Morphological classification of medical images using nonlinear support vector machines

Christos Davatzikos, Dinggang Shen, Zhiqiang Lao, Zhong Xue, Bilge Karacali
Section for Biomedical Image Analysis, Department of Radiology
University of Pennsylvania, Philadelphia, Pennsylvania
Correspond to: christos@rad.upenn.edu

Abstract

The wavelet decomposition of a high-dimensional shape transformation posed in a mass-preserving framework is used as a morphological signature of a brain image. Population differences with complex spatial patterns are then determined by applying a nonlinear support vector machine pattern classification method to the morphological signatures. By considering measurements from the entire image, and not only from isolated anatomical structures, and by using a highly non-linear classifier, this method has achieved very high classification results in a variety of tests.

INTRODUCTION

Morphological representations of anatomical images are often obtained using variants of high-dimensional shape transformations on a template that represents a "typical anatomy". Using these high-dimensional morphological representations in conjunction with machine learning techniques can result in powerful diagnostic tools. It can also assist in better understanding which morphological variables best reflect group differences, e.g. patients and normal controls. However, such an approach faces two fundamental difficulties. First, the high dimensionality of image data (in the extreme, one measurement per voxel can be obtained) makes classification techniques vulnerable to noise and can significantly reduce classification accuracy, as numerous studies in machine learning have shown. Second, anatomical differences or associations between anatomical and clinical parameters are typically nonlinear. As a hypothetical example, it is conceivable that similar rates of atrophy might put people with a small hippocampus at greater risk for cognitive impairment, compared to people that start with a larger hippocampus. Nonlinear effects of anatomical changes on cognitive measures are well known. Therefore, linear methods for classification are unlikely to be able to fully capture group differences.

In this paper, we describe a method that attempts to overcome both of the aforementioned limitations. By applying the wavelet transform to a tissue-preserving implementation of a shape transformation [1, 2], we reduce data dimensionality and organize the information in a hierarchical way, from a global and coarse to a local and fine scale. Moreover, by using a nonlinear support vector machine classifier [3], we attempt to capture nonlinear relationships between anatomical and clinical parameters.

METHODS

Shape transformation using HAMMER

In this paper, we adopt an approach referred to as Hierarchical Attribute Matching Mechanism for Elastic Registration (HAMMER), which was published in detail in [2] and briefly summarized here. HAMMER uses an *attribute vector*, i.e., a collection of attributes that reflect the anatomy around a particular voxel from a local to a global scale. If the attribute vector is rich and distinctive enough, it can differentiate between anatomically different points that have similar image intensities. Moreover, HAMMER uses a hierarchical deformation strategy, in which points with distinctive attribute vectors initially influence the warping process, followed by other points that have more ambiguous matches. This adds robustness and helps avoid local minima of the underlying cost function.

RAVENS mass-preserving framework

In principle, one can use the shape transformation that warps individuals to the template, in order to perform statistical analysis on morphological variables. This is certainly true when the shape transformation is extremely accurate and so can capture the finest details of an individual morphology. In practice, however, this is not necessarily the case. Shape transformations are subject to errors introduced by limitations of current deformable registration methods, and by the fact that inter-individual differences in brain morphology simply make it difficult, or in many cases impossible, to define anatomical correspondence. Such errors and ambiguities in the shape transformation can significantly affect subsequent statistical analysis.

In order to partly overcome this limitation, we have adopted the framework of mass-preserving shape transformations, described in detail in [1, 4, 5]. In this framework, we warp individual images to conform with a template, while the total amount of tissue in any arbitrarily defined region is preserved. This is accomplished by increasing or decreasing the density of the tissue whenever the shape transformation contracts or expands the tissue, respectively. This approach guarantees that the total tissue mass within the region is preserved.

The result of the mass-preserving shape transformation is three tissue density maps, one for gray matter (GM), one for white matter (WM), and one for cerebrospinal fluid (CSF). We collect the values of these

0-7803-8388-5/04/$20.00

three maps on all voxels within the brain into a long vector, which we call a brain morphological signature (BMS). We use the BMS to perform morphological classification in a high-dimensional space.

Support vector machine (SVM)-based classification.

SVM has emerged as one of the most powerful pattern classification methods during the past decade. A good reference for SVM classifiers is [3]. For the sake of completeness of this paper, we now summarize the basic principles of SVM. We first start with the linear case, which is simpler to describe.

Assume that we want to build a linear classifier that best separates two populations in a high-dimensional space. This classifier is described by a hyperplane whose position and orientation must be determined, with the help of a pre-classified training set. The optimal parameters of the dividing hyperplane are determined via an iterative constrained quadratic optimization scheme, in which the training samples of one group are forced to be on one side of the hyperplane and the samples of the other group are forced to be on the opposite side. This problem is solved via a variety of nonlinear programming techniques [3], which results in a number of "active" constraints, i.e., constraints that determine the solution. These constraints correspond to the samples that are very close to or are on the interface between the two groups. These are called "support vectors". The rest of the training samples do not contribute to the expression of the dividing hyperplane. This reveals a very important aspect of SVM, which is one of the reasons for its effectiveness as a classifier. The hyperplane is determined only by a relatively small number of samples that are close to the opposite group; the samples that are far away have no influence on the results, because it is clear as to which group they belong. The classifier inherently focuses on the subtleties of the morphological differences between the two groups and not on gross differences that are not difficult to detect, and is therefore more effective. In practice, however, it is impossible to prevent the two groups from overlapping to some degree. Therefore, the constraints are relaxed to permit some training samples to be on the wrong side of the hyperplane.

Linear classifiers have only limited power to separate groups, particularly when the statistical distributions in the high-dimensional feature space are complex. In our experiments we have verified that this is indeed the case with the kinds of problems that are being considered. Therefore, we use nonlinear SVM classification which is based on the same principle, but it is not based on a hyperplane that divides the two groups of samples, but rather on a more general hypersurface. This is accomplished by mapping the data to a high-dimensional space where the classification is achieved via a linear classifier as described earlier, and then by mapping the results back to the original feature space. This results in a non-planar hypersurface that adapts to an even greater degree to the subtleties of the interface between the two groups, and thus is more effective. Details about nonlinear SVM can be found in [3].

Classification using wavelet decomposition and feature reduction of the RAVENS maps.

Wavelet decomposition. The wavelet decomposition has been used successfully in many applications, including medical imaging. It is a very effective way of representing image information in a hierarchical way, and it offers the potential for significant reduction in the dimensionality of the data. We apply Daubechies wavelet decomposition on the original RAVENS maps only once for each RAVENS map image, thereby obtaining a scale-space representation of the volumetric information provided by these maps. A typical wavelet decomposition is shown in Fig. 1, where as we move from the top left to the bottom right we obtain more localized and higher frequency information from that initially collected from the RAVENS maps

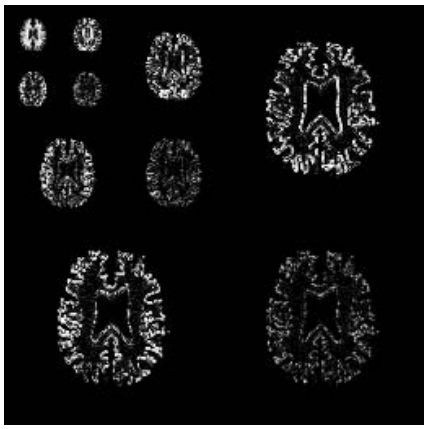

Fig. 1. A typical RAVENS tissue density map of the white matter, after wavelet decomposition. The top-left image is the low frequency global image content, whereas other images show increasingly localized anatomic detail.

Feature reduction. The wavelet decomposition reorganizes the information provided by the RAVENS maps in a hierarchical way, from relatively global and low-frequency information to relatively localized and high-frequency information. However, it does not reduce the dimensionality of the data. Therefore, applying a classification algorithm on the wavelet decomposition would also be sensitive to noise. However, due to its hierarchical nature, the wavelet decomposition of the RAVENS maps offers the possibility for significant data reduction. We achieve this via an algorithm previously described [6]. In particular, a standard variance-based feature discrimination technique, such as between-to within-class variance ratio [7], is first employed to define the discrimination measure for each wavelet feature as shown in Equation. 1.

$$Q(C_i \quad , \quad C_j) = \frac{\eta(\sigma(C_i) + \sigma(C_j))}{|m(C_i) - m(C_j)|} \qquad (1)$$

where $\eta = 3.0$, $m(C_i)$ and $m(C_j)$ are means for class C_i and C_j, and $\sigma(C_i)$ and $\sigma(C_j)$ are standard deviations of classes C_i and C_j.

The wavelet features are then ranked according to their discrimination measures. Finally, a collection of the most pertinent features are selected for image classification. This feature reduction method is simple and computationally very efficient, although not necessarily mathematically optimal, strictly speaking. The strictly optimal way of data reduction would be by iteratively selecting different sets of features and testing the SVM classification rate for each of them, until the best feature set is found. This approach, however, would clearly be computationally infeasible for our problem. In our implementation, a set of 2000 of the most pertinent features for each tissue usually yields satisfactory results.

Displaying Group Differences.

Although nonlinear classification methods are generally effective in resolving subtle and spatially complex group differences, they do not easily lend themselves to intuitive interpretation of the result, as opposed to statistical parametric maps of voxel-based analysis. This problem is largely due to the nonlinearity of the classifiers, which implies that group differences depend on the morphology itself, and they cannot be summarized with a single image. To further elucidate this issue, we construct a *hypothetical example*: evaluating the risk of developing a clinical condition, such as dementia, might depend not only on the rate of change of the hippocampus and the entorhinal cortex, but also on the size of these structures. For example, it could be that if the hippocampus is relatively small, then the rate of change might be a good predictor of risk of developing dementia, whereas if the hippocampus is large, other morphological characteristics might have higher predictive value. This means that one would need to display one image that reflects group differences for all possible brain morphologies. This is clearly not possible in practice.

A second difficulty is introduced by the dimensionality reduction (feature selection) that takes place before pattern classification, which is necessary for dealing with very high-dimensionality data, such as 3D images. For example, as we discussed earlier, we select about 2,000 features from the wavelet decomposition of each RAVENS map. Because of this feature selection process, we can no longer reconstruct the original brain, but only certain aspects of it that are represented by this limited set of variables.

In order to get around these problems and be able to display group differences in a way that is not only quantitative, but also suitable for visual interpretation of group differences determined quantitatively, we have developed the following procedure:

1. For every "support vector", i.e. for every brain that lies close to the hypersurface dividing two groups, we follow the gradient of the decision function until we reach the opposite side of the hypersurface, which is entirely in the second group. For example, in a male/female classification experiment, a fraction of the male brains will be the ones that influence the dividing hypersurface between the two groups. For each one of these brains, following the gradient of the decision function (which assigns brains to one group or the other) gives the fastest path that will "make a male brain look like a female brain, given the particular morphological characteristics of that specific male brain". This path varies from brain to brain, due to the nonlinearity of the classifier, as discussed above.

2. From the paths determined in Step 1, apply the inverse wavelet transform and construct images that highlight group differences for each of these brains.

3. Average these group difference images for all support vectors, so that a single map can be constructed.

4. Find all local maxima of the clusters that are formed in Step 3; these local maxima show the regions which are most informative in terms of resolving group differences.

5. Overlay these local maxima on the brain image used as template, for anatomical reference purposes.

Of course, Step 3 could be omitted. However, in that case one would need to show one difference map for every single brain.

EXPERIMENTS

In this section we provide experiments that demonstrate the performance of our approach using magnetic resonance images of healthy older adults who are participants in the Baltimore Longitudinal Study of Aging (BLSA). These individuals range in age from 56 to 85 and undergo yearly structural scans, among other evaluations, using a standard SPGR protocol. Details about the subjects and the image acquisition parameters can be found in [8].

Response in the absence of any effect: Experiment 1.

In order to test that there is no bias in our classifier, we tested the hypothesis that under no effect, the response should be detection of no group differences. Accordingly, we performed a random permutation experiment. Specifically, we randomly assigned 153 brain images from our database to two different groups. Using the leave-one-out method, we then tested our method. We repeated this experiment using 20 random permutations. The average classification rate was 48%, which agrees very well with our expectation that this experiment should lead to a nearly "coin-toss" success rate (which would be 50%).

Simulated atrophy

We performed two experiments in which we simulated morphological effects in two different ways on images from older adults, thereby generating a second set of the same number of images displaying systematic morphological differences in certain parts of the brain. Our goal was to test whether our classification method would correctly discriminate between these two groups, providing results that revealed brain differences in a systematic way. These

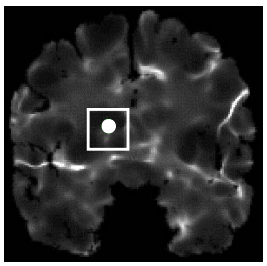

Fig. 2. The background shows a representative RAVENS map. The square was the area of simulated atrophy, thereby resulting in two groups that differed by a systematic reduction of their ravens maps. Applying our classification method determined the region highlighted with a white disk to be the most relevant region for distinguishing between the two groups, as one would expect.

experiments are described next.

Experiment 2. We introduced systematic atrophy on the RAVENS maps of 40 brain images by reducing the intensity values of the RAVENS maps in a cubic region centered on a manually selected voxel, as shown in Fig. 2. Since a RAVENS map uses a mass-preserving scheme via deforming the subject from its own space to the space of the template, it is then straightforward that reducing the intensity value with a certain region in the RAVENS map is equivalent to reducing the volume of the region's corresponding region in the subject space. We introduced atrophy of 10%, 20%, 30% and 40% within a cube of dimensions 19 × 19 × 19 mm^3, then a cube of 38 × 38 × 38 mm^3, and finally a cube of 57 × 57 × 57 mm^3. We used the leave-two-out method, and trained a classifier on 38 of these 40 images, then tested the classification result on the left out 2 subjects. We repeated this procedure, each time leaving two of the subjects out. Table 1 shows the resulting detection rates. The bright dot in Fig. 2 shows the region that was found as explained in the section "Displaying group differences".

mm^3	**10%**	**20%**	**30%**	**40%**
19×19×19	75%	80%	85%	90%
38×38×38	80%	82.5%	85%	92.5%
57×57×57	82.5%	87.5%	95%	100%

Table 1 Correct classification rates for Experiment 2. Different columns correspond to different levels of simulated atrophy. Different rows correspond to different spatial extents of simulated atrophy. As anticipated, increased level or extent of atrophy leads to increased separation of the two groups, and therefore to better classification.

Experiment 3. Experiment 2 tested classification performance under the assumption that a single region of atrophy is what separates the two groups. In reality, morphological differences between groups exhibit more complex spatial patterns. In order to test our classification scheme on a more realistic case scenario, we systematically introduced morphological differences to 10 brain images in 5 different regions: for the purposes of this simulation, the corpus callosum and the left and right lateral ventricles were expanded, while the right temporal lobe and the left hippocampus were contracted. We simulated 3 different levels of expansion/contraction, changing those portions of the brain in each group of ten by 5%, 10% and 15%, respectively. In other words, for each atrophy level we generated two groups of 10 subjects each, which provided pairs of images that were otherwise identical except for the morphologic differences in the 5 regions noted above. We then used the leave-two-out method by training our classifier on 18 subjects and testing it on the two left out, then repeated this procedure for all subjects. The resulting accuracy was 100% for all levels of atrophy, tested using the leave-two-out method as above.

Other experiments. We have also applied this methodology to studies of sex differences, aging, and schizophrenia, which have shown great promise. The results of these studies will be reported in separate papers.

CONCLUSION

We have presented a methodology for classification of anatomical images. A high-dimensional shape transformation is first used to align images with a standardized template. A tissue mass preserving framework is used to guarantee that volumetric measurements can be performed after the shape transformation, taking into account the fact that the latter changes the very anatomy being measured. Resulting are tissue density maps, with higher tissue density representing larger volume of the respective anatomical region and vice versa. Information in the density maps is organized using the wavelet transform and reduced via feature reduction methods. A nonlinear support vector machine classifier is then constructed from about 6,000 parameters maintained after feature reduction. The main goal of this method is to be used as an image-based diagnostic tool. Moreover, by examining trajectories that take support vectors into the opposite side of the dividing hypersurface, this method can elucidate the anatomical regions that are most important in differentiating between two groups.

REFERENCES

1. Davatzikos, C., A. Genc, D. Xu, and S.M. Resnick, "Voxel-Based Morphometry Using the RAVENS Maps: Methods and Validation Using Simulated Longitudinal Atrophy". *NeuroImage*, 2001. **14**: p. 1361-1369.
2. Shen, D. and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration". *IEEE Trans on Med. Imaging*, 2002. **21**(11): p. 1421-1439.
3. Burges, C.J.C., "A tutorial on support vector machines for pattern recognition". *Data Mining and Knowledge Discovery*, 1998. **2**(2): p. 121-167.
4. Goldszal, A.F., C. Davatzikos, D. Pham, M. Yan, et al., "An image processing protocol for the analysis of MR images from an elderly population". *J. Comp. Assist. Tomogr.*, 1998. **22**(5): p. 827-837.
5. Davatzikos, C., "Measuring biological shape using geometry-based shape transformations". *J. Image and Vision Comp*, 2001. **19**: p. 63-74.
6. Shen, D. and H.H.S. Ip, "Discriminative wavelet shape descriptors for invariant recognition of 2-D patterns". *Pattern Recognition*, 1999. **32**(2): p. 151-165.
7. Devijver, P.A. and J. Kittler, *Pattern recognition: a statistical approach*. 1982, Englewood Cliffs, NJ: Prentice-Hall.
8. Resnick, S.M., A.F. Goldszal, C. Davatzikos, S. Golski, et al., "One-year age changes in MRI brain volumes in older adults". *Cerebral Cortex*, 2000. **10**: p. 464-472.