Classification in DTI using Shapes of White Matter Tracts

Nagesh Adluru, Chris Hinrichs, Moo K. Chung, Jee-Eun Lee, Vikas Singh
Erin D. Bigler, Nicholas Lange, Janet E. Lainhart, Andrew L. Alexander

Abstract— **Diffusion Tensor Imaging (DTI) provides unique information about the underlying tissue structure of brain white matter in vivo, including both the geometry of fiber bundles as well as quantitative information about tissue properties as characterized by measures such as tensor orientation, anisotropy, and size. Our objective in this paper is to evaluate the utility of *shape representations* of white matter tracts extracted from DTI data for classification of clinically different population groups (here autistic vs control). As a first step, our algorithm extracts fiber bundles passing through approximately marked regions of interest on affinely aligned brain volumes. The subsequent analysis is entirely based on the geometric modeling of the extracted tracts. A key advantage of using such an abstraction is that it allows us to capture invariant features of brains allowing for efficient large sample size studies. We demonstrate that with the use of an appropriate representation of the tract shapes, classifiers can be built with reasonable prediction accuracies *without* making heavy use of the spatial normalization machinery needed when using voxel based features.**

I. INTRODUCTION

Neuroimaging researchers increasingly rely on diffusion tensor imaging (DTI) for new insights into the tissue microstructure and organization of brain white matter *in vivo*. DTI can provide information about the axon bundles of the white matter such as preferred orientation, information about local tissue structure using properties of tensors at each voxel. Using these tensors at each voxel fiber bundles are extracted using various techniques like solving stream-line equations [1], clustering [2], monte-carlo based methods [3]. Typically fiber bundles are traced from seed points in regions with anatomical interest. Most approaches to group analysis in the clinical DTI literature have relied on voxel based morphometry (VBM) like approaches which require computationally expensive non-linear diffeomorphic registration which can introduce artifcats in connectivity structures.

Recent advances have been made in using *tract* based morphometry (TBM) in which they abstract away from voxels and work in the space of streamlines [4], [5]. The main challenge with TBM as mentioned in their paper is the requirement of registering tracts across subjects with *non-linear* deformation that might change the actual anatomical significances of the tracts. Using *shape representations* of tracts can alleviate some of the problems in TBM. [6] presented a *shape* based normalization and provided evidence for tangible improvements in statistical power. Tract shape

Corresponding author: N. Adluru (adluru@wisc.edu). E. D. Bigler is with Dept. of Psychology, Brigham Young Univ. N. Lange is with Depts. of Psychiatry and Biostatistics, Harvard Univ., J. E. Lainhart is with Dept. of Psychiatry, Univ. of Utah. The rest of the authors are with the Univ. of Wisconsin, Madison.

modeling has also been used to segment out corresponding region specific tracts across subjects *without* requiring "registration" of seed points used in tractography [7]. All these studies indicate the power of using *shape representations* of tracts for DTI group analysis. Such geometric modeling of tracts has been argued to be able to account for errors in normalization [8]. Such techniques also offer a moderate level of invariance to quality variations in the low level data (e.g., different scanning parameters) common in multi-center studies. This is indeed attractive as it can open doors to analyze larger datasets (possibly acquired on different scanners) thus increasing the generalizability of statistical conclusions. Partly motivated by these new results, in this paper we seek to make use of the tractography data in the setting of supervised learning in building a two-class predictor where we address the following question:

1) Can the shapes of tracts be used to classify autistic subjects from controls *without* requiring tract-tract registration?

To study this question we use white matter tracts seeded in the splenium of the corpus callosum. Corpus callosum has been found to be important in connection with autism ([9], [10]) and pathways in this region have a well known anatomy [7]. Using support vector machine (SVM) as a binary classifier and histogram based shape representation of the tracts we are able to obtain reasonable prediction accuracy of upto 75% with cross-validation. A schematic overview of our classification system is shown in Fig. 1. The subjects are first aligned using simple affine transformations. The fiber bundles passing through seeds identified in the splenium (see Fig. 2) are then aligned using the affine parameters. We build histogram representations of these fiber bundles using $3D$ version of shape context [11]. To avoid over-fitting and improve classification accuracies non-linear principal component analysis (PCA) ([12]) with histogram intersection kernel [13] is used to project the bins into a low-dimensional feature space. These projections are then input to SVM as feature vectors for binary classification.

II. METHODOLOGY

A. Data Pre-processing

The DTI data from 73 male subjects were used in this study including

- 41 subjects with high functioning autism spectrum disorders.
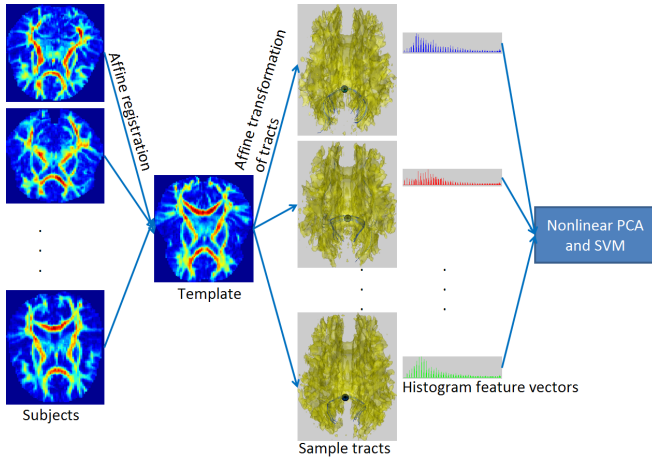- 32 control subjects matched for age, handedness, IQ, and head size.

Fig. 1. Schematic overview of the shape based classification system. Subjects are registered to a template using affine transformations. These affine transformations are used to align the tracts passing through the splenium. The tracts are represented using $3D$ shape context. These histogram based feature vectors are projected into a low dimensional space using non-linear PCA and classified using binary SVM.

Since brain shapes (and hence shapes of tracts) can be affected by factors like age, head size etc., the control subjects were matched for age, handedness, IQ and head size with those of subjects with autism.

Tractography in aligned space tends to suffer from compounding tensor interpolation errors. Hence tracts were extracted in the native space and were then aligned using affine transformations. A second order Runge-Kutta streamline algorithm with TENsor Deflection (TEND) [1] was used to extract the tracts. This gives more robust tracts compared to other streamline algorithms because it uses the entire tensor at each voxel to "deflect" the tracts in contrast to just using the major eigen vector. Fig. 2 shows some sample tracts passing through the splenium for two subjects. The alignment parameters were obtained as follows.
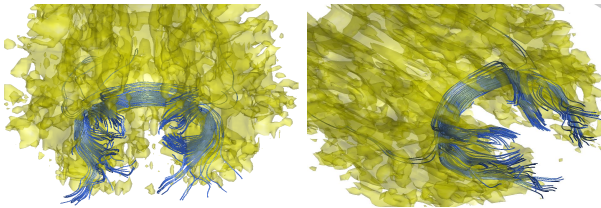

Fig. 2. Sample tracts (blue) passing through splenium of the corpus callosum for two subjects.

The DTI data from a 16 year old control subject was used as an initial template. The fractional anisotropy (FA) map for this subject was aligned to the MNI-152 white matter prior probability map using an affine transformation and mutual information based cost function with $2mm$ isotropic resolution over a $91 \times 109 \times 91$ matrix. The FA maps for the other 72 subjects were aligned to this single subject template using a 12-parameter affine transformation with FLIRT (http://www.fmrib.ox.ac.uk/fsl/). The aligned FA maps were then averaged to create an FA

template. The FA maps for each subject were again aligned to this template using affine transformation. This secondary normalization step reduced the potential bias issues of using a single subject as a template.

After briefly describing the classification framework in the following section, we explain the feature extraction.

B. Binary Support Vector Classifier

Support vector machines [14] have demonstrated empirical superiority with many theoretical guarantees on their generalization capacity among pattern classification algorithms. Given labeled training data of the form $\{(y_i, \mathbf{x}_i)\}_{i=1}^{N}$, with $y_i \in \{-1, +1\}$, $\mathbf{x}_i \in \mathcal{R}^n$ the algorithm finds a hyperplane $\mathbf{w} \cdot \mathbf{x} - b = 0$ by minimizing classification error while maximizing the margin using "boundary examples" or *support vectors* as:

$$\operatorname*{argmin}_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i \quad (1)$$

such that: $y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 - \xi_i, \ \forall i = \{1, 2, \dots, N\}$

where $\mathbf{w}$ is the normal vector, $\frac{b}{\|w\|}$ is the offset from origin and ξ_i are the slack variables. Using Langrange multipliers the following dual formulation is maximized so that the actually classification can be done by computing dot products with only support vectors:

$$\operatorname*{argmax}_{\alpha} W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{ij} \alpha_i\alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (2)$$

subject to: $0 \leq \alpha_i \leq C$ and $\sum \alpha_i y_i = 0$

where $C(> 0)$ is the tradeoff between regularization and constraint violation. Once we obtain the optimal alpha values (α) the decision function is $sign(h(\mathbf{x}))$, where:

$$h(\mathbf{x}) = \sum_{l=1}^{m} \alpha_l y_l \mathbf{x}^T \mathbf{x}_l + b \quad (3)$$

where $\{\mathbf{x}_l\}_{l=1}^{m}$ is the set of support vectors. One of the key design aspects in using SVMs is good feature extraction, i.e. effective representation of $\mathbf{x}$ for the data. It is even more crucial with medial data where the number of training examples is usually orders of magnitude low compared to typical machine learning tasks. The goal is to keep the dimensionality of $\mathbf{x}$ low while retaining the most useful information to separate the two classes. Hence we use robust histogram based representation of the tract data. To keep the dimensionality low we perform non-linear principal component analysis using histogram intersection kernel [13]. The following two sections describe the histogram representation and non-linear PCA respectively.

C. 3D Shape Context

Typically features used for classification tasks in brain imaging have been raw voxel-level measurements. For each subject all voxels with their values are used as features with some simple preprocessing (like down-sampling [15]) to more complex processing like segmentation using discrimination and robustness measures (DRM) in [16]. Spatial
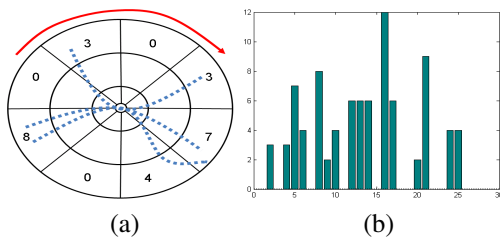
Fig. 3. Shape context. Demonstration in $2D$ for ease of understanding. (a) The log-polar binning used in $2D$ shape context [11]. Sample $2D$ streamlines are shown as dotted blue curves. Counts for outermost radial bins are also shown. By reading the bins in the order pointed by the red curve in (a) we obtain the histogram in (b).

Fig. 4. (a) Sample tracts passing through the splenium. A sample $2D$ version of binning-frame is centered at the splenium. (b) The histogram representation for the tracts.

consistency cues are emphasized in a typical feature selection process. They have been used both explicitly and implicity [17]. All of such approaches assume a nearly perfect normalization step since the feature values do not encode any high-level information.

The motivation for our feature extraction is to use the progress made in streamline tractography. By using tracts we are in essense using *anisotropic* spatial prior in extracting features from DTI data. As mentioned earlier we need good shape representation of tracts that can capture the differences between the two classes. Since we do not expect the subjects to be perfectly aligned, for each subject we build a "context" of the tracts passing through the splenium (see Fig. 2). This is based on the $3D$ extension of $2D$ shape context, a very successful descriptor used for shape matching for object recognition [11]. We divide the $3D$ volume into azimuthal-log-polar bins and count the number of tract control points in each bin. Fig. 3 illustrates the idea in $2D$ for ease of understanding. Fig. 4 (a) shows some sample tracts of a subject with a sample $2D$ version of binning-frame centered at the splenium. Fig. 4 (b) shows the histogram representation for the tracts of that subject. Using medium resolutions ($10°$ bins for azimuthal and polar angles and about 20 bins for logspace radii) helps us cope with slight misalignments since the histograms capture tract control point distributions of tracts which are high-level features as opposed voxel-level features. Such geometric modeling of tracts has been argued to be able to account for errors in normalization [8]. There they use "continuous medial representation (cm-rep)" of the white matter tracts for performing tract based spatial statistics (TBSS) [18]. Histogram based shape modeling offers attractive advantages both in robustness of matching and also computational efficiency [13]. As described in the following section histogram based features allow using histogram intersection kernel to perform efficient non-linear principal component analysis. To our best knowledge, ours is the first effort to study such histogram based modeling of DTI feature extraction in a *supervised* classification framework.

D. Non-linear PCA using Intersection Kernel

The dimensionality (n) of the features ($\mathbf{x}$) used for a classification influences the complexity of the learnt classifier. The higher the dimensionality the more complex a classifier can be. Even thoug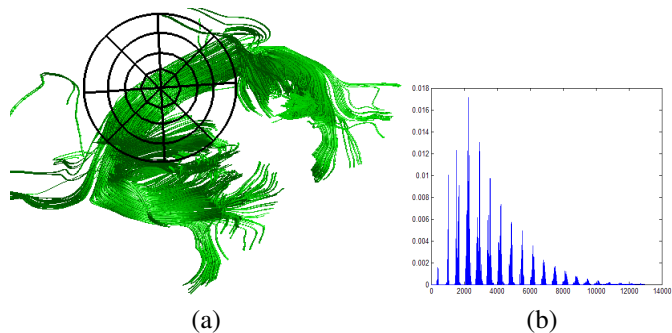h complex classifiers can provide higher classification accuracies they tend to over-fit the data and lose the generalization ability, indicated by decreased prediction accuracies with cross validation. This is referred to as "curse of dimensionality" which states for a given sample size of training data there is a maximum number of features above which the performance of the classifier decreases. A practical heuristic is to upper bound the dimensionality as $n \leq \frac{N}{10}$ where N is the number of training samples. There are several ways to reduce the dimensionality of the feature space. By using $10°$ bins for azimuthal and polar angles and about 20 bins for logspace radii our histogram representation has $12,960$ bins.

Principal component analysis (PCA) is a dimensionality reduction technique used to select the "basis" features by finding most uncorrelated features. The set of basis features can be obtained by solving for eigen values of the covariance matrix of the data $\{\mathbf{x}_i\}_{i=1}^N$[19]. That is one has to solve:

$$\lambda \mathbf{v} = \mathcal{C}\mathbf{v}, \text{ where } \mathcal{C} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T \qquad (4)$$

where $\mathbf{v}$ is the matrix of eigen vectors and λ is the diagonal matrix with eigen values. By selecting the top L largest eigen values the dimensionality can be reduced to L. In some cases where the correlation between features might be high, the above described linear PCA is not powerful enough to reduce the dimensionality. There are several non-linear extensions to perform component analysis like using Hebbian networks, multi-layer perceptrons etc. [20]. But many of those techniques are quite slow. [12] introduced non-linear extension of PCA using "kernel trick" where they replace the dot product in Eq. (4) with a kernel that can simulate dot product in a high-dimensional mapping of the data. The additional technical details of centering the data points before computing the kernel matrix can be found in [12]. We use histogram intersection kernel to compute the kernel matrix. The intersection kernel between two data points is given as:

$$k(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \min(x(i), z(i)) \qquad (5)$$

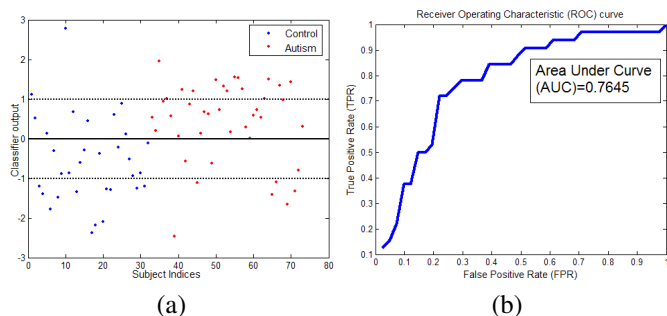Using top 6 eigen values we project most histograms onto a 6 dimensional feature space.

Fig. 5. (a) Classifier output ($h(\mathbf{x}) - b$, Eq. (3)) values for the two classes. The thick line is the classification boundary and the dotted lines are the margins. Values above the thick line indicate autism and those below indicate control subjects. Examples inside the margins are harder examples. (b) ROC curve shows that our classifier can perform reasonably well with an area under curve (AUC) of 0.7645. Average specificity and sensitivity is 71.88%. The values are estimated using leave-one-out cross validation.

III. EXPERIMENTAL EVALUATION

Since we have only 73 examples we evaluate our classifier performance using cross-validation scheme. In a k fold cross-validation, the dataset is partitioned into k subsets and each time $k-1$ samples are used for training and 1 for testing. The process is iterated k times until all the examples are tested once. The prediction accuracies over test subsets are then averaged to report the overall performance of the classifier. The measures reported using cross-validation scheme usually have more statistical significance. We perform leave-one-out cross-validation where $k = 73$. That is the evaluation is done 73 times each time using 72 examples for training and 1 example for testing until all the examples are tested. The average accuracy over 73 folds is 75.34% with specificity and sensitivity of 71.88%. The classifier output values and the corresponding receiver operating characteristic (ROC) curve are shown in Fig. 5. The average area under curve (AUC) is 0.7645.

IV. DISCUSSIONS AND FUTURE WORK

Applying machine learning techinques for brain imaging is attracting a significant interest of researchers [21]. Predictors can assist doctors and radiologists to diagonise diseases cost-effectively and efficiently. This paper presents an effort to apply tractography data of DTI to build a binary classifier with resonable prediction accuracy. Tractography data is usually based on anatomical regions of interest. We intend to include tractography data from other anatomically relevant regions for autism like temporal lobe (arcuate fasciculus) to achieve higher accuracies. Autistic patients have problems in understanding language and arcuate fasciculus is thought to be having pathways between regions involved in language understanding. We also intend to work with larger datasets since shape based representations can permit using data from multiple labs more effectively.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Lazar, D. Weinstein, J. Tsuruda, K. Hasan, and et. al., "White matter tractography using diffusion tensor deflection," *Human Brain Mapping*, vol. 18, no. 4, 2003.

[2] P. Savadjiev, J. S. Campbell, G. B. Pike, and K. Siddiqi, "Streamline flows for white matter fibre pathway segmentation in diffusion MRI," in *MICCAI*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 135–143.

[3] M. D. King, D. G. Gadiana, and C. A. Clarka, "A random effects modelling approach to the crossing-fibre problem in tractography," *NeuroImage*, vol. 44, no. 3, pp. 753–768, February 2009.

[4] L. O'Donnel, C.-F. Westin, and A. Golby, "Tract-based morphometry for white matter group analysis," *NeuroImage*, vol. 45, pp. 832–844, 2009.

[5] C. B. Goodlett, T. P. Fletcher, J. H. Gilmore, and G. Gerig, "Group analysis of DTI fiber tract statistics with application to neurodevelopment," *NeuroImage*, vol. 45, no. 1, 2009.

[6] H. Sun, P. Yushkevich, H. Zhang, P. Cook, J. Duda, T. Simon, and J. C. Gee, "Shape-based normalization of the corpus callosum for DTI connectivity analysis," *TMI*, vol. 26, 2007.

[7] M. Bastin, J. Piatkowski, A. Storkey, L. Brown, A. MacLullich, and J. Clayden, "Tract shape modelling provides evidence of topological change in corpus callosum genu during normal ageing," *NeuroImage*, vol. 43, pp. 20–28, 2008.

[8] P. Yushkevich, H. Zhang, T. Simon, and J. Gee, "Structure-specific statistical mapping of white matter tracts using the continuous medial representation," in *MMBIA07*, 2007.

[9] J. E. Lee, D. Hsu, A. Alexander, M. Lazar, E. Bigler, and J. Lainhart, "A study of underconnectivity in autism using DTI: W-matrix tractography," in *ISMRM*, 2008.

[10] X. Ke, S. Hong, T. Tang, H. Huang, B. Zou, H. Li, Y. Hang, and Z. Lu, "An analysis method of the fiber tractography of corpus callosum in autism based on DTI data," in *IEEE ICBBE*, 2008, pp. 2409–2413.

[11] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, 2002.

[12] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[13] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *CVPR*, 2008, pp. 1–8.

[14] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, 1995.

[15] P. Vemuri, J. Gunter, M. Senjem, J. Whitwell, and et. al., "AD diagnosis in individual subjects using struct. MR images," *Neu.Im.*, 2008.

[16] Y. Fan, D. Shen, R. Gur, R. Gur, and C. Davatzikos, "COMPARE: Classification of morphological patterns using adaptive regional elements," *Tran. Med. Imag.*, 2007.

[17] V. Singh, L. Mukherjee, and M. Chung, "Cortical surface thickness as a classifier: Boosting for autism classification," in *MICCAI*, 2008, pp. 999–1007.

[18] S. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, and et. al., "Nonlinear component analysis as a kernel eigenvalue problem," *Neu. Img.*, vol. 31, pp. 1487–1505, 2006.

[19] I. Jolliffe, *Principal Component Analysis*. Springer, 2002.

[20] K. Diamantaras and S. Kung, *Principal component neural networks: theory and applications*. New York, NY, USA: John Wiley & Sons, Inc., 1996.

[21] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: A tutorial overview," *NeuroImage*, vol. 45, pp. S199–S209, March 2009.