Local semi-supervised approach to brain tissue classification in child brain MRI

Nataliya Portman*, Paule-J Toussaint, Alan C. Evans

McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, QC, Canada

Abstract

Most segmentation methods in child brain MRI are supervised and are based on global intensity distributions of major brain structures. The successful implementation of a supervised approach depends on availability of an age-appropriate probabilistic brain atlas. For the study of early normal brain development, the construction of such a brain atlas remains a significant challenge. Moreover, using global intensity statistics leads to inaccurate detection of major brain tissue classes due to substantial intensity variations of MR signal within the constituent parts of early developing brain.

In order to overcome these methodological limitations we develop a local, semi-supervised framework. It is based on Kernel Fisher Discriminant Analysis (KFDA) for pattern recognition, combined with an objective structural similarity index (SSIM) for perceptual image quality assessment. The proposed method performs optimal brain partitioning into subdomains having different average intensity values followed by SSIM-guided computation of separating surfaces between the constituent brain parts. The classified image subdomains are then stitched slice by slice via simulated annealing to form a global image of the classified brain.

In this paper, we consider classification into major tissue classes (white matter and grey matter) and the cerebro-spinal fluid and illustrate the proposed framework on examples of brain templates for ages 8 to 11 months and ages 44 to 60 months. We show that our method improves detection of the tissue classes by its comparison to state-of-the-art classification techniques known as Partial Volume Estimation.

Keywords: Kernel Fisher Discriminant Analysis, Structural Similarity, brain tissue classification, early brain development, NIH Objective-2, intensity variability, low contrast

*Corresponding author

Email addresses: nataliyaportman@gmail.com (Nataliya Portman), paule.toussaint@mcgill.ca (Paule-J Toussaint), alan.evans@mcgill.ca (Alan C. Evans)

URL: http://www.bic.mni.mcgill.ca/~alan (Alan C. Evans)
1. Introduction

1.1. Motivation

The motivation for a new classification method comes from MRI study of early normal brain development (Almli et al., 2007, Evans, 2006) designed to create a database of healthy controls comprised of T1-weighted (T1w), T2-weighted (T2w) and proton-density weighted (pdw) images. MR dataset of early brain development (from birth to 4.5 years of age) poses a challenge to using automatic classification procedures such as FAST (Collins et al., 1999), ANIMAL+INSECT (Collins et al., 1999) and SPM (Collins et al., 1999). These methods rely on young adult brain atlases that do not adequately reflect dynamic changes in brain anatomy through early childhood (Almli et al., 2007, Fonov et al., 2011).

Furthermore, there is a number of factors inherent to early developing brain MR images that hamper automated classification process. They include high tissue intensity variation, low contrast-to-noise ratio between tissue types and partial volume artifact meaning that a brain voxel possesses a signal average of two or three different tissue classes.

However, the development of segmentation techniques in newborn and infant brain MRI over the past decade has shown that atlas-based segmentation methodology for adult brain MR images can be applicable to child brain MRI when tuned to the specific properties of the dataset under study. There are three major child MRI segmentation frameworks, Expectation-Maximization (EM) (Van Leemput et al., 1999), Registration-based method (Collins et al., 1999) and Adaptive Label Fusion (Weisenfeld and Warfield, 2009).

A barrier to implementation of these methods is the dependency on a pediatric brain atlas with accurate measures of spatial boundary uncertainty for tissue classes that the early childhood dataset does not possess. Up to date, the creation of such standard atlas that would cover the entire developmental epoch still remains a significant challenge since it requires semi-automated segmentation of large datasets.

Another methodological issue with EM-based and Adaptive Label Fusion methods in applications to infant brain MRI is using global intensity statistics that requires for each tissue type to produce similar intensities in different parts of the image.

1.2. Handling intra-class variability

It is important to find a strategy to cope with a highly variable signal intensity since the same tissue type having different intensities in different parts of the brain is prone to misclassification. An effective remedy to this problem was offered by Pappas (Pappas, 1992). The author devised an adaptive clustering algorithm that estimates the mean intensity function (representing the pure space-dependent signal) and the corresponding posterior probabilities of the classes for all pixels in a 2D image with a varying window size. In this way, adaptability of the mean intensity function from the initially global estimates of the class means to the local characteristics of the image was achieved. However,
the implementation of this adaptive clustering method to a 3D image is computationally expensive due to the slow process of successive updates of mean class intensities for each pixel within the window making the extension of this technique to 3D MRI segmentation unfeasible.

To tackle high intra-class intensity variability in EM framework, Prastawa et al. (Prastawa et al., 2005) used a technique developed by (Van Leemput et al., 1999) incorporating bias field correction step into EM algorithm. The segmentation results were refined by pruning the training samples used to estimate the probability distributions of classes for each voxel by Parzen windowing (Wells et al., 1996). The successful EM classification of child brain MRI is achieved at a high computational cost associated with overcoming restrictions of the global Gaussian mixture modelling.

(Tohka et al., 2004) included of the partial volume effect (PVE) model (Choi et al., 1991) in EM framework that estimated mixing tissue proportions within each voxel and improved detection GM, WM and the CSF in small structures and areas between sulcal CSF and the surrounding GM. The PVE algorithm has been successfully applied to single or multi-channel adult brain scans. In this work, we investigate the performance of PVE technique in segmentation of young child’s brain MRI.

Another approach is local segmentation by brain splitting introduced by (Xue et al., 2007). The brain was subdivided into regions with different statistical properties of WM samples using k-means clustering followed by a Voronoi tessellation. The EM segmentation performed on each Voronoi region prevented overestimation of cortical GM. However, the number of subvolumes in Xue’s brain splitting algorithm is arbitrary.

The label fusion technique handles tissue intensity variability by means of constructing a probabilistic atlas from a small cohort of newborn brain MRI and incorporating it into the fused segmentation model as a spatial prior (Warfield et al., 2000). This introduces dependency on a brain atlas that our MR dataset of early brain development does not possess.

1.3. The proposed framework

We propose a novel local atlas-independent framework inspired by modern trends in Computer Vision such as Kernel Fisher Discriminant Analysis (KFDA) for pattern recognition and structural similarity index (SSIM; Wang et al., 2004) for perceptual image quality evaluation. KFDA is related to kernel-based classifiers such as Support Vector Machines (SVM) introduced by (Vapnik, 1998). KFDA has shown competitive performance compared to other state-of-the-art classifiers such as SVM and AdaBoost (Mika et al., 1999, Rätsch et al., 2001).

In the proposed framework, the KFDA discriminant criterion is modified by including a regularization term that penalizes intensity differences in the neighbourhood of a brain voxel. A regularized solution in the abstract high-dimensional feature space yields connected, and therefore more meaningful, spatial tissue patterns.

We build on the global version of the multi-channel KFDA-based classification
algorithm introduced in [Portman and Evans, 2013] and further refine the algorithm to improve its performance. Namely, we propose a local approach that handles within-class intensity variations by optimal partitioning of the brain into overlapping subdomains having different average intensities.

We aim to classify age-specific pediatric templates into two major tissue classes (GM, WM) and the CSF. The classified representative templates can then be used for the classification of a subject brain MRI of a developmental age similar to the age range of the template.

In the absence of a “ground truth”, we assess the quality of classification via SSIM that predicts image quality as perceived by the Human Visual System (HVS) [Wang and Bovik, 2009] which is regarded as an optimal information extractor that seeks to identify objects in the image. As such, the HVS must automatically identify structural distortions and compensate for the nonstructural distortions (e.g., image corruption by noise). The SSIM is an objective image quality metric that simulates this functionality and computes the degree of structural similarity between reference and distorted images. It has been shown that the SSIM is well-matched to the subject ratings of image databases, and therefore, it is a good approximation to the perceived image quality [Wang et al., 2003]. In this work, the SSIM finds a new role as

Given the objectivity and effectiveness of the SSIM we apply it for comparison of Partial Volume Estimation [Tohka et al., 2004] and KFDA-based classification algorithms, as well as for automatic monitoring of the quality of classification. That is, we compute the structural closeness of classified 3D brain images containing GM, WM and CSF mean intensity values with their counterparts seen in an MR image type of a highest contrast and regarded as references. In a test example of a brain template for ages 8 to 11 months shown in Figure ?? we rely on T1w as the most informative of all MR image types.

1.4. Paper organization

In the following we will describe the algorithm step by step, namely, optimal brain partitioning, modified KFDA-based classification guided by SSIM, and image stitching. A brief background on KFDA is provided in Appendix A. Finally, we will report on classification results for brain template for ages 8 to 11 months, compare the performance of PVE and KFDA methods and discuss possible improvements.

2. Method

2.1. Optimal brain partitioning

We start with a mathematical formalization of the brain splitting problem. Let $X$ be a discrete random variable (r.v.) representing the bins $X_i$ of the histogram containing $n_i$ interior brain voxels with marginal distribution $\{p_i\} = \{\frac{n_i}{N}\}$, $Y$ be the voxel-to-voxel image partition containing $N$ voxels with uniform distribution $\{q_j\} = \{\frac{1}{N}\}$ and $\hat{Y}$ be a random clustering over $Y$ into $2^k$ clusters containing $N_j$ voxels at the partitioning step $k$. 
For each of \(2^{k-1}\) subvolumes at the partitioning step \(k\) (referred to as \(Y\), for simplicity) we aim to find the partition \(\hat{Y} = \{\hat{Y}_i\}_{i=1}^2\) that maximizes \(MI(X, \hat{Y})\) defined by

\[
MI(X, \hat{Y}) = -\sum_{i=1}^{2} p_i \log p_i + \sum_{j=1}^{2} \sum_{i=1}^{2} p_{ij} \log p_{ij}. \tag{2.1}
\]

In equation (2.1), \(\{p_i\}\) is the marginal probability distribution of the histogram bins, \(\{p_{ij}\}\) is a joint probability distribution of r.v. \(X\) and \(\hat{Y}\) and \(\{p_{ij}\}\) is the conditional probability of the r.v. \(X\) taking the value \(X_i\) given that the r.v. \(\hat{Y}\) is the jth cluster \(\hat{Y}_j\).

\[
p_i = p(X = X_i) = \frac{n_i}{N}, \tag{2.2}
\]

\[
p_{ij} = p(X = X_i)p(\hat{Y} = \hat{Y}_j|X = X_i) = \frac{|\hat{Y}_j \cap X_i|}{N} = \frac{n_{ij}}{N}, \tag{2.3}
\]

\[
p_{ij} = p(X = X_i|\hat{Y} = \hat{Y}_j) = \frac{|\hat{Y}_j \cap X_i|}{N} = \frac{n_{ij}}{N_j}. \tag{2.4}
\]

Here, \(n_{ij}\) is the number of shared voxels between the cluster \(\hat{Y}_j\) and the histogram bin \(X_i\), and \(N_j\) is the number of voxels in \(\hat{Y}_j\). Substituting (2.3) and (2.4) for \(p_{ij}\) and \(p_{ij}\) into equation (2.1) we obtain

\[
MI(X, \hat{Y}) = -\sum_{i=1}^{2} \frac{n_i}{N} \log \left(\frac{n_i}{N}\right) + \sum_{j=1}^{2} \sum_{i=1}^{2} \frac{n_{ij}}{N} \log \left(\frac{n_{ij}}{N_j}\right). \tag{2.5}
\]

The \(MI\) can be rewritten in terms of entropy functions as follows

\[
MI(X, \hat{Y}) = H(X) - H(X|\hat{Y}). \tag{2.6}
\]

When \(\hat{Y}\) is a single intensity voxel \(Y_j\) and \(X\) takes values from a set \(\{X_i\}_{i=1}^2\) then there is no uncertainty that this voxel belongs to one of the bins. So, \(H(X|Y)\) is equal to 0 implying that

\[
MI(X, Y) = H(X). \tag{2.7}
\]

**Brain partitioning algorithm.**

1. Given \(2^{k-1}\) brain volumes (or nodes of a binary tree at the level \(k-1\)) initially partition into \(2^k\) subvolumes by cutting each of the volumes into \(\hat{Y}_1\) and \(\hat{Y}_2\) that contain three subsequent slices and the rest of the brain slices, respectively, with sagittal, coronal and axial planes. For each direction (sagittal, coronal, axial) compute \(MI\) (2.5) between the histogram bins \(\{X_1, X_2\}\) and the initial clusters.

2. For each subvolume \(Y\) and each direction, create a set of new clusters by moving the cutting plane one slice at a time (until only three slices remain in \(Y_2\)). Find the clustering that maximizes \(MI\) over a set of all possible clusters \(\hat{Y} = \{\hat{Y}_1, \hat{Y}_2\}, \hat{Y}_1 \in Y, i = 1, 2\) in all directions.

\[
\hat{Y}^* = \arg \max_{\{\hat{Y}_1, \hat{Y}_2\}} MI(X, \hat{Y}).
\]
3. Given the total number $M$ of brain subvolumes that possess $MI$ from the top level 1 down to the current level $k$ of the binary tree compute the total $MI_t$ acquired in the partitioning process. $MI_t$ is the weighted sum of the $MI_i$ associated with the subdomains $i$, $i = 1, \ldots, M$.

$$MI_t(X, \hat{Y}^*) = \sum_{i=1}^{M} \frac{N_i}{N} MI_i = \sum_{i=1}^{M} \frac{N_i}{N} H_i(X) - \sum_{i=1}^{M} \frac{N_i}{N} H_i(X|\hat{Y}^*) = H_t(X) - H_t(X|\hat{Y}).$$

(2.8)

The weights $\frac{N_i}{N}$ are relative volumes of the image subdomains, where $N$ is the number of voxels comprising the entire 3D brain volume.

4. Compute the Mutual Information Ratio ($MIR$) curve using equation (2.8)

$$MIR(X, \hat{Y}^*) = \frac{MI_t(X, \hat{Y}^*)}{MI_t(X, Y)},$$

(2.9)

where $MI_t(X, Y) = H_t(X)$ due to (2.7).

5. Compute signal-to-noise ratio (SNR) averaged over newly obtained $2^k$ brain subvolumes to control growth of the MIR. Construct the SNR curve with respect to the number of subvolumes $2^i$, $i = 1, 2, \ldots, k$ and normalize it to the range of MIR values.

6. Move down to the level $k$ of the binary tree and repeat the procedure from step 1 until the following stopping criterion is satisfied.

- **Stopping criterion:** Find the abscissa of the intercept between the MIR and SNR curves that provides an optimal number of subvolumes.

Note: With the growing number of partitioned brain subvolumes the acquisition of information increases. Therefore, the $MIR$ is an increasing function with respect to the number of subvolumes. As the subdomains decrease in size in the process of brain splitting, the SNR decreases. The subvolume sizes should allow samples of brain voxels large enough to distinguish between the constituent tissue classes. In order to control the brain subvolume sizes and deduce an optimal number of subvolumes we compute the SNR. In this way, the balance between the ability of the proposed method to classify image subdomains and the SNR is established.

**Algorithm implementation.**

Figure 2.1 shows the graphs of the SNR and MIR curves as functions of the number of subdomains in the T1w template for ages 44 to 60 months. The abscissa of the intercept yields the optimal number of 22 subdomains for this template. This number falls between 16 and 32 subdomains of the levels 4 and 5 of the binary tree. To determine these 22 brain subvolumes we followed a simple decision rule. We sorted the 32 subdomains in descending order according to their size and selected the first 22 subdomains with a larger size and therefore a higher SNR.

To maintain the continuity of the classified image subdomains across their boundaries, we added two slices to each planar boundary of each subdomain.
Figure 2.1: MIR and SNR curves across 7 levels of the binary tree of the brain partitioning process in the T1w template for ages 44 to 60 months.

thus creating an overlap of 4 slices between adjacent brain subdomains. Since the partitioning algorithm is intensity-based, the brain regions and their total number vary for different brain scans. Partitioning of brain MRI with a higher intensity variation yields a larger number of brain regions as demonstrated in an example below. The T1w infant brain template for ages 8 to 11 months was subdivided into 40 subvolumes (Figure 2.2a). Local assessment of the contrast-to-noise ratio (CNR) between the CSF and G+WM showed that CNR significantly varies across brain subdomains as seen in Figure 2.2b. It tends to be lower near the base of the brain and higher in the central part of the brain.

In a similar fashion, we can compute GM/WM CNR in image subdomains of the brain template. Local dependency of CNR graphs provides a better understanding of natural MR signal intensity variation throughout an individual brain shedding light on subvolumes with poor CSF and/or WM signal detection. Since the proposed method is kernel-based the kernel parameters (e.g., bandwidth of the Gaussian kernel) can be tuned over a certain range of values in accordance with CNR of brain subvolumes to ensure successful segmentation.

2.2. Modified KFDA optimality criterion

We reap the benefits of the KFDA approach to brain tissue classification of infant brain MRI. To mention a few, KFDA allows a simple closed form solution, improves classification by taking into account all MR signal observations (samples) in the input space, and automatically identifies PV voxels that lie near the
boundary between the tissue classes. The competitiveness of the KFDA method compared to other state-of-the-art classifiers (Mika et al., 1999) has motivated our exploration of KFDA to find high accuracy segmentation solutions in child brain MRI.

We implement a binary formulation of KFDA in the feature space $\mathcal{H}$ provided in Appendix A. The classical kernel Fisher discriminant given by (A.10) does not assume spatial correlations between neighbouring intensity voxels in $\mathcal{H}$. To increase robustness to misclassification caused by the presence of noise in MR images we introduce a spatial regularization term that penalizes local kernel-projected intensity differences in the feature space $\mathcal{H}$. Since the graph that defines the local relationships between the brain voxels is a 3D grid with each interior node having 26 neighbours we define the neighbourhood matrix as fol-
\[ H_{ij} = \begin{cases} 1, & \text{if } (i, j) \text{ is an edge;} \\ -d_{ij}, & \text{if } i = j, \text{ the degree of vertex } i; \\ 0, & \text{otherwise.} \end{cases} \]

Then
\[ \tilde{V}^T H \tilde{V} = - \sum_{(i,j) \in E} (V_i - V_j)^2, \quad \forall \tilde{V} \in \mathbb{R}^l, \tag{2.10} \]

where \( E \) is a set of edges \( \{V_i, V_j\} \).

Let \( \tilde{V} = \bar{w} \cdot \phi(\bar{I}) \) be the kernel projection of the input data \( \bar{I} \) onto the optimal direction \( \bar{w} \) in \( \mathcal{H} \). \( \tilde{V} \) can be rewritten as \( \tilde{V} = \sum_{i=1}^l \alpha_i K(\bar{I}, \bar{I}) \) due to the expansion of \( \bar{w} = \sum_{i=1}^l \alpha_i \Phi(\bar{I}_i) \) in \( \mathcal{H} \) spanned by the mapped training samples \( \Phi(\bar{I}_i), i = 1, 2, ..., l \). We modified the KFDA optimality criterion by adding the penalty term of the form \( V^T HV = \alpha^T kHk^T \alpha \), where \( k \) is the kernel matrix of size \( l \times l \)
\[ \hat{\alpha} = \arg \max \alpha \left( \frac{\alpha^T M \alpha + \lambda \alpha^T kHk^T \alpha}{\alpha^T (N + \beta I) \alpha} \right). \tag{2.11} \]

In this setup, the penalty function forces misclassified voxels closer to another class cluster centroid.

The problem (2.11) can be reformulated as
\[ \hat{\alpha} = \arg \min \alpha \left( -\frac{1}{2} (\alpha^T M \alpha + \lambda \alpha^T kHk^T \alpha) \right) \text{ subject to } \alpha^T (N + \beta I) \alpha = 1. \tag{2.12} \]

The constrained optimization problem (2.12) can be solved algebraically using the method of Lagrange multipliers. We constructed the Lagrangian function
\[ L = -\frac{1}{2} (\alpha^T M \alpha + \lambda \alpha^T kHk^T \alpha) + \frac{1}{2} \gamma (\alpha^T (N + \beta I) \alpha - 1) \tag{2.13} \]

and by computing the gradient of the Lagrangian \( L \) with respect to \( \alpha \) we obtained an eigen-value problem
\[ (N + \beta I)^{-1}(M + \lambda kHk^T) \alpha = \gamma \alpha. \tag{2.14} \]

Then the solution to (2.11) is a leading eigen-vector of \( (N + \beta I)^{-1}(M + \lambda kHk^T) \).

### 2.3. Objective assessment of classification quality via SSIM

A modified version of the KFDA criterion given by (2.11) depends on a regularization coefficient \( \lambda \) that influences the quality of classification. In order to find the “best” value of \( \lambda \) we constructed a sequence of \( \lambda \)-values \{\( \lambda_i = 0.000025 \cdot i \), \( i = 1, ..., n \)\} and computed the Structural Similarity Index (SSIM) (Wang et al. 2002, 2004) between the classified and structural MR input data for each \( \lambda \)-value in the sequence. In this application, the SSIM quantifies the degree of structural similarity between spatial tissue patterns seen in ideal (reference) and distorted (classified) images.
Figure 2.3: Dependency of the CSF classification on $\lambda$ in the template subvolume near the base of the brain (ages 8-11 months): CSF patterns for increasing values of $\lambda$ from 0 to 0.0001 in a brain subvolume consisting of 6 slices.

We implemented the SSIM to evaluate the performance of our classification algorithm in the absence of ground truth. We evaluated how well GM, WM and CSF boundaries are captured by our algorithm versus the boundaries visually extracted from input T1w images.

For each partitioned brain subvolume we solved a 2-class separation problem and created $n$ subvolumes with the mean T1w intensity values for the two tissue types (G+WM and CSF, and GW and WM). These are the classified $\lambda$-dependent image subdomains to be compared with T1w reference. We computed the SSIM between each classified and T1w brain slices and then found the mean SSIM taken over all slices of the brain subvolume. Thus, we obtained $n$ mean Structural Similarity Indices (MSSIM) and chose the $\lambda$ value corresponding to the largest MSSIM. In this way, we were able to automatically control the quality of classification.

Given below is a formula for the mean SSIM

$$MSSIM = \frac{1}{N} \sum_{i=1}^{N} SSIM(x_i, y_i), \quad SSIM(x_i, y_i) = l(x_i, y_i) \cdot c(x_i, y_i) \cdot s(x_i, y_i),$$

where $N$ is the total number of interior brain voxels, $x_i$ and $y_i$ are local image patches$^2$ and $l(x_i, y_i)$, $c(x_i, y_i)$, $s(x_i, y_i)$ are the luminance, contrast and

---

2A sliding window that moves across the entire brain slice pixel by pixel. For the MSSIM the background patches have been excluded.
structure comparison measures defined by
\[ l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}; \quad c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}; \quad s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}. \]

Here, \( \mu_x(\mu_y) \), \( \sigma_x(\sigma_y) \) and \( \sigma_{xy} \) represent the local mean, standard deviation and cross-correlation estimates of the local image patches \( x \) and \( y \), respectively, and \( C_1, C_2, C_3 \) are small constants that have been derived from the properties of the visual system (Wang et al., 2002).

Figure 2.3 illustrates the dependency of the CSF pattern identified by KFDA on the value of the regularization parameter \( \lambda \). The value of \( \lambda = 0.00005 \) reveals most of the connected CSF tissue surrounding sulci, thus making the sulcus contours visible. The computed MSSIMs between the classified subdomains and the corresponding T1w template for \( \lambda \)-values from the sequence \( \{0.000025 \cdot i\}_{i=0}^{4} \) show that the CSF structure corresponding to \( \lambda = 0.00005 \) is most similar to its counterpart in T1w reference. Therefore, the regularization term builds more of the connected CSF component by forcing initially identified G+WM voxels into the CSF class if predominant neighbouring intensities are the CSF samples.

### 2.4. KFDA implementation

To proceed with the KFDA implementation, we solve an eigen-value problem (2.14) in a high-dimensional feature space \( \mathcal{H} \) whose dimension is defined by the number of brain voxels in the image subdomain \( N \). Given a vector-valued image function with labels \( l = \{-1, +1\} \) for the two tissue classes (CSF and G+WM, or GM and WM) as an input
\[
(\vec{I}(i, j, k), l) = (t1w(i, j, k), t2w(i, j, k), pdw(i, j, k), l),
\]
where \( 1 \leq i \leq M_1, 1 \leq j \leq M_2, 1 \leq k \leq M_3 \) are voxel coordinates of the interior brain subvolume, the input data are implicitly and nonlinearly transformed to \( \mathcal{H} \)
\[
\vec{\phi}: \vec{I} \in R^3 \rightarrow \vec{I}^*.
\]

Next, having chosen a kernel function \( K \) and computed a leading eigen-vector \( \alpha \) of the matrix given in equation (2.14), we find an optimal projection of the mapped data in \( \mathcal{H} \)
\[
\vec{w} \cdot \vec{\phi}(\vec{I}) = \sum_{m=1}^{N} \alpha_m K(\vec{I}_m, \vec{I}) + b.
\]

We performed KFDA in two steps.

**Step1: Classification into the CSF and G+WM.** Through extensive experimentation we determined that the sigmoidal kernel function \( K(\vec{I}_m, \vec{I}) = \tanh(a(\vec{I}_m \cdot \vec{I}) + b) \) (with \( a \) and \( b \) being user-specified parameters) is the best choice for delineation of the CSF. Figure 2.4 shows a test example of a template brain subvolume for ages 8 to 11 months that consists of 7 axial slices with initial CSF and G+WM labels obtained using global PVE. The values of the image vector-function \( \vec{I} \) are plotted in the 3D intensity space shown in Figure 2.4b.
Each point in the 3D input space carries specific stereotaxic coordinates \((i, j, k)\) of a brain voxel. The initial CSF and G+WM clusters are coloured in blue and red, respectively.

An optimal projection of the input data \(\vec{w} \cdot \vec{\phi}(\vec{I})\) with initial classification shown in red (G+WM) and blue (CSF) is given in Figure 2.4c. The non-linear data transformation via sigmoidal kernel preserves the sparsity of the input CSF pattern and the density of the G+WM pattern. In this Figure, the X-axis represents column-wise enumeration of the interior brain voxels from 1 to \(N\) and the Y-axis represents the projected values \(\vec{w} \cdot \vec{\phi}(\vec{I}_i)\), \(1 \leq i \leq N\). When calculated with the offset \(b\) they are positive for one class and negative for another. Peaks of both class clusters correspond to the voxels that lie deep inside the tissue volumes and the voxels next to the decision line \(y = 0\) lie close to the boundary between the CSF and G+WM.

Voxel categorization. Since KFDA computes an optimal decision surface between the CSF and G+WM it easily identifies PV voxels that lie near or on the boundary between the classes.

A close-up look at the distribution of the kernel-projected data reveals overlapping CSF and G+WM voxels and class outliers (Figure 2.5a). We identified the following voxel categories,

- G+WM overlapping voxels (in the negative CSF range) in black,
- CSF overlapping voxels (in the positive G+WM range) in green,
- CSF outliers in cyan,
- G+WM outliers in yellow,
- CSF and G+WM tissue prototypes in blue and red, respectively.

A visualization of the overlapping set (coloured in black and green) in the stereotaxic space given in Figure 2.5b shows that the overlapping voxels are located within the boundary regions between G+WM and the CSF. We recognize a particularly problematic brain area around the sulcus where the CSF is usually poorly detected due to the presence of PVE. Therefore, it is reasonable to assume that the overlapping voxels contain a mix of both tissue classes. The outliers located away from the decision boundary and their initial class centroids are most likely to change their initial membership and the rest of the brain voxels constitute tissue prototypes that truly represent CSF and G+WM tissues.

Voxel categorization is useful since we can use the predictive power of KFDA to assign the most likely class membership to overlapping voxels (containing PVE). For this purpose, we treat the overlapping voxels as a testing set and tissue prototypes as a training set. For the classification of the overlapping set we implemented \(k\)-nearest neighbours (KNN) classifier. It determines the class membership based on the class majority rule in the vicinity of each voxel defined by Euclidean distances to \(k\) nearest tissue prototypes in \(\mathcal{H}\). The outliers comprise a separate testing set and they are classified using Mahalanobis distance.
Figure 2.4: (a) T1w, T2w and pdw input data from a pediatric template for ages 8 to 11 months with initial CSF and G+WM labels obtained using global PVE, (b) CSF and G+WM clusters in 3D intensity space based on the initial labeling, (c) Optimal projection of the input data given in (b) onto the vector of maximal information discrimination in the feature space. The kernel is sigmoid with $a = 8$ and $b = -0.0005$. 

13
It is tempting to classify all kernel-projected samples into 2 classes based on their sign (positive or negative). However, this usually leads to an overestimate of the CSF. To monitor the quality of classification produced by predictive labelling of the testing set we devised the following SSIM-guided classification procedure.

An algorithm for SSIM-guided computation of the decision surface.

1. Classify the outlier set using Mahalanobis distance
2. Compute the resulting classified image with tissue class means and the $MSSIM_{Mahal}$
3. Classify the overlapping set using KNN classifier with different values of the number of neighbours $k$
4. For each value of the parameter $k$, compute the resulting classified image and the $MSSIM_{knn}$. Choose the classification that corresponds to the maximal $MSSIM$
5. Compare $MSSIM_{Mahal}$ and $MSSIM_{knn}$ and return the classification that corresponds to a larger value.

This algorithm is general enough to handle a variety of class intensity distributions. For a test example provided by Figure 2.4 there are only a few class...
outliers and their assignment to different classes will not make a visible difference to the initial classification. Due to a large overlapping set, application of the KNN classifier contributes most to the final KFDA classification of the test example. However, there are cases when there is a large number of the class outliers and very few overlapping voxels. In this case, Mahalanobis classification of the outlier set suffices to apply. Figures 2.6(a-b) demonstrate steps 1-4 of

![Figure 2.6: (a) Mahalanobis classification of the outlier set, (b) KNN classification of the overlapping set, (c) decision surface delineating the CSF cluster in the input space, (d) classification in the stereotaxic space corresponding to a larger MSSIM value $\text{MSSIM}_{knn} = 0.8404$.](image)

the algorithm, and Figure 2.6(c) shows the separating surface in the input space corresponding to the KNN classification with an optimal value of the parameter $k$ shown in Figure 2.6(b). The discriminating boundary between the CSF and G+WM clusters is a plane. The CSF cluster identified by the SSIM-guided algorithm contains 1917 voxels, a significant increase over the initial CSF volume consisting of 515 voxels. It is shown in stereotaxic coordinates in Figure 2.6(d).

**Step 2:** Classification into GM and WM.
Having delineated the CSF we classify G+WM into GM and WM. Experiments with various kernel functions showed that the Gaussian radial basis function
Figure 2.7: (a-b) G+WM cluster in the T1w template brain subvolume (Figure 2.4.a) with initial GM and WM labels obtained by PVE, (c) Non-linear mapping and optimal projection of the labeled G+WM input data using a Gaussian RBF with $\sigma = 0.5$, (d) Voxel categories: GM (cyan) and WM (yellow) outliers and GM overlapping voxels (green), (e) GM (blue) and WM (red) clusters in the feature space, (f) Decision surface in the input space.

$$K(\vec{I}, \bar{I}) = \exp \left( -\frac{(\vec{I} - \bar{I})^T(\vec{I} - \bar{I})}{2\sigma^2} \right)$$ is the best choice to model the non-linear structure of WM and GM clusters. The G+WM of the test example with initial GM and WM labels is given in Figures 2.7(a-b). The initial classification represents a significant underestimate of WM. The small WM structure in the temporal lobe of low signal intensity is visible but not detected by the PVE method. We
performed KFDA of the labeled G+WM input data and displayed the projected GM and WM distributions in $H$ as shown in Figure 2.7.e. Here, WM and GM peaks correspond to the voxels located deep inside the GM and WM. Unlike the CSF and G+WM case, GM and WM distributions contain only GM overlapping voxels that fall in the positive range of WM and class outliers (Figure 2.7.d). The tissue prototypes that form the training set are color-coded in red for WM and blue for GM.

We then applied the algorithm for SSIM-guided computation of the decision surface and identified 819 WM voxels in the feature space in addition to the initial 482 WM voxels shown in red in Figure 2.7.e. The corresponding non-linear decision surface that attempts to mimic the shape of the GM cluster in the input space is displayed in Figure 2.7.f. Figure 2.7.f also demonstrates the importance of providing multichannel data for the input. This test example possesses a wide range of T1w WM intensities and a narrow range of T2w WM intensities. Without the T2w imaging data it would not be possible to detect myelinated WM with a lower signal intensity. SSIM comparison of KFDA and PVE classification results shows that KFDA yields spatial tissue patterns that are structurally closer to their counterparts in T1w reference subdomain (Figure 2.8).

2.5. Stitching of brain subdomains

We applied a Simulated Annealing technique to stitch the local classifications together into a cohesive global picture of the classified brain. First of all, for each brain slice we collected the constituent subimages. Each pair of overlapping subimages contained a joint image region $I_s$ of size $4 \times n\text{cols}$ or $n\text{rows} \times 4$ that is to be optimally estimated from two sets of observations $I_l, I_r$ or $I_{up}, I_{low}$ as illustrated in Figures 2.10.a-b.

We define an undirected graph $G = (V, E)$ to represent a lattice structure of the image $I_s$, where $V$ are nodes of the graph and $E$ are pairs of neighbouring nodes or edges. For every pixel within $I_s$ there is a hidden node. The node is a random variable taking values from the state space $S = \{1 (CSF), 2 (GM), 3 (WM), 4 (BG)\}$. The nodes are divided into 2 sets (Figure 2.10).
• $Y_{ij}$ - the observed random variable with a label $s \in S$ at a pixel coordinate $(i, j)$,

• $X_{ij}$ - a hidden node.

Figure 2.9: Markov Random Field image model.

Given neighbouring node interactions it is natural to model $Y$ as a realization of a Markov Random Field $X$ with respect to the graph $G = (V, E)$. The hidden nodes $X_{ij}$ and $X_{mn}$ are connected in a lattice way if $i \neq j$ and $m \neq n$. Each possible label coupling and label weights are modelled by the edge potential $\Psi(X_{ij}, X_{mn})$ and the node potential $\Phi(X_{ij})$, respectively. The posterior probability density of the joint label configuration becomes

$$p(X|Y) = \frac{1}{Z(Y)} \prod_{(X_{ij}, X_{mn}) \in E} \Psi(X_{ij}, X_{mn}) \prod_{(i,j) \in V} \Phi(X_{ij}), \quad (2.15)$$

where the first product is over all pairs of neighbouring nodes and the second one over all nodes. $Z(Y)$ is a normalizing constant that sums the probabilities over all possible configurations of the variables to 1.

To find labelling $X$ of the joint region $I_s$ with the highest probability $p(X|Y)$ we estimate $\Psi(X_{ij}, X_{mn})$ and $\Phi(X_{ij})$ for every node $X_{ij}$ from two sets of observations $Y$ (e.g., $Y = \{I_l, I_r\}$). $\Psi(X_{ij}, X_{mn})$ is defined on $S \times S$ and for every $X_{ij}$ it is given in the form of a $4 \times 4$ matrix with entries

$$\Psi(X_{ij} = s_k, X_{mn} = s_l) = \begin{cases} 1, & \text{if } (Y_{ij}, Y_{mn})_{I_r} = (Y_{ij}, Y_{mn})_{I_l} = (s_k, s_l); \\ 0.5, & \text{if } (Y_{ij}, Y_{mn})_{I_l} = (s_k, s_l) \\ & \text{and } (Y_{ij}, Y_{mn})_{I_r} \neq (s_k, s_l) \text{ or vice versa}; \\ 0.01, & \text{if } (s_k, s_l) \text{ is not observed in } I_r \text{ and } I_l. \end{cases}$$

Here, $(s_k, s_l) \in S \times S$, the space of all possible combinations of label pairs. The
node potential assigns weight to every possible value of the node $X_{ij}$ as follows

$$\Phi(X_{ij} = s_k) = \begin{cases} 
  w, & \text{if } Y_{ijI_l} = Y_{ijI_r} = s_k; \\
  w \cdot 0.5, & \text{if } Y_{ijI_l} = s_k; \\
  & \text{and } Y_{ijI_r} \neq s_k \text{ or vice versa}; \\
  w \cdot 0.01, & \text{if } s_k \text{ is not observed in } I_r \text{ and } I_r.
\end{cases}$$

Here, we also take into account nodal locations, namely, the rightmost nodes in $I_r$ and the leftmost nodes in $I_l$ (similarly, the uppermost nodes in $I_{up}$ and the lowermost nodes in $I_{low}$) are more likely to have values observed in $I_r$ and $I_l$ (or $I_{up}$ and $I_{low}$), respectively. In this way, we ensure the continuity of label propagation from non-overlapping regions to the joint region $I_s$. The additional weight function $w$ is defined as

$$w(X_{ij}) = \begin{cases} 
  \frac{3}{4}, & \text{if } X_{ij} \text{ is the leftmost/rightmost node in } I_l/ \text{or } I_r; \\
  \frac{1}{4}, & \text{or the uppermost/lowermost node in } I_{up}/I_{low}; \\
  1, & \text{otherwise}
\end{cases}$$

With this setup, we implemented the iterative Simulated Annealing (SA) algorithm (Grenander 1996) to find the most probable joint region for each pair of overlapping classified subimages in a brain slice. We initialized $I_s$ as a region containing the first two columns/rows of $I_l/I_{up}$ and the last two columns/rows of $I_r/I_{low}$. Such initialization provides a good approximation to the optimal solution and speeds up the convergence of the algorithm.

An example of SA application is shown in Figure 2.10c. The leftmost columns of $I_r$ and $I_l$ appear slightly different in presence of the CSF and GM. The rightmost columns of $I_l$ and $I_r$ only differ in the value of a single central pixel. The optimal labelling $I$ preserves the label configuration of its first and last columns as they appear in their respective overlapping regions $I_l$ and $I_r$.

We started with the upper left subimage and stitched its neighbouring subimages vertically and horizontally. Then for each of the neighbouring subimages we identified overlapping ones in both horizontal and vertical directions, and glued them together using the SA algorithm. By progressive stitching of constituent image parts we assembled the entire brain slice as illustrated in Figure 2.10d. The proposed SA algorithm yields seamless estimates of the joint regions.
Figure 2.10: (a) Left and right image region overlap, (b) upper and lower image region overlap, (c) Maximum a posteriori estimate of the label configuration $I$ in presence of observations $(I_r, I_l)$, (d) Sequential steps of the stitching algorithm in direction from top left to bottom right.
3. Results and Discussion

We applied the local KFDA-based method to segment the brain template for ages 8 to 11 months initially classified into GM, WM and the CSF using PVE. The proposed approach leads to a significant improvement in the CSF detection throughout the brain almost doubling the initial CSF volume (see Figures 3.1 and 3.2). The initial CSF cluster determined by PVE consisted of 26717 voxels and has increased to 53681 voxels with application of KFDA. The

| Domain | PVE    | KFDA   | Domain | PVE    | KFDA   |
|--------|--------|--------|--------|--------|--------|
| 1      | 0.9286 | 0.9206 | 21     | 0.8481 | 0.8848 |
| 2      | 0.8963 | 0.9000 | 22     | 0.9496 | 0.9595 |
| 3      | 0.8357 | 0.8500 | 23     | 0.8890 | 0.8975 |
| 4      | 0.8682 | 0.8768 | 24     | 0.9226 | 0.9197 |
| 5      | 0.9014 | 0.8958 | 25     | 0.8769 | 0.8776 |
| 6      | 0.8776 | 0.8709 | 26     | 0.9155 | 0.9496 |
| 7      | 0.8623 | 0.8676 | 27     | 0.9602 | 0.9628 |
| 8      | 0.8850 | 0.8753 | 28     | 0.8642 | 0.8675 |
| 9      | 0.8337 | 0.8530 | 29     | 0.8978 | 0.8897 |
| 10     | 0.8874 | 0.9254 | 30     | 0.8795 | 0.8765 |
| 11     | 0.8375 | 0.8349 | 31     | 0.8261 | 0.8525 |
| 12     | 0.8129 | 0.8545 | 32     | 0.8733 | 0.8709 |
| 13     | 0.9027 | 0.9034 | 33     | 0.8722 | 0.8746 |
| 14     | 0.8928 | 0.8978 | 34     | 0.9743 | 0.9759 |
| 15     | 0.9399 | 0.9382 | 35     | 0.9135 | 0.9254 |
| 16     | 0.9033 | 0.983  | 36     | 0.8501 | 0.8581 |
| 17     | 0.9433 | 0.9416 | 37     | 0.8952 | 0.8927 |
| 18     | 0.8768 | 0.8828 | 38     | 0.8191 | 0.8379 |
| 19     | 0.8828 | 0.8878 | 39     | 0.8198 | 0.8456 |
| 20     | 0.8688 | 0.8807 | 40     | 0.9566 | 0.9652 |

Total MSSIM

0.8668 0.8758

local MSSIM assessment of the performance of KFDA and PVE given in Table 3 shows that KFDA outperforms PVE in 28 subdomains out of 40 and overall. KFDA ”looses” to PVE in lower contrast subdomains 5, 15 and 17 and 9 other subdomains within the middle CNR range presumably due to overestimation of the CSF in these subvolumes.

We also applied our segmentation method to the 3D brain template for ages 44 to 60 months with initialization obtained by a label transfer from the brain template for ages 4.5 to 8.5 years. Figure 3.3 illustrates the performance of KFDA using classified brain slices. Local MSSIM assessment of the quality of
classifications by initial label transfer, KFDA and PVE classifications suggest that KFDA improves initial classification and outperforms PVE in all but 7 brain subvolumes (Table 3). However, in high contrast subvolumes KFDA tends to overestimate WM and CSF leading to poorer structural similarity estimates. Overall, in comparison to PVE, KFDA detects brain tissue classes throughout the entire brain with a higher accuracy.
Table 3.2: Local comparison of the initial label transfer, PVE and KFDA classifications via MSSIM. Pediatric brain template for ages 44 to 60 months.

| Domain | Initial | PVE   | KFDA  | Domain | Initial | PVE   | KFDA  |
|--------|---------|-------|-------|--------|---------|-------|-------|
| 1      | 0.7809  | 0.7846| 0.8243| 12     | 0.8444  | 0.8856| 0.8718|
| 2      | 0.8051  | 0.8454| 0.8389| 13     | 0.7904  | 0.8247| 0.8320|
| 3      | 0.8405  | 0.8405| 0.8759| 14     | 0.8142  | 0.8304| 0.868 |
| 4      | 0.7815  | 0.8212| 0.8422| 15     | 0.8560  | 0.8804| 0.8825|
| 5      | 0.8230  | 0.8671| 0.8601| 16     | 0.8004  | 0.8231| 0.8467|
| 6      | 0.7422  | 0.7422| 0.8146| 17     | 0.7785  | 0.7785| 0.8292|
| 7      | 0.7071  | 0.7384| 0.7830| 18     | 0.7755  | 0.8234| 0.8479|
| 8      | 0.8266  | 0.8864| 0.8676| 19     | 0.7600  | 0.7600| 0.8373|
| 9      | 0.9257  | 0.9482| 0.9400| 20     | 0.7213  | 0.7299| 0.8058|
| 10     | 0.7925  | 0.8511| 0.8348| 21     | 0.8347  | 0.8571| 0.8747|
| 11     | 0.7529  | 0.8373| 0.8128| 22     | 0.7416  | 0.8071| 0.8251|
| means  | 0.7952  | 0.8256| 0.8458|

Figure 3.3: (a),(e) Initial CSF, GM and WM labels (MSSIM=0.8454) in (b),(f) the T1w reference subdomain (left temporal lobe near the base of the brain) from the brain template for ages 44 to 60 months and the corresponding (c),(g) PVE (MSSIM=0.8419) and (d),(h) KFDA (MSSIM=0.8911) classifications. MSSIM was computed over 14 slices of the brain subdomain enclosed in the red frame.

This example demonstrates the capability of the proposed method to identify small brain structures hardly visible in low contrast subdomains of reference images.
4. Summary

In this paper, we developed a powerful KFDA-based framework that overcomes methodological limitations of existing segmentation approaches in infant brain MRI such as global modelling of tissue intensity distributions and dependency on probabilistic brain atlas. The proposed method

- uses a general class separability criterion since it does not impose tissue intensity distribution models on the input data and captures a wide range of tissue cluster non-linearities,
- avoids a build-up of various techniques for refinement of segmentation results and accommodation of intra-class intensity variability,
- takes advantage of multi-channel brain imaging data and avoids inconsistencies that arise when segmenting each image type separately,
- classifies MR brain data locally and improves detection of spatial tissue patterns in low contrast subdomains,
- adapts initial classification to the tissue intensity distributions within an individual brain scan,
- allows comparison of classification algorithms and automatic monitoring of the quality of classification in the absence of the ground truth via the objective image quality metric SSIM.

Our framework is semi-supervised since we used tissue label transfer from an older brain MRI to a younger brain anatomy for initialization. The initialization provides the best guess on the spatial locations of tissue classes that is updated in accordance with tissue intensity histograms of brain imaging data.

We explored the potential of KFDA in applications to brain tissue classification of infant brain MRI, in particular, the NIH pediatric database. We observed that even with poor initial estimates of the class clusters in the brain template for ages 8 to 11 months KFDA compensates for the underestimates and detects most of the tissues visible in MRI. Overall, application of the KFDA-based method yields a more accurate quantification of brain tissue volumes from infant brain MRI.

The proposed method is mathematically rich and offers avenues for the further advancement of KFDA-based methodology. To mention a few,

- The classification problem in the absence of ground truth can be stated in a mathematically rigorous way, namely, as an optimization of MSSIM over a set of overlapping voxels in the stereotaxic space. The Gibbs field model of the unknown label configuration defined on this set with the Gibbs interaction energy in the form of the negative of MSSIM increment will allow optimal estimation of labelling via SA. The classification results obtained in this way do not correspond to the global maximum of MSSIM.
• A more competitive perception quality measure can be used for the evaluation of the classification quality. The information content-weighted SSIM (Wang and Qiang, 2011) is an advanced version of SSIM that assigns local information content weights to image components containing more information. Such a metric will emphasize tissue boundaries in the classified image and measure the structural similarity between classified and T1w(reference) images more accurately.

• A more general kernel function in the form of a linear combination of Gaussians with different bandwidths can be explored.

• Further understanding of variation in tissue class estimates between neighbouring subimages resulting from the choice of the width of the overlapping area may lead to seamless brain assembly (Pelletier et al., 2014), (Koen et al., 2010) thus avoiding the need for SA application.

The proposed method is applicable to brain tissue classification of multichannel brain MRI for ages 8 months and older.

5. Acknowledgements

This project was funded in whole or in part by the Montreal Neurological Institute in the form of a postdoctoral fellowship, the National Institute of Child Health and Human Development, the National Institute on Drug Abuse, the National Institute of Mental Health, and the National Institute of Neurological Disorders and Stroke (Contract #s N01-HD02-3343, N01-MH9-0002, and N01-NS-9-2314, -2315, -2316, -2317, -2319 and -2320). This manuscript expresses the views of the authors and may not reflect the opinions or views of the NIH.

Appendix A. Fundamentals of KFDA

Originally, Kernel Fisher Discriminant Analysis was proposed for a binary classification problem (Mika et al., 1999) and later it has been extended to multiclass classification. Presented below is a classical binary KFDA. Let $X$ be an input set, an arbitrary subset in $R^n$ and let $Y = \{-1, +1\}$ be class label set. We refer to an input pair $(x, y)$, where $x \in X$ and $y \in Y$ as a sample. The assumption is that all samples are drawn randomly and independently from unknown probability distribution over $X \times Y$. Let $X_- = \{x_1, x_2, ..., x_{l_1}\}_-$ and $X_+ = \{x_{l_1+1}, x_{l_1+2}, ..., x_{l_1+l_2}\}_+$ be samples from two different classes (referred to as negative and positive for a reason that will become clear later) with means $\mu_-$ and $\mu_+$. KFDA finds a non-linear direction of maximal information discrimination in the input space by first mapping the data $X_-$ and $X_+$ non-linearly into a high-dimensional feature space $\mathcal{H}$ through an implicit mapping $\phi : X \rightarrow \mathcal{H}$ and computing an optimal separating hyperplane there that corresponds to a non-linear separating surface in the input space. The power of KFDA lies in ability
to capture complex non-linear structures of clusters in the input space. To understand a kernel "trick" associated with KFDA we first formulate a linear discriminant optimality criterion. In the context of our classification problem in 3D input space, we aim to find the direction \( \vec{w} \) of input data projection \( \vec{w}^T \cdot \vec{x} \in \mathbb{R}^1 \) such that

- between-class variance \( \vec{w}^T S_B \vec{w} \) is maximized, where
  \[
  S_B = (\mu_- - \mu_+)(\mu_- - \mu_+)^T,
  \]
- within-class variance \( \vec{w}^T S_W \vec{w} \) is minimized, where
  \[
  S_W = \sum_{i \in \{-1, 1\}} \sum_{x \in X_i} (x - \mu_i)(x - \mu_i)^T.
  \]

Overall, the objective is to maximize a linear discriminant

\[
J(w) = \frac{\vec{w}^T S_B \vec{w}}{\vec{w}^T S_W \vec{w}}.
\]

To find an optimal non-linear direction of data projection, we first non-linearly transform the data with the implicit mapping \( \phi \) to the abstract feature space \( \mathcal{H} \) and compute the linear discriminant in this feature space. Then, in \( \mathcal{H} \) we have

\[
\begin{align}
\mu_-^\phi &= \frac{1}{l_1} \sum_{i=1}^{l_1} \phi(x_{i_-}), & \mu_+^\phi &= \frac{1}{l_2} \sum_{i=1}^{l_2} \phi(x_{i_+}), \\
S_B^\phi &= (\mu_-^\phi - \mu_+^\phi)(\mu_-^\phi - \mu_+^\phi)^T, \\
S_W^\phi &= \sum_{i \in \{-1, 1\}} \sum_{x \in X_i} (\phi(x) - \mu_i^\phi)(\phi(x) - \mu_i^\phi)^T.
\end{align}
\]

Thus, the linear discriminant in \( \mathcal{H} \) is

\[
J(w) = \frac{\vec{w}^T S_B^\phi \vec{w}}{\vec{w}^T S_W^\phi \vec{w}}.
\]

Since \( \mathcal{H} \) is a high-dimensional space whose dimension would be equal to the total number of interior brain voxels in applications to MR brain images it is impossible to solve [A.3] directly. We seek a formulation of the optimality criterion in terms of dot-products \( <\phi(x), \phi(z)>_H \) of the mapped samples \( x, z \in X \) since by Mercer’s theorem [Shawe-Taylor and Cristianini, 2004] we can compute these dot-products without explicit knowledge of the mapping \( \phi \) via kernel functions. The kernel \( K : X \times X \to \mathbb{R} \) is a symmetric function that defines the dot product \( K(x, z) = <\phi(x), \phi(z)>_H \) for \( \forall x, z \in X \). Possible choices for \( K \) are Gaussian radial basis function (RBF), \( K(x, z) = \exp(-\|\phi(x) - \phi(z)\|^2) \), polynomial kernels \( k(x, z) = (<\phi(x), \phi(z)>)^d \) and sigmoidal functions \( k(x, z) = \tanh(a <\phi(x), \phi(z)> + b) \).
Therefore, we make use of the kernel trick and rewrite (A.4) in computable form. Let the total number of samples be \( l = l_1 + l_2 \). Given that any solution \( w \in H \) lies in the span of training samples \( \{ \phi(x_i) \} \), \( i = 1, 2, ..., l \) (Saitoh, 1988)

\[
w = \sum_{i=1}^{l} \alpha_i \phi(x_i).
\]  

(A.5)

Using equations (A.5) and (A.1) we obtain

\[
w^T \mu_\phi = \frac{1}{l_1} \sum_{j=1}^{l_1} \sum_{k=1}^{l_2} \alpha_j K(x_j, x_{k-}) = \alpha^T M_-, \tag{A.6}
\]

where \( M_- \) is an \( l \times 1 \) vector with entries \( M_-^{ij} = \frac{1}{l_1} \sum_{k=1}^{l_1} K(x_j, x_{k-}) \). Similarly, we find

\[
w^T \mu_+ = \alpha^T M_+. \tag{A.7}
\]

By using equations (A.2), (A.6) and (A.7) the numerator of \( J(w) \) in (A.4) can be rewritten as

\[
w^T S_B \phi w = \alpha^T (M_- - M_+) (M_- - M_)^T \alpha = \alpha^T M \alpha. \tag{A.8}
\]

By using equations (A.1), (A.5) and (A.3) the denominator of \( J(w) \) in (refisher) becomes

\[
w^T S_W \phi w = \alpha^T [k_-(I - 1_{l_1})k_2^T + k_+(I - 1_{l_2})k_1^T] \alpha = \alpha^T N \alpha, \tag{A.9}
\]

where \( k_- \) and \( k_+ \) are kernel matrices of sizes \( l \times l_1 \) and \( l \times l_2 \), respectively, with entries \( (k_m)_{ij} = K(x_i, x_{jm}) \), \( m \in \{-1, +1\} \), \( I \) is the identity matrix of size \( l_n \times l_n \) and \( 1_{l_n} \) is the matrix with all entries \( \frac{1}{l_n} \), \( n = 1, 2 \). Combining (A.8) and (A.9) we aim to find a vector \( \alpha \) in \( H \) that maximizes

\[
J(\alpha) = \frac{\alpha^T M \alpha}{\alpha^T (N + \beta I) \alpha}. \tag{A.10}
\]

Here, \( \beta \) is a small positive regularization parameter added to the variance entries of the within-class variance-covariance matrix \( N \) to ensure its positive-definiteness since \( N \) may be singular. The decision boundary between the classes in the input space is a set of points that satisfies

\[
h(x) = \sum_{i=1}^{l} \alpha_i K(x_i, x) + b = 0, \text{ where } b \text{ is an offset.} \tag{A.11}
\]

Since \( h(x) \) depends on the kernel choice, the classification result will also be kernel-dependent. As such, it is important to choose the kernel function that best captures a non-linear behaviour of class clusters based on their natural occurrences in the input space.
References

Adams, R. and Bischof, L. (1994). Seeded region growing. *IEEE Trans. Pattern Analysis Machine Intelligence*, 16(6):641–647.

Almli, C., Rivkin, M., and McKinstry, R. B. D. C. G. (2007). The nih mri study of normal brain development (objective-2): Newborns, infants, toddlers, and preschoolers. *NeuroImage*, 35(1):308–325.

Amunts, K., Lepage, C., Borgeat, L., and et al. (2013). Bigbrain: An ultrahigh-resolution 3d human brain mode. *Science*, 340(1472):1472–1475.

Ashburner, J. and Friston, K. (1997). Multimodal image coregistration and partitioning a unified framework. *NeuroImage*, 6:209–217.

Bouix, S., Martin-Fernandez, M., Ungar, L., and et al. (2007). On evaluating brain tissue classifiers without a ground truth. *NeuroImage*, 36:1207–1224.

Choi, H., Haynor, D., and Kim, Y. (1991). Partial volume tissue classification of multichannel magnetic resonance images-a mixel model. *IEEE Transactions on Medical Imaging*, 10(3):395–407.

Cocosco, C., Zijdenbos, A., and Evans, A. (2003). A fully automatic and robust brain mri tissue classification method. *Medical Image Analysis*, 7(4):513–527.

Collins, D., Neelin, P., Peters, T., and Evans, A. (1994). Automatic 3d inter-subject registration of mr volumetric data in standardized talairach space. *J Comput Assist Tomogr.*, 18(2):192–205.

Collins, L., Zijdenbos, A., Baarè, W., and Evans, A. (1999). Animal+ insect: improved cortical structure segmentation. In *Information Processing in Medical Imaging*, pages 210–223. Springer, Berlin Heidelberg.

Evans, A. B. D. C. G. (2006). The nih mri study of normal brain development. *NeuroImage*, 30:184–202.

Evans, A., Janke, A., Collins, L., and Baillet (2012). Brain templates and atlases. *NeuroImage*, 62:911–922.

Evans, A., Kamber, M., Collins, L., and Macdonald, D. (1994). An mri-based probabilistic atlas of neuroanatomy. In *Magnetic Resonance Scanning and Epilepsy Plenum*, NATO ASI Series, pages 263–274. Springer, New York.

Fonov, V., Evans, A., and Botteron, K. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1):313–327.

Grenander, U. (1996). *Elements of Pattern Theory*. The Johns Hopkins University Press, Baltimore, Maryland.

Kim, S.-J., Magnani, A., and Boyd, S. (2006). Optimal kernel selection in kernel fisher discriminant analysis. In *ICML 2006 Proceedings of the 23rd international Conference on Machine learning*, pages 465–472.
Koen, E., Garroway, C., Wilson, P., and Bowman, J. (2010). The effect of map boundary on estimates of landscape resistance to animal movement, online publication. *PLoS ONE*, 5(7).

Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Consell, S., Doria, V., Serag, A., Gousias, I., Boardman, J., Rutherford, M., Edwards, A., Hajnal, J., and Rueckert, D. (2011). A dynamic 4d probabilistic atlas of the developing brain. *NeuroImage*, 54(4):2750–2763.

MacDonald, D., Kabani, N., Avis, D., and Evans, A. (2000). Automated 3-d extraction of inner and outer surfaces of cerebral cortex from mri. *NeuroImage*, 12:340–356.

Mika, S., G., R., J., W., and et al. (1999). Fisher discriminant analysis with kernels. In *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pages 41–48.

Murgasova, M., Dyet, L., Edwards, D., Rutherford, M., Hajnal, J., and Rueckert, D. (2007). Segmentation of brain mri in young children. *Academic Radiology*, 14(11):1350–1366.

Nyúl László, G., Udupa, J., and Zhang, X. (2000). New variants of a method of mri scale standardization. *IEEE Transactions on Medical Imaging*, 19(2):143–150.

Pappas, T. N. (1992). An adaptive clustering algorithm for image segmentation. *IEEE Transactions on Signal Processing*, 40(4):901–914.

Pelletier, D., Clark, M., Anderson, M., Rayfield, B., Wulder, M., and et al. (2014). Applying circuit theory for corridor expansion and management at regional scales: Tiling, pinch points, and omnidirectional connectivity, online publication. *PLoS ONE*, 9(1).

Portman, N. and Evans, A. (2013). Novel vector-valued approach to automatic brain tissue classification. In *Medical Computer Vision, Recognition Techniques and Applications in Medical Imaging*, pages 70–81, Berlin Heidelberg. Springer.

Prastawa, M., Gilmore, J., Lin, W., andGerig, G. (2005). Automatic segmentation of mr images of the developing newborn brain. *Medical Image Analysis (MedIA)*, 9(5):457–466.

Rätsch, G., Onoda, T., and Müller, K.-R. (2001). Soft margins for adaboost. *Machine Learning*, 42:287–320.

Rigau, J., Feixas, M., Sbert, M., Bardera, A., and Boada, I. (2004). Medical image segmentation based on mutual information maximization. In *7th Int. Con. on Medical Image Computing and Computer Assisted Intervention*, volume 3216 of *Lecture Notes in Computer Science*, pages 135–142.
Saitoh, S. (1988). *Theory of Reproducing Kernels and its Applications*. Longman Scientific & Technical, Harlow, England.

Seghier, M., RamLackhansingh, A., Crinion, J., Leff, A., and Price, C. (2008). Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *NeuroImage*, 41(4-3):1253–1266.

Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge University Press, Cambridge.

Sled, J., Zijdenbos, A., and Evans, A. (1998). A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Transactions on Medical Imaging*, 17(1):87–97.

Smith, S. (2002). Fast robust automated brain extraction. *Human brain mapping*, 17(3):143–155.

Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., Bannister, P., De Luca, M., Drobnjak, I., Flitney, D., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J., and Matthews, P. (2004). Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage*, 23(S1):208–219.

Tohka, J., Zijdenbos, A., and Evans, A. (2004). Fast and robust parameter estimation for statistical partial volume models in brain mri. *NeuroImage*, 23(1):84–97.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999). Automated model-based tissue classification of mr images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908.

Vapnik, N. (1998). *Statistical learning theory*. Wiley, New York.

Wang, Z. and Bovik, A. (2009). Mean squared error: love it or leave it? a new look at signal fidelity measures. *Signal Processing Magazine, IEEE*, 26(1):98–117.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Wang, Z. and Qiang, L. (2011). Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20(5):1185–1198.

Wang, Z., Sheikh, H., and Bovik, A. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84.

Wang, Z., Simoncelli, E., and Bovik, A. (2003). Multiscale structural similarity for image quality assessment. In *Conference Record of the 37th Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1398–1402.
Warfield, S., Kaus, M., Jolesz, F., and Kikinis, R. (2000). Adaptive, template moderated, spatially varying statistical classification. *Medical Image Analysis*, 4(1):43–55.

Weisenfeld, N. and Warfield, S. (2009). Automatic segmentation of newborn brain mri. *Neuroimage*, 47(2):564–572.

Wells, W., Grimson, W., Kikinis, R., and Jolesz, F. (1996). Adaptive segmentation of mri data. *IEEE Transactions on Medical Imaging*, 15(4):429–441.

Xue, H., Srinivasan, L., Jianga, S., Rutherforda, M., Edwardsa, A., Rueckert, D., and Hajnal, J. (2007). Automatic segmentation and reconstruction of the cortex from neonatal mri. *NeuroImage*, 38(3):461–477.

Zijdenbos, A., Dawant, B., Margolin, R., and Palmer, A. (1994). Morphometric analysis of white matter lesions in mri images: method and validation. *IEEE Transactions on Medical Imaging*, 13(4):716–724.