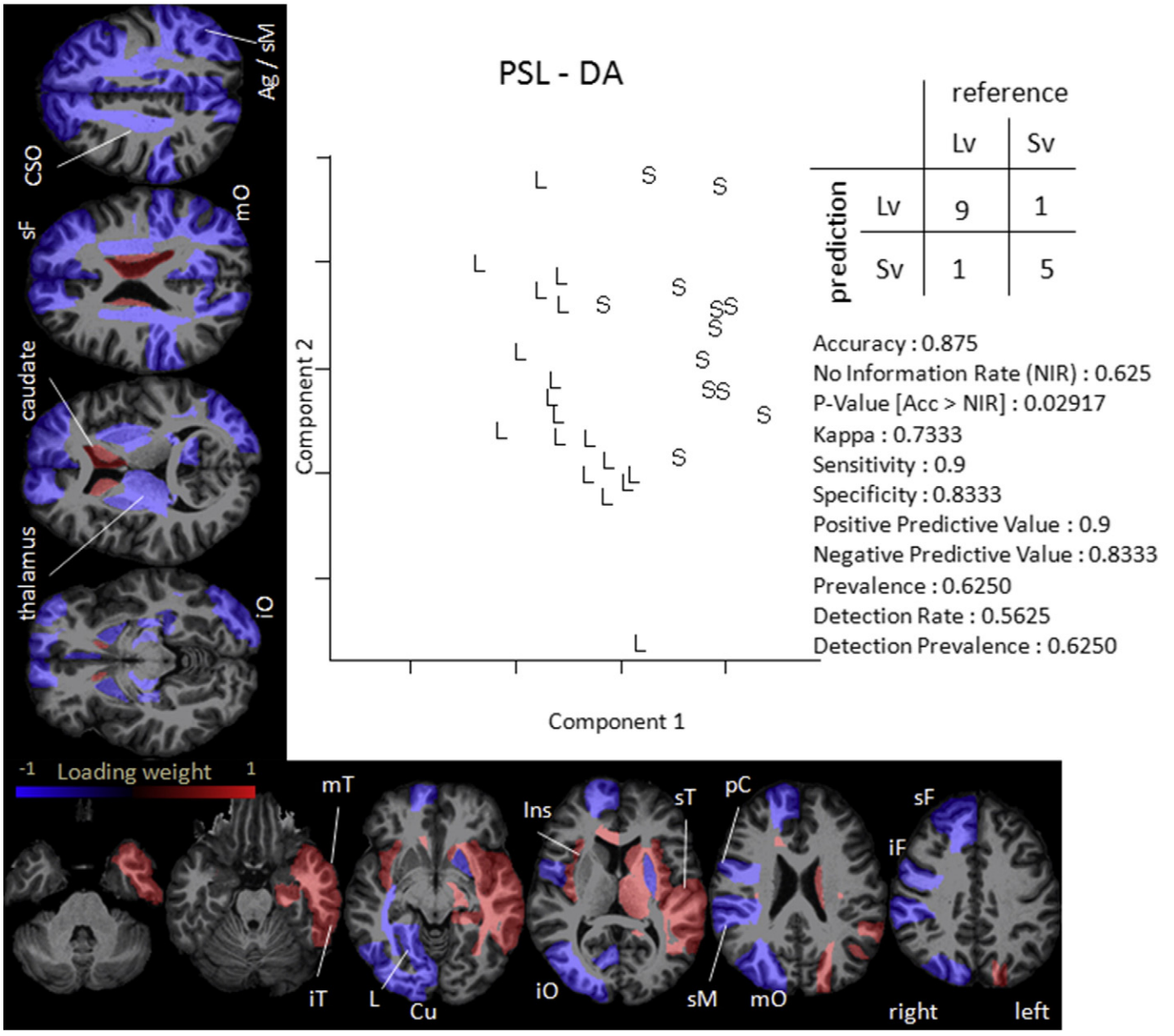

Fig. 7. PLS-DA. The classifier selected and combined anatomical features characteristics of each variant (components) in a model that achieved reasonable classification accuracy. sT, mT, iT: superior, middle, and inferior temporal gyri; iO, mO: inferior and middle occipital gyrus; Ag: angular gyrus; sM: supra marginal gyrus; ins: insula; L: lingual gyrus; Cu: cuneus; iF, sF: inferior and superior frontal gyrus, pC: pre-Central gyrus; CSO: centrum semi ovalle.

shown in Fig. 2 has a limited value. If we want to incorporate anatomical phenotype information for clinical use, alternative analytical frameworks, in which individual data are retained and not population-averaged, are needed. Using PCA, we explored the interpretation of individual images within a context of the population data. In addition, a concept of similarity search based on the feature vectors was demonstrated (Figs. 4). The distances among individuals could be used for similarity search. For example, suppose that patient # 5 in Fig. 2 is a patient of interest. Based on the nearest 10 cases in the PCA plot, one can conclude that this is probably an SvPPA patient because 70% of his "neighbors" are Sv, 20% are Lv, and 10% are controls. This type of image search can provide evidence-based clinical reports (Sinha et al., 2001); the PACS data can be searched purely on anatomical features, and statistical reports of associated clinical phenotypes can be provided, aiding the diagnosis and prognosis.

Of course, the searching approach described above considers only the anatomical phenotypes, which may not encode complete information about the pathology. One potential extension of the PCA is to add other dimensions of information, such as clinical data, to create classifiers. For example, using PLS-DA, we achieved a reasonable accuracy for classifying individuals according the PPA variants (Fig. 7). The idea of creating classifiers based on image features is not new and was successfully applied before to various neurodegenerative diseases, such as Alzheimer's, achieving a level of classification accuracy that suppress the human analysis (Cuingnet, 2011; Eskildsen, 2013; Fritzsche, 2010; Gray, 2013; Klöppel, 2008a,b). While the level of the classification accuracy is one of the hallmarks of the success, our main interest in this paper is to examine if the resultant anatomical features that would discriminate the patient populations actually make sense in the light of clinicians' accumulated knowledge, which they did (Gorno-Tempini, 2011, 2004; Mesulam, 2009, 2012). This indicates that the classifier extracts and smartly combines important anatomical features. These features can be easily translated, sought, and confirmed by the qualitative analysis, which is not always possible using VBA.

4.3. Limitations and perspectives

In the present study, we chose to use just the regional volumes, obtained from T1-WI analysis, as the image feature because the pathology in question causes brain atrophy primarily. However, this methodology is extensible to any other image contrast, such as T2-WIs, DTI derived-contrasts, SWIs, MRSI, or even to correlation matrices derived from resting state fMRI, since they share the same anatomical space as shown in a previous publication (Faria, 2012). The different image features, that can be contrasts such as FA, MD, or metabolite concentrations, or connectivity maps, can be used isolated or combined as a matrix of structures by image features that would characterize each individual. Then the same type of analysis can be performed and would be useful in cases where different domains of the anatomical phenotype affected (Zhang, 2011).

Finally, we do not expect perfect diagnostic accuracy with the discriminant analysis; first, as described above, we cannot assume that the anatomy contains enough information about the pathology. Second, our analysis does not include several critical information, such as clinical severity or distance from the onset. In addition, the diagnosis by clinicians, which is used as the gold standard for the classification, is known to contain errors and it is reasonable that the diagnoses change over time. In this context, the goal of this study was to maximize the usefulness of the anatomical information delineated by MRI by systematically integrating the anatomical and clinical information from past cases. We achieved this goal by demonstrating effective structuration of image data through image-vector conversion, which provides new opportunities to mine existing clinical databases for medical decision support.

This study is, however, still preliminary to judge the usefulness of the proposed approach in real clinical settings. To replicate the situations physicians routinely encounter in daily practices, a larger number of cases that contain more heterogeneous pathology are required. Another practical, yet, highly important question is if automated image analysis tools are robust enough to accurately analyze clinical data in PACS, which are often suboptimal in terms of image resolutions and contrasts. Although our tool has been tested for pediatric (Faria, 2010) and geriatric populations (Oishi, 2009; Djamanakova, 2013), protocol impacts have not been rigorously tested. Also, in order to determine the regional pattern of failure we need extensive manual delineation of hundreds of tridimensional structures in a large database. It is in our future scope to improve and test the robustness through technical advancement such as multiple atlas segmentation.

Supplementary data related to this article can be found online at http://doi.dx.org/10.1016/j.nicl.2015.01.008.

Acknowledgments

We thank all the patients and volunteers that agree to participate. We thank NIH (grants P41RR15241, RO1AG20012, and RO1NS058299 (SM), RO1HD065955 (KO), RO3EB014357 (AVF), R01 DC011317 and R01 DC 03681 (AH)) for financial support.

References

Aljabar, P., et al., 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage 46 (3), 726–738. http://dx.doi.org/10.1016/j.neuroimage.2009.02.01819245840.

Ceritoglu, C., et al., 2009. Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging. Neuroimage 47 (2), 618–627. http://dx.doi.org/10.1016/j.neuroimage.2009.04.05719398016.

Chung, M.K., et al., 2001. A unified statistical approach to deformation-based morphometry. NeuroImage 14 (3), 595–606. http://dx.doi.org/10.1006/nimg.2001.086211506533.

Cuingnet, R., et al., 2011. 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. http://dx.doi.org/10.1016/j.neuroimage.2010.06.01320542124.

Diaz-de-Grenu, L.Z., et al., 2014. A brief history of voxel-based grey matter analysis in Alzheimer's disease. J. Alzheimers Dis. 38 (3), 647–659. http://dx.doi.org/10.3233/JAD-13036224037033.

Djamanakova, A., et al., 2013. Diffeomorphic brain mapping based on T1-weighted images: improvement of registration accuracy by multichannel mapping. J. Magn. Reson. Imaging 37 (1), 76–84. http://dx.doi.org/10.1002/jmri.2379022972747.

El-Kwae, E.A., Xu, H., Kabuka, M.R., 2000. Content-based retrieval in picture archiving and communication systems. J. Digit. Imaging 13 (2), 70–81. http://dx.doi.org/10.1007/BF0316837110843252.

Eskildsen, S.F., et al., 2013. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. Neuroimage 65, 511–521. http://dx.doi.org/10.1016/j.neuroimage.2012.09.05823036450.

Faria, A.V., et al., 2010. Atlas-based analysis of neurodevelopment from infancy to adulthood using diffusion tensor imaging and applications for automated abnormality detection. NeuroImage 52 (2), 415–428. http://dx.doi.org/10.1016/j.neuroimage.2010.04.23820420929.

Faria, A.V., et al., 2012. Atlas-based analysis of resting-state functional connectivity: evaluation for reproducibility and multi-modal anatomy-function correlation studies. Neuroimage 61 (3), 613–621. http://dx.doi.org/10.1016/j.neuroimage.2012.03.07822498656.

Fritzsche, K.H., et al., 2010. Automated MR morphometry to predict Alzheimer's disease in mild cognitive impairment. Int. J. Comput. Assist. Radiol. Surg. 5 (6), 623–632. http://dx.doi.org/10.1007/s11548-010-0412-020440655.

Gorno-Tempini, M.L., et al., 2004. Cognition and anatomy in three variants of primary progressive aphasia. Ann. Neurol. 55 (3), 335–346. http://dx.doi.org/10.1002/ana.1082514991811.

Gorno-Tempini, M.L., et al., 2011. Classification of primary progressive aphasia and its variants. Neurology 76 (11), 1006–1014. http://dx.doi.org/10.1212/WNL.0b013e31821103e621325651.

Gray, K.R., et al., 2013. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. Neuroimage 65, 167–175. http://dx.doi.org/10.1016/j.neuroimage.2012.09.06523041336.

Greenspan, H., Pinhas, A.T., 2007. Medical image categorization and retrieval for PACS using the GMM-KL framework. IEEE Trans. Inf. Technol. Biomed. 11 (2), 190–202. http://dx.doi.org/10.1109/TITB.2006.87419117390989.

Heckemann, R.A., et al., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33 (1), 115–126. http://dx.doi.org/10.1016/j.neuroimage.2006.05.06116860573.

Hsu, W., et al., 2012. Context-based electronic health record: toward patient specific healthcare. IEEE Trans. Inf. Technol. Biomed. 16 (2), 228–234. http://dx.doi.org/10.1109/TITB.2012.218614922395637.

Hussein, R., et al., 2004. DICOM structured reporting: part 2. Problems and challenges in implementation for PACS workstations. RadioGraphics 24 (3), 897–909. http://dx.doi.org/10.1148/rg.24303572215143239.
Iosifescu, D.V., et al., 1997. An automated registration algorithm for measuring MRI subcortical brain structures. NeuroImage 6 (1), 13–25. http://dx.doi.org/10.1006/nimg.1997.02749245652.
Jia, H., Yap, P.T., Shen, D., 2012. Iterative multi-atlas-based multi-image segmentation with tree-based registration. Neuroimage 59 (1), 422–430. http://dx.doi.org/10.1016/j.neuroimage.2011.07.03621807102.
Klein, A., Hirsch, J., 2005. Mindboggle: a scatterbrained approach to automate brain labeling. NeuroImage 24 (2), 261–280. http://dx.doi.org/10.1016/j.neuroimage.2004.09.01615627570.
Klöppel, S., et al., 2008a. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. Brain 131 (11), 2969–2974. http://dx.doi.org/10.1093/brain/awn23918835868.
Klöppel, S., et al., 2008b. Automatic classification of MR scans in Alzheimer's disease. Brain 131 (3), 681–689. http://dx.doi.org/10.1093/brain/awm31918202106.
Laboratory of Brain Anatomical MRI Center for Imaging Science at Johns Hopkins University, 2014. MRI Studio: an image processing program. Available from: https://http://www.mristudio.org.
Lao, Z., et al., 2004. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. NeuroImage 21 (1), 46–57. http://dx.doi.org/10.1016/j.neuroimage.2003.09.02714741641.
Liu, T., Shen, D., Davatzikos, C., 2004. Deformable registration of cortical structures via hybrid volumetric and surface warping. NeuroImage 22 (4), 1790–1801. http://dx.doi.org/10.1016/j.neuroimage.2004.04.02015275935.
Mesulam, M., et al., 2009. Quantitative template for subtyping primary progressive aphasia. Arch. Neurol. 66 (12), 1545–1551. http://dx.doi.org/10.1001/archneurol.2009.28820008661.
Mesulam, M.M., et al., 2012. Quantitative classification of primary progressive aphasia at early and mild impairment stages. Brain 135 (5), 1537–1553. http://dx.doi.org/10.1093/brain/aws08022525158.
Miller, M.I., et al., 2005. Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. Proc. Natl. Acad. Sci. U S A 102 (27), 9685–9690. http://dx.doi.org/10.1073/pnas.050389210215980148.
Mori, S., et al., 2008. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. NeuroImage 40 (2), 570–582. http://dx.doi.org/10.1016/j.neuroimage.2007.12.03518255316.
Müller, H., et al., 2004. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. Int. J. Med. Inform. 73 (1), 1–23. http://dx.doi.org/10.1016/j.ijmedinf.2003.11.02415036075.
Oishi, K., et al., 2008. Human brain white matter atlas: identification and assignment of common anatomical structures in superficial white matter. NeuroImage 43 (3), 447–457. http://dx.doi.org/10.1016/j.neuroimage.2008.07.00918692144.
Oishi, K., et al., 2009. Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and Alzheimer's disease participants. Neuroimage 46 (2), 486–499. http://dx.doi.org/10.1016/j.neuroimage.2009.01.00219385016.
Orphanoudakis, S.C., Chronaki, C.E., Vamvaka, D., 1996. I2Cnet: content-based similarity search in geographically distributed repositories of medical images. Comput. Med. Imaging Graph. 20 (4), 193–207. http://dx.doi.org/10.1016/S0895-6111(96)00013-48954228.
Radiological Society of North America, 2014a. Radiological lexicon. Available from: https://http://www.rsna.org/RadLex.aspx.
Radiological Society of North America, Medical Image Research Center, 2014b. Available from: https://http://www.rsna.org/MIRC.aspx.
Rahman, M.M., Bhattacharya, P., Desai, B.C., 2007. A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. IEEE Trans. Inf. Technol. Biomed. 11 (1), 58–69. http://dx.doi.org/10.1109/TITB.2006.88436417249404.
Riddle, W.R., et al., 2004. Characterizing changes in MR images with color-coded Jacobians. Magn. Reson. Imaging 22 (6), 769–777. http://dx.doi.org/10.1016/j.mri.2004.01.07815234445.
Robinson, G.P., et al., 1996. Medical image collection indexing: shape-based retrieval using KD-trees. Comput. Med. Imaging Graph. 20 (4), 209–217. http://dx.doi.org/10.1016/S0895-6111(96)00014-68954229.
Rohlfing, T., Russakoff, D.B., Maurer Jr., C.R., 2004. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Trans. Med. Imaging 23 (8), 983–994. http://dx.doi.org/10.1109/TMI.2004.83080315338732.
Shen, D., Davatzikos, C., 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging 21 (11), 1421–1439. http://dx.doi.org/10.1109/TMI.2002.80311112575879.
Shi, F., et al., 2010. Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. NeuroImage 51 (2), 684–693. http://dx.doi.org/10.1016/j.neuroimage.2010.02.02520171290.
Sinha, U., Taira, R., Kangarloo, H., 2001. Structure localization in brain images: application to relevant image selection. Proc AMIA Symp 622–62611837219.
Thompson, P.M., et al., 2000. Growth patterns in the developing brain detected by using continuum mechanical tensor maps. Nature 404 (6774), 190–193. http://dx.doi.org/10.1038/3500459310724172.
Unay, D., Ekin, A., Jasinschi, R.S., 2010. Local structure-based region-of-interest retrieval in brain MR images. IEEE Trans. Inf. Technol. Biomed. 14 (4), 897–903. http://dx.doi.org/10.1109/TITB.2009.203815220064763.
Wang, L., et al., 2007. Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. IEEE Trans. Med. Imaging 26 (4), 462–470. http://dx.doi.org/10.1109/TMI.2005.85392317427733.
Wang, L., et al., 2013. 4D segmentation of brain MR images with constrained cortical thickness variation. PLOS One 8 (7), e64207. http://dx.doi.org/10.1371/journal.pone.006420723843934.
Wang, Q., et al., 2010. Groupwise registration based on hierarchical image clustering and atlas synthesis. Hum. Brain Mapp. 31 (8), 1128–1140. http://dx.doi.org/10.1002/hbm.2092320063349.
Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging 23 (7), 903–921. http://dx.doi.org/10.1109/TMI.2004.82835415250643.
Wilson, S.M., et al., 2009. Automated MRI-based classification of primary progressive aphasia variants. Neuroimage 47 (4), 1558–1567. http://dx.doi.org/10.1016/j.neuroimage.2009.05.08519501654.
Wu, G., Qi, F., Shen, D., 2007. Learning best features and deformation statistics for hierarchical registration of MR brain images. Inf Process Med Imaging 20, 160–17117633697.
Zhang, D., et al., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 55 (3), 856–867. http://dx.doi.org/10.1016/j.neuroimage.2011.01.00821236349.