

A NEW DEEP-LEARNING APPROACH FOR EARLY DETECTION OF SHAPE VARIATIONS IN AUTISM USING STRUCTURAL MRI

Marwa Ismail¹, Gregory Barnes², Matthew Nitzken¹, Andrew Switala¹,
Ahmed Shalaby¹, Ehsan Hosseini-Asl¹, Manuel Casanova³, Robert Keynton¹,
Ashraf Khalil⁴, and Ayman El-Baz¹

¹BioImaging Laboratory, Bioengineering Department, University of Louisville, USA.

²Department of Neurology, University of Louisville, USA.

³Departments of Pediatrics and Biomedical Sciences, University of South Carolina, USA.

⁴Department of Electrical and Computer Engineering, Abu Dhabi University, Abu Dhabi, UAE.

ABSTRACT

This paper introduces a novel shape-based computer-aided diagnosis (CAD) system using magnetic resonance (MR) brain images for autism diagnosis at different life stages. To improve the classification robustness, the system fuses the shape features extracted from the cerebral cortex (Cx) and cerebral white matter (CWM). Fusion is conducted based on the findings suggesting that Cx changes in autism are related to CWM abnormalities. The CAD system starts with segmenting Cx and CWM using a 3D joint model that combines intensity, shape, and spatial information. Then, Spherical Harmonic (SPHARM) is applied to the re-constructed meshes of Cx to derive 4 metrics for each mesh point; normal curvature, mean curvature, gaussian curvature, and Cx surface reconstruction error. To analyze the CWM shape, distance maps of its gyri are computed and three more shape features are extracted for these gyri. Finally, all the extracted shape features are fed to a multi-level deep network for feature fusion and diagnosis. The CAD system has been evaluated using subjects from the ABIDE database (8–12.8 years), achieving an accuracy of 93%, and from NDAR/Pitt database (16–51 years), achieving an accuracy of 97%. Also in order to show the capability of the system for early diagnosis, it has been tested on NDAR/IBIS database for infants, resulting in an accuracy of 85%. These initial results on the 3 databases hold the promise of efficient autism diagnosis.

Index Terms— Cx, CWM, SPHARM, Gyri, deep fusion

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neuro-developmental disorder that affects personal linguistic, behavioral, and social skills. Autism has many symptoms, most prominently, social impairment and repetitive behaviors [1].

In the literature, several imaging modalities have investigated the changes associated with autism such as structural MRI (sMRI), diffusion tensor imaging (DTI), and functional MRI (fMRI), [2, 3]. Many acquisition sites in the

US have been collecting sMRI data, such as NDAR/IBIS, NDAR/Pitt [4], and ABIDE [5], with some sites collecting data using other modalities such as DTI data [4]. This paper focuses on sMRI data, overviewing its literature. For example, enlargement in cerebral gray matter (CGM) was confirmed for autistic toddlers by the longitudinal study [6] that divided Cx into 34 regions to obtain lobe volume measures. A ROI-based volumetric (RBV) study [7] was conducted on infants at high risk from the NDAR/IBIS database in comparison to those of low risk. Volumetric measures were obtained using Fourier harmonics to parameterize the head contour, and no significant differences could be observed between groups. Ecker et al. [8] investigated changes in the brains of autistic adults using support vector machines (SVM) which classified subjects with ASD with a sensitivity of 88%, and compared the findings to those of RBV. A voxel-based approach was adopted [9] to investigate the brains of autistic children and found abnormalities in the posterior cingulate cortex. The RBV approaches are age and gender-sensitive and in need of age-correction coefficients. Another type of study, called “shape-based morphometry” (SBM), mainly addresses topological shape features. For example, the study conducted on autistic children [10] showed that the cortical thickness in autistic subjects decreased in the cerebral lobes. Another study [11] on adolescents obtained the cortical surface, and demonstrated an accelerated cortical thinning in the autistic brains in two areas of the left hemisphere. SBM approaches account for the brain’s inherent topological features, however most of the work at present uses the meshes generated from raw data without parameterization or alignment, which slows computations and hinders comparing information. **In addition, SBM studies on infants with autism are rarely found in the literature.** To overcome these limitations, this paper proposes an automated shape analysis approach for autism diagnosis. The motivation behind fusing CWM and Cx features is that recent studies suggest that neuropathological findings in regards to minicolumnopathy and dysplastic Cx changes magnify themselves by abnormal connections of CWM [12].

2. METHODS

This proposed CAD system starts with brain segmentation into CWM and Cx. Shape analysis is then performed on the reconstructed meshes, from which 8 features are extracted that are fused to classify autistic and control brains. The proposed method is illustrated in Fig. 1.

2.1. Segmentation of CWM and Cx from MR images

The framework is based on the 3D joint model in [13, 14] that integrates shape, intensity, and spatial information. An input brain image, \mathbf{g} , co-aligned to the training data base, and its map, \mathbf{m} , are described with a joint probability model: $P(\mathbf{g}, \mathbf{m}) = P(\mathbf{g}|\mathbf{m})P(\mathbf{m})$, where $P(\mathbf{m}) = P_{\text{sp}}(\mathbf{m})P_{\text{V}}(\mathbf{m})$. $P_{\text{sp}}(\mathbf{m})$ denotes a weighted shape prior, and $P_{\text{V}}(\mathbf{m})$ is a Gibbs probability distribution with potentials \mathbf{V} , [15].

Adaptive Shape Model, $P_{\text{sp}}(\mathbf{m})$: segmentation starts by creating an atlas from 15 data sets (not from test subjects), co-aligned by 3D affine transformations [16], and spanning different ages from both control and autistic subjects [13, 14].

First-Order Intensity Model, $P(\mathbf{g}|\mathbf{m})$: The first-order visual appearance of each brain label is modeled and approximated using the linear combinations of discrete Gaussians (LCDG) approach [13, 14].

MGRF Model With Higher-order Cliques, $P_{\text{V}}(\mathbf{m})$: In order to account for inhomogeneities, especially for infant scans where there is poor contrast between different tissue types, spatial information is considered [17] using a higher-order Markov Gibbs Random Field (MGRF) spatial model, which adds the families of the triple and quad cliques to the pairwise cliques. Let \mathbf{C}_a denote a family of a -order cliques of an interaction graph with nodes in the 3D lattice sites $p = (x, y, z)$ and edges connecting the interacting sites. Label interactions are modeled by a homogeneous MGRF over the 26-neighborhoods of voxels:

$$P_{\text{V}}(\mathbf{m}) = \frac{1}{Z_{\text{V}}} \exp \left(\sum_{a=1}^A \sum_{\mathbf{c} \in \mathbf{C}_a} V_a(\mathbf{m}(x, y, z) : (x, y, z) \in \mathbf{c}) \right) \quad (1)$$

where A clique families describe the geometry of the graph interactions, $\mathbf{V} = [V_a : \{0, \dots, L\} \rightarrow (-\infty, \infty) : a = 1, \dots, A]$ is a collection of Gibbs potential functions V_a for the families \mathbf{C}_a , and the partition function Z_{V} normalizes the probabilities. An initial region map \mathbf{m} allows for analytically approximating the maximum likelihood estimates of the potentials and computing the voxel-wise probabilities of the region labels. The second, third, and forth-order potentials for the region map label $m_{\mathbf{p}_i}$ are given by:

$$V_a(m_{\mathbf{p}_1}, m_{\mathbf{p}_2}) = \begin{cases} V_{2:a:\text{eq}} & \text{if } m_{\mathbf{p}_1} = m_{\mathbf{p}_2} \\ -V_{2:a:\text{eq}} & \text{otherwise} \end{cases} \quad (2)$$

where $V_{2:a:\text{eq}} = -V_{2:a:\text{ne}} = 4(F_{a:\text{eq}}(\mathbf{m}^\circ) - \frac{1}{2})$, and $\mathbf{F}(\mathbf{m}^\circ) = [\rho_a F_a(\mu_1, \dots, \mu_s | \mathbf{m}^\circ) : (\mu_1, \dots, \mu_s) \in \{0, \dots, L\}^s; a = 1, \dots, A]$ is the collection of scaled relative frequencies of co-occurrences of configurations (μ_1, \dots, μ_s) of the labels in the cliques of each family \mathbf{C}_a over a given training map \mathbf{m}° .

$$V_a(m_{\mathbf{p}_1}, m_{\mathbf{p}_2}, m_{\mathbf{p}_3}) = \begin{cases} V_{3:a:\text{eq}_3} & \text{if } m_{\mathbf{p}_1} = m_{\mathbf{p}_2} = m_{\mathbf{p}_3} \\ -V_{3:a:\text{eq}_3} & \text{otherwise} \end{cases} \quad (3)$$

where $V_{3:a:\text{eq}_3} = -V_{3:a:\text{eq}_2} = \frac{16}{3}(F_{a:\text{eq}_3}(\mathbf{m}^\circ) - \frac{1}{4})$

$$V_a(m_{\mathbf{p}_1}, m_{\mathbf{p}_2}, m_{\mathbf{p}_3}, m_{\mathbf{p}_4}) = \begin{cases} V_{4:a:\text{eq}_4} & \text{if 4 labels} \\ V_{4:a:\text{eq}_3} & \text{if 3 labels} \\ -(V_{4:a:\text{eq}_3} + V_{4:a:\text{eq}_4}) & \text{otherwise} \end{cases} \quad (4)$$

where

$$\begin{aligned} V_{4:a:\text{eq}_4} &= \lambda^* (F_{a:\text{eq}_4}(\mathbf{m}^\circ) - \frac{1}{8}) \\ V_{4:a:\text{eq}_3} &= \lambda^* (F_{a:\text{eq}_3}(\mathbf{m}^\circ) - \frac{1}{8}) \\ V_{4:a:\text{eq}_2} &= \lambda^* (F_{a:\text{eq}_2}(\mathbf{m}^\circ) - \frac{1}{8}) = -(V_{4:a:\text{eq}_4} + V_{4:a:\text{eq}_3}) \end{aligned}$$

and

$$\lambda^* = \frac{\sum_{a=1}^A \left((F_{a:\text{eq}_4}(\mathbf{m}^\circ) - \frac{1}{8})^2 + (F_{a:\text{eq}_3}(\mathbf{m}^\circ) - \frac{1}{2})^2 + (F_{a:\text{eq}_2}(\mathbf{m}^\circ) - \frac{3}{8})^2 \right)}{\sum_{a=1}^A \left(\frac{7}{64} (F_{a:\text{eq}_4}(\mathbf{m}^\circ) - \frac{1}{8})^2 + \frac{1}{4} (F_{a:\text{eq}_3}(\mathbf{m}^\circ) - \frac{1}{2})^2 + \frac{15}{64} (F_{a:\text{eq}_2}(\mathbf{m}^\circ) - \frac{3}{8})^2 \right)}$$

Finally, the region map \mathbf{m} is improved using Iterative Conditional Mode (ICM) algorithm [18] that maximizes the probabilities of the 3D joint model, Algorithm 1.

Algorithm 1 Steps for the proposed approach

1. Construct the shape database by co-aligning images.
 2. Align the input with the database. Choose the atlas subject with the highest similarity to be the reference.
 3. For each voxel in the input, get its shape prior probability by transforming each voxel to the reference domain.
 4. Initialize a 3D window and search inside it for voxels with corresponding grey level in all training sets. Create probabilities based on the labels from the results.
 5. Approximate $P(\mathbf{g})$ of the images using an LCDG with four dominant modes.
 6. Form region map \mathbf{m} using marginal estimated density and prior shape.
 7. Find Gibbs potentials for the MGRF model from \mathbf{m} .
 8. Improve \mathbf{m} using ICM algorithm [18].
-

2.2. Shape feature extraction

After segmenting Cx and CWM, shape analysis is conducted to obtain the CAD system. Four shape features have been extracted from the 4 cerebral lobes of Cx (frontal, occipital, parietal, and temporal) per hemisphere. Four other shape features from CWM have also been extracted, resulting in a total number of 64 features per scan (8 features per lobe per hemisphere). Details of shape features are outlined below.

Cx shape features: to extract the Cx shape features, accurate approximation of its 3D shape is required as well as having the ability to compare different brain subjects. One way to achieve so is SPHARM analysis [19] that approximates the cortex surface, which would allow for comparing correspondences as well as having faster computations. It also makes the reconstruction insensitive to data acquisition from different sites. In order to apply SPHARM, a mesh manifold is first generated from the segmented scans using a modified version of the TETGEN algorithm [20]. When constructing the meshes, there is a restriction on the number of nodes to

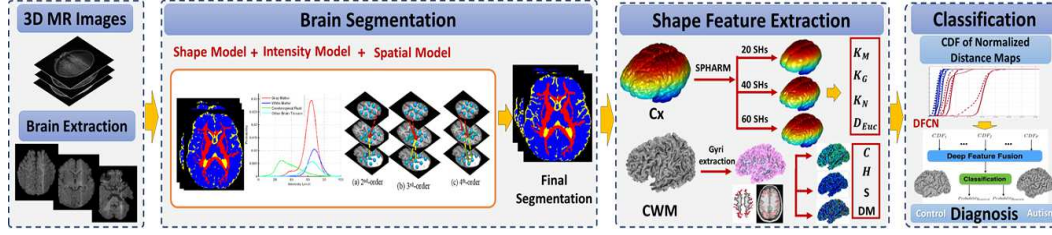


Fig. 1. Illustration of the proposed CAD system.

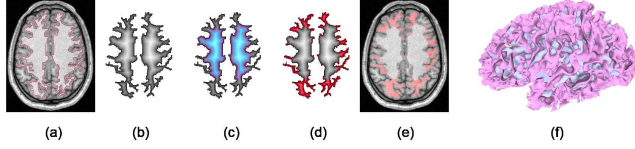


Fig. 2. (a) Boundary of segmented CWM. (b) First distance map. (c) Second distance map. (d), (e) Extracted CWM gyri. (f) 3D visualization of the extracted gyri.

be always the same for all subjects. This is because the shape of each mesh differs from the other, while there must be a consistent number of nodes to align meshes. Alignment is done using SPHARM-based registration [19]. Each node of the lobe mesh (4 per hemisphere) is then characterized with 4 features. These metrics are: Cx reconstruction error, by computing the Euclidean distance between the mesh's original and reconstructed nodes D_{Euc} ; Gaussian curvature K_G (product of the principal curvatures); mean curvature K_M (the average of the principal curvatures); and normal surface curvature K_N (maximum of the two principal curvatures).

CWM shape features: to analyze the CWM shape, its gyri are first extracted using a level-set based method. The distance map is calculated inside the segmented 3D CWM by a fast marching level set method [21]. Two waves are propagated; one finds voxels located at a distance less than or equal to T from CWM boundary, Fig. 2(b). The other one is then propagated to find voxels located at a distance less than or equal to T from voxels located at a distance T from CWM boundary, Fig. 2(c). Finally, voxels that are visited by the second wave are removed from the ones visited by the first wave. The remaining part is CWM gyri, Fig. 2(d), (e).

From the extracted gyral CWM, distance maps, DM , were calculated to be used as a shape feature for gyral CWM. In addition, curvature-based analysis is conducted for the 3D brain CWM meshes. Three features are calculated: the mean curvature MK , the curvedness C , which distinguishes highly folded regions from less folded ones [22], and the sharpness S , which quantifies the sharpness of folding [22].

2.3. Deep Fusion Classification Network (DFCN)

The goal of this stage is to obtain a global diagnosis from the collected shape features, 8 per lobe per hemisphere (64 per scan). This could be done using the raw features in a vertex-wise manner, but this is inefficient since the number of nodes for each mesh is 48K. This will result in having around

400K points per subject which is time consuming. To avoid this, features are represented using their cumulative distribution function (CDF) values that would retain all information as well as reduce data dimensionality. The CDFs are calculated with the minimum increment obtained from sorting all measures and selecting the min difference between consecutive points to capture all information in the distribution.

In order to build the classifier, the number of features per subject is 64 (8/lobe/hemisphere), and the length of each feature's CDF vector is 4000. This huge number of points, 64×4000 , required using a deep network. The model is built from a stack of auto-encoders (AEs) and output layer of softmax regression for each feature [23], and is composed of 2 stages. In the first stage, the CDF for each feature of the 64 is learnt separately using an AE with non-negativity constraint (NCAE) using sparsity (KL-divergence) to capture most prominent variations. Second stage is a supervised back-propagation method that minimizes the total loss (negative log-likelihood) for given training labeled data. Finally, the 64 high-level features extracted from each stacked NCAE (SNCAE) are concatenated as a new high-level feature, and fed into another SNCAE for fusion and final diagnosis, by taking 2 probabilities (autism, control) in its output layer.

3. EXPERIMENTAL RESULTS AND CONCLUSIONS

The proposed framework was tested on 94 subjects from different databases. 21 autistic subjects and their matched controls were analyzed from the Kennedy Krieger Institute (KKI)(8–12.8 years) [5]. Also 16 autistic adolescent and adult subjects from the University of Pittsburgh Collaborative Program of Excellence in Autism [4] (NDAR/Pitt) were used with their matched controls (16–51 years old). Also, 10 infants aged 6 months with autism and their matched controls from the NDAR/IBIS database [4] were selected for the study.

The proposed segmentation method was first validated on all the subjects from the three databases, and evaluated with their manually segmented ground truth created by an MR expert. The performance was evaluated using the Dice similarity coefficient (DSC) [24] by comparing it to the ground truth segmentation. This approach is also compared to iBEAT [25] package, and to FSL package [26]. The average accuracies for CWM and Cx segmentation in infant subjects using the proposed approach were 94.7% and 93.8% respectively, whereas

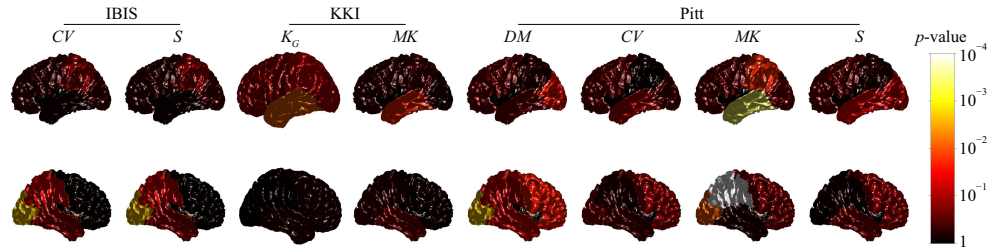


Fig. 3. Individual lobes showing strong evidence against ASD and control having the same mean value for each feature with a significant diagnosis×lobe or diagnosis×hemisphere interaction.

Table 1. Classification accuracy per lobe per hemisphere for statistically significant features using NCAE. 'I' stands for NDAR/IBIS, 'K' for KKI, and 'P' for NDAR/Pitt.

Metric	Frontal		Temporal		Parietal		Occipital	
	L	R	L	R	L	R	L	R
<i>C</i>	N/A	N/A	N/A	N/A	N/A	75% (I)	85% (I)	N/A
<i>DM</i>	N/A	92%(P)	N/A	N/A	N/A	N/A	91.5%(P)	94%(P)
<i>MK</i>	N/A	N/A	93%(P), 85.7%(K)	N/A	93.5%(P)	92.5%(P)	N/A	91%(P)
<i>S</i>	N/A	N/A	94%(P)	N/A	N/A	80%(I)	N/A	85%(I)
<i>K_G</i>	N/A	N/A	88%(K)	N/A	90%(K)	N/A	N/A	N/A

they were 73.3% and 81.6% using iBEAT and 80.4% and 89.5% using FSL. Subjects from the KKI database were segmented with accuracies of 95.8% and 96.7% for CWM and Cx respectively using the proposed approach, whereas accuracies were 85% and 88% using iBEAT and 88.4% and 92.3% using FSL. Finally, subjects from NDAR/Pitt database were segmented with accuracies of 96.1% and 97.8% for CWM and Cx respectively using the proposed approach, whereas accuracies were 87.3% and 89.6% using iBEAT and 92.4% and 93.2% using FSL. The proposed technique generated notably better results than other packages with infant subjects, in particular, owing to the higher-order MGRF model that distinguishes subtle differences in such images with poor contrast. Before training the features using the deep fusion model, we needed to test their significance and whether they will be discriminatory. Statistical analysis was thus performed on each dataset and each measurement separately. The same methodology was used throughout: a linear mixed effects model with factors diagnosis, lobe, and hemisphere in a full factorial design, age at time of scan as a covariate (KKI and NDAR/Pitt only), and random intercept for each individual, nested within sex. The dependent variable was the mean value of feature, computed from its respective CDF. Statistically significant effects of interest were, for the NDAR/IBIS dataset, diagnosis×lobe interaction for *C* and *S* measures. Post hoc testing of individual lobes suggested that both *C* and *S* differed between ASD and control in the right parietal and right occipital lobes. In the KKI data, there was a significant diagnosis×hemisphere interaction for *K_G*, with differences in the left temporal and left parietal lobes, and diagnosis×lobe interaction for *MK*, with differences in the left temporal lobe. All four of the CWM metrics had significant dependency on the diagnosis×lobe interaction in the

NDAR/Pitt dataset. Significant post hoc findings were found in the right frontal and both occipital lobes for *DM*; in left temporal, both parietal, and right occipital lobes for *MK*; and in left temporal lobe for *S*. Fig. 3 shows p-values of the diagnosis dependent difference in each individual lobe. As there were some features showing statistical significance for each database, this was encouraging to go for diagnosis, where each set was analyzed separately. First, the proposed deep network was trained and tested with only the discriminatory features using leave-one-subject-out approach. Deep fusion parameters were trained on (n-1) subjects, as the test wasn't included. Experiment is then updated with different training subjects and a new test. The classification accuracy for each significant feature per lobe per hemisphere found in each of the 3 databases was computed, and all accuracies are listed in Table 1. The high accuracies achieved were consistent with the statistical analysis. Next, discriminatory features from all 8 lobes for each database were fused, seeking higher classification accuracy. Fusing features from NDAR/IBIS, KKI, and NDAR/Pitt databases yielded classification accuracies of 81.25%, 90.47%, and 93.75% respectively. Finally, these fusing results were encouraging to train and test the deep fusion network with all the 64 extracted features. Final accuracies after integrating all features and testing for each database were: 85% for NDAR/IBIS, 92.8% for KKI, and 96.88% for NDAR/Pitt. While investigating the reason for improving accuracies, it is noted that some features other than those reported in Table 1 were close to be significant, which have improved accuracies when integrated with significant features.

To conclude, this paper proposed an autism diagnosis approach that analyzes the brain's local areas, which could help understand the autism spectrum. Future work targets integrating features from other modalities such as DTI and fMRI.

4. REFERENCES

- [1] NIMH, *Autism Spectrum Disorder*, Office of Science Policy, Bethesda, 2015.
- [2] Manuel F Casanova, Ayman El-Baz, and Jasjit S Suri, *Autism Imaging and Devices*, CRC Press, 2016.
- [3] Manuel F Casanova, Ayman S El-Baz, and Jasjit S Suri, *Imaging the brain in autism*, Springer, 2013.
- [4] Dan Hall et al., “Sharing heterogeneous data: the National Database for Autism Research,” *Neuroinformatics*, vol. 10, no. 4, pp. 331–339, 2012.
- [5] A. Di Martino et al., “The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism,” *Mol. Psychiatry*, vol. 19, no. 6, pp. 659–667, 2014.
- [6] Cynthia M Schumann et al., “Longitudinal magnetic resonance imaging study of cortical development through early childhood in autism,” *J. Neurosci.*, vol. 30, no. 12, pp. 4419–4427, 2010.
- [7] Heather Cody Hazlett et al., “Brain volume findings in 6-month-old infants at high familial risk for autism,” *Am. J. Psychiatry*, 2012.
- [8] Christine Ecker et al., “Investigating the predictive value of whole-brain structural mr scans in autism: a pattern classification approach,” *Neuroimage*, vol. 49, no. 1, pp. 44–56, 2010.
- [9] Lucina Uddin et al., “Multivariate searchlight classification of structural magnetic resonance imaging in children and adolescents with autism,” *Biological psychiatry*, vol. 70, no. 9, pp. 833–841, 2011.
- [10] Antonio Hardan et al., “A preliminary longitudinal magnetic resonance imaging study of brain volume and cortical thickness in autism,” *Biol. Psychiatry*, vol. 66, no. 4, pp. 320–326, 2009.
- [11] Gregory L Wallace et al., “Longitudinal cortical development during adolescence and young adulthood in autism spectrum disorders: Increased cortical thinning but comparable surface area changes,” *J. Am. Acad. Child Adolesc. Psychiatry*, 2015.
- [12] Manuel F. Casanova et al., “Minicolumnar abnormalities in autism,” *Acta Neuropathol.*, vol. 112, no. 3, pp. 287–303, 2006.
- [13] Marwa Ismail et al., “Segmentation of infant brain mr images based on adaptive shape prior and higher-order mrf,” in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4327–4331.
- [14] Marwa Ismail et al., “Detection of white matter abnormalities in mr brain images for diagnosis of autism in children,” in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. IEEE, 2016, pp. 6–9.
- [15] Ayman El-Baz, Georgy Gimelfarb, and Jasjit S Suri, *Stochastic modeling for medical image analysis*, CRC Press, 2015.
- [16] P. A. Viola and W. M. Wells III, “Alignment by maximization of mutual information,” *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.
- [17] A Alansary et al., “Map-based framework for segmentation of mr brain images based on visual appearance and prior shape,” *MIDAS J.*, vol. 1, pp. 1–13, 2013.
- [18] Julian Besag, “On the statistical analysis of dirty pictures,” *J. R. Stat. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.
- [19] Moo K Chung et al., “Tensor-based cortical surface morphometry via weighted spherical harmonic representation,” *IEEE Trans. Med. Imag.*, vol. 27, no. 8, pp. 1143–1151, 2008.
- [20] The CGAL Project, *CGAL User and Reference Manual*, CGAL Editorial Board, 4.7 edition, 2015.
- [21] D. Adalsteinsson et al., “A fast level set method for propagating interfaces,” *J. Comput. Phys.*, vol. 118, no. 2, pp. 269–277, 1995.
- [22] Rudolph Pienaar et al., “A methodology for analyzing curvature in the developing brain from preterm to adult,” *IJIST*, vol. 18, no. 1, pp. 42–68, 2008.
- [23] Y. Bengio et al., “Greedy layer-wise training of deep networks,” *Adv. Neural Inform. Process. Syst.*, vol. 19, pp. 153, 2007.
- [24] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [25] Dai Yakang et al., “iBEAT: a toolbox for infant brain magnetic resonance image processing,” *Neuroinformatics*, vol. 11, no. 2, pp. 211–225, 2013.
- [26] M. Jenkinson et al., “FSL,” *Neuroimage*, vol. 62, pp. 782–790, 2012.