Detection of corpus callosum malformations in pediatric population using the discriminative direction in multiple kernel learning

Denis Peruzzo[1,2]⋆, Filippo Arrigoni[1], Fabio Triulzi[1,3], Cecilia Parazzini[4], and Umberto Castellani[2];

[1] Scientific Institute IRCCS "Eugenio Medea", Bosisio Parini (LC), Italy
[2] University of Verona, Verona (VR), Italy
[3] Fondazione IRCCS "Ca' Granda", Ospedale Maggiore Policlinico, Milan (MI), Italy
[4] Children Hospital "Vittore Buzzi", Milan (MI), Italy;

Abstract. In this paper we propose a Multiple Kernel Learning (MKL) classifier to detect malformations of the corpus callosum (CC) and apply it in a pediatric population. Furthermore, we extend the concept of *discriminative direction* to the linear MKL methods, implementing it in a single subject analysis framework.
The CC is characterized using different measures derived from Magnetic Resonance Imaging (MRI) data and the MKL approach is used to efficiently combine them. The discriminative direction analysis highlights those features that lead the classification for each given subject. In the case of a CC with malformation this means highlighting the abnormal characteristics of the CC that guide the diagnosis.
Experiments show that the method correctly identifies the malformative aspects of the CC. Moreover, it is able to identify dishomogeneus, localized or widespread abnormalities among the different features.
The proposed method is therefore suitable for supporting neuroradiologists in the decision-making process, providing them not only with a suggested diagnosis, but also with a description of the pathology.

Keywords: magnetic resonance imaging, multiple kernel learning, brain imaging, computer-aided diagnosis.

1 Introduction

The Corpus Callosum (CC) is a thick white-matter (WM) bundle, made of myelinated axons, connecting homotopic cortical regions of the two cerebral hemispheres. It is the largest commissural structure, playing a fundamental role both in motor and cognitive functions. On the midsagittal cerebral plane, it shows a broad-arched shape, with variable thickness along its major axis. A wide range of malformative patterns (i.e. complete or partial agenesis, hypoplasia, thinning, thickening) can affect the CC and are frequently associated with different rate of mental retardation and motor impairment.

⋆ Corresponding author.

The diagnosis of CC malformations is currently made with Magnetic Resonance Imaging (MRI) by neuroradiologists trained to evaluate its shape. Several features (i.e. area, length, curvature, thickness) are visually inspected in order to detect possible alterations and malformative patterns. Although useful, such approach suffers from the limitation of being subjective and only qualitative.

Automatic classifiers, such as Support Vector Machine (SVM), can be used to support the diagnostic task providing mathematical tools to describe and classify the CC. In particular, Multiple Kernel Learning (MKL) methods [1] efficiently combine different descriptor functions, thus including all the different aspects of the CC usually taken into account by the neuroradiologist. However, automatic classifiers are usually developed as "black box" tools, that provide the subject classification, but not a description or a list of the aspects that led the choice.

Several works have been proposed in literature to highlight and analyze the discriminative information. They are usually based on the analysis of the separating hyperplane [6], on the kernel weight analysis for MKL [2], or on the analysis of the neighborhood points in the feature or sampling space [4]. The common aspect of these works is that they derive group differences. On the contrary, starting from the work of Golland and colleagues [4] about the *discriminative direction* analysis, we aim to develop a method tailored on the single subject, able to highlight the specific features that guide the classification.

In this work, we use a MKL classifier to automatically discriminate between malformed and normal CC. The MKL approach allows to efficiently combine different CC measures, as performed by neuroradiologists. In particular, we apply the Group-Lasso MKL method [7], thus incorporating a feature selection procedure in the classifier training. Finally, we extend the definition of discriminative direction to the linear MKL methods and exploit it not only to provide the subject diagnosis, but also a quantitative description of the CC malformations.

2 SVM and MKL

Given a set of observation $\boldsymbol{x}_i \in \mathbb{R}^n$ and labels $l_i \in \{-1;+1\}$, and a kernel function $K : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ the SVM methods build a classifier by implicitly mapping the training data into a higher dimensional space and determining the linear classifier that maximize the margin between the two labels. The separating hyperplane w and the decision function $f(\boldsymbol{x})$ can be derived from the training data:

$$\boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \Phi_K(\boldsymbol{x}_i) \qquad f(\boldsymbol{x}) = \langle \boldsymbol{x} \cdot \boldsymbol{w} \rangle = \sum_{i=1}^{N} \alpha_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{1}$$

where the function $\Phi_K : \mathbb{R}^n \mapsto \mathbb{F}$ is the mapping function implicitly defined by the kernel function K and the coefficients α_i are computed while maximizing the classifier margin.

The MKL technique [1] allows to deal with multiple features and/or multiple kernels in the same classifier. The original kernel function $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is replaced by a new one $K_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\eta})$, which is obtained combining different kernels. In particular, for the *linear*-MKL methods its general formulation is [1]:

$$K_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\eta}) = \sum_{m=1}^{P} \eta_m K_m(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m) \tag{2}$$

where P is the number of different kernels and $\boldsymbol{\eta} \in \mathbb{R}^P$ contains the weight associated to each kernel. In this conditions, the separating hyperplane and the decision function become the linear combination of the individual kernel ones:

$$\boldsymbol{w}_\eta = \sum_{i=1}^{N} \alpha_i y_i \Phi_{K_{\boldsymbol{\eta}}}(\boldsymbol{x}_i; \boldsymbol{\eta}) = \sum_{m=1}^{P} \eta_m \cdot \boldsymbol{w}_m \tag{3}$$

$$f_\eta(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i y_i \sum_{m=1}^{P} \eta_m K_m(\boldsymbol{x}_i^m, \boldsymbol{x}^m) + b = \sum_{m=1}^{P} \eta_m f_m(\boldsymbol{x}^m) \tag{4}$$

In this study we employ the *Group Lasso* MKL (GL-MKL) method [7]. Its advantage is that it allows an explicit tuning of the sparsity effect on the involved weights by introducing a dedicated parameter p in model. When $p = 1$ a *competitive* approach is introduced among features by emphasizing the selection effect of MKL. Conversely, when $p = 2$ a *cooperative* approach is encouraged by exploiting complementary information.

2.1 Discriminative direction for MKL

According to the original definition given in [4], the *discriminative direction* is the direction (in the sample space) that affects the output of the classifier while introducing as little irrelevant change as possible into the input vector.

Given a point in the original space $\boldsymbol{x}$, the discriminative direction $d\boldsymbol{x}$ introduces a displacement in the projection space $\mathbb{F}$ defined as $d\boldsymbol{z} = \Phi_K(\boldsymbol{x} + d\boldsymbol{x}) - \Phi_K(\boldsymbol{x})$. The displacement can be decomposed into two components: its projection onto $\boldsymbol{w}$ and its deviation from $\boldsymbol{w}$:

$$\boldsymbol{p} = \frac{\langle d\boldsymbol{z} \cdot \boldsymbol{w} \rangle}{\langle \boldsymbol{w} \cdot \boldsymbol{w} \rangle} \cdot \boldsymbol{w} \qquad \boldsymbol{e} = d\boldsymbol{z} - \boldsymbol{p} = d\boldsymbol{z} - \frac{\langle d\boldsymbol{z} \cdot \boldsymbol{w} \rangle}{\langle \boldsymbol{w} \cdot \boldsymbol{w} \rangle} \cdot \boldsymbol{w} \tag{5}$$

Formally, the *discriminative direction* minimizes the divergence component e, leading to the following optimization problem:

$$\begin{aligned} &\mathit{minimize}\ \mathcal{E}(d\boldsymbol{x}) = \|\boldsymbol{e}\|^2 = \langle d\boldsymbol{z} \cdot d\boldsymbol{z} \rangle - \tfrac{\langle d\boldsymbol{z} \cdot \boldsymbol{w} \rangle^2}{\langle \boldsymbol{w} \cdot \boldsymbol{w} \rangle} \\ &\mathit{with}\ \|d\boldsymbol{x}\|^2 = \epsilon \end{aligned} \tag{6}$$

The elements of the optimization problem can be computed using the MKL kernel function introduced in Equation 2:

$$
\begin{aligned}
\langle \boldsymbol{w}_\eta \cdot \boldsymbol{w}_\eta \rangle &= \sum_{m=1}^{P} \eta_m \cdot \langle \boldsymbol{w}^m \cdot \boldsymbol{w}^m \rangle \\
\langle d\boldsymbol{z} \cdot \boldsymbol{w}_\eta \rangle &= \nabla f_{K_\eta}(\boldsymbol{x}) d\boldsymbol{x} = \sum_{m=1}^{P} \eta_m \cdot \nabla f_m(\boldsymbol{x}^m) \\
\langle d\boldsymbol{z} \cdot d\boldsymbol{z} \rangle &= d\boldsymbol{x}^T H_{K_\eta}(\boldsymbol{x}) d\boldsymbol{x}
\end{aligned}
\tag{7}
$$

Where $\nabla f_{K_\eta}(\boldsymbol{x})$ contains the gradient of the decision function f_{K_η} and $H_{K_\eta}(\boldsymbol{x})$ is one of the equivalent off-diagonal quarters of the Hessian of K_η.

$\nabla f_{K_\eta}(\boldsymbol{x})$ is formally computed as the linear combination of the gradients of the single kernel decision functions $\nabla f_m(x^m)$. However, if the single kernels are independent of each other (i.e. they operate on separate components of the input vector $\boldsymbol{x}$), then the gradient of the whole decision function can be seen as a concatenation of the gradients of the single ones. Following the same procedure, the matrix $H_{K_\eta}(\boldsymbol{x})$ is a diagonal matrix where the blocks on the main diagonal contain the single kernel $H_m(\boldsymbol{x}^m)$ matrices and the off-diagonal blocks are all zero due to the reciprocal independence among kernels.

Substituting the elements in the eq. 6, the following problem relative to the MKL discriminative direction is obtained

$$
\begin{aligned}
&minimize\ \mathcal{E}(d\boldsymbol{x}) = d\boldsymbol{x}^T (H_{K_\eta}(\boldsymbol{x}) - \|\boldsymbol{w}\|^{-2} \nabla f_{K_\eta}^T \nabla f_{K_\eta}) d\boldsymbol{x} \\
&\quad with\ \|d\boldsymbol{x}\|^2 = \epsilon
\end{aligned}
\tag{8}
$$

its solution is the smallest eigenvector of matrix:

$$
Q_K(\boldsymbol{x}) = H_{K_\eta}(\boldsymbol{x}) - \|\boldsymbol{w}\|^{-2} \nabla f_{K_\eta}^T \nabla f_{K_\eta} \tag{9}
$$

As in the original formulation, an analytical solution can be derived when the matrix $H_{K_\eta}(\boldsymbol{x})$ is a multiple of the identity matrix $H_{K_\eta}(\boldsymbol{x}) = cI$. Given the structure of $H_{K_\eta}(\boldsymbol{x})$, this condition is satisfied when the matrix $H_m(\boldsymbol{x}^m)$ of each kernel satisfies the same condition, i.e. if an analytical solution exist independently for each kernel included in the MKL classifier. In such case the discriminative direction is obtained concatenating the solutions computed for each kernel weighted by the corresponding kernel weights.

$$
\begin{aligned}
d\boldsymbol{x}^* &= [\eta_1 \cdot d\boldsymbol{x}_1^{*T}, \eta_2 \cdot d\boldsymbol{x}_2^{*T}, .., \eta_P \cdot d\boldsymbol{x}_P^{*T}]^T \\
\epsilon(d\boldsymbol{x}^*) &= [\eta_1 \cdot \epsilon(d\boldsymbol{x}_1^*)^T, \eta_2 \cdot \epsilon(d\boldsymbol{x}_2^*)^T, .., \eta_P \cdot \epsilon(d\boldsymbol{x}_P^*)^T]^T
\end{aligned}
\tag{10}
$$

Analytical solutions for the discriminative direction problem have been provided in [4] for linear and radial basis function (rbf) kernels. In the case of linear kernel the solution is exactly the hyperplane vector $\boldsymbol{w}$, whereas for the rbf kernel the solution is:

$$
\begin{aligned}
d\boldsymbol{x}^* &= -\frac{2}{\gamma}\sum_{i=1}^{N}\alpha_i y_i e^{-\frac{\|\boldsymbol{x}-\boldsymbol{x}_i\|^2}{\gamma}}(\boldsymbol{x}-\boldsymbol{x}_i) \\
\epsilon(d\boldsymbol{x}^*) &= \frac{2}{\gamma} - \frac{\|\nabla f_K^T(\boldsymbol{x})\|^2}{\|\boldsymbol{w}\|^2}
\end{aligned}
\tag{11}
$$

It is important to highlight that a zero error can not be achieved in the rbf kernel case because it has no corresponding solution in the original space.

The discriminative direction can be used to draw a trajectory in the sampling space that moves a given subject from one group subspace to the other one. The integral of the trajectory provides a quantitative description of feature changes required for that sample to change its classification. In our case, the integral of the trajectory gives a measure of the degree of the malformation for each features, allowing to infer not only on the diagnosis, nut also on the malformation degree of the given subject.

Extended version of the original discriminative direction have been proposed in literature to account for the underlying manifold structure of the shape descriptor spaces [8] and their spatial relations [3]. MKL discriminative direction can be extended also to these methods.

3 Materials and methods

The dataset includes 104 children (mean age and SD = 7.1 ± 2.2 years, male/female = 68/36). The study was approved by the local Ethics Committee , the research institute IRCCS "Eugenio Medea" (Bosisio Parini, Italy), and parents of all the participants provide their signed informed consent. Two experienced pediatric neuroradiologists reviewed in consensus the MRI exams and identified 52 subjects with a malformed CC and 52 subjects with a normal shaped CC.

MRI scans were acquired using a 3.0T Philips Achieva scanner with a 32 channels head coil. For CC evaluations and measurements a 3D fast field echo T1-weighted sequence covering the entire brain was used (TR = 8.2 ms, TE = 3.8 ms, flip angle = 8, FOV = 210 x 210 mm^2, matrix size = 210 x 210 x 170, voxel size 1 x 1 x 1 mm^3, turbo factor = 210, TI = 910 ms).

Images are pre-processed using the FMRIB Software Library (FSL). Briefly, they are corrected for the magnetic field inhomogeneities, brains are extracted, rigidly registered to the MNI atlas and the WM probability maps are computed.

The CC mask is extracted using an automatic segmentation method proposed in [5]. From the CC mask, a 2D perimeter is computed using a $\beta - spline$ model on the midsaggital slice. Two well known anatomical markers are subsequently defined on the perimeter and identified as the points with the highest perimeter curvature. The first marker is the rostrum vertex and it is localized in the anterior part of the CC, the second one is the splenium vertex and it is in the posterior one. The two anatomical vertices naturally divide the perimeter into two sections, a superior and an inferior one. Subsequently, the CC skeleton is defined as the line connecting the two vertices and equidistant from the

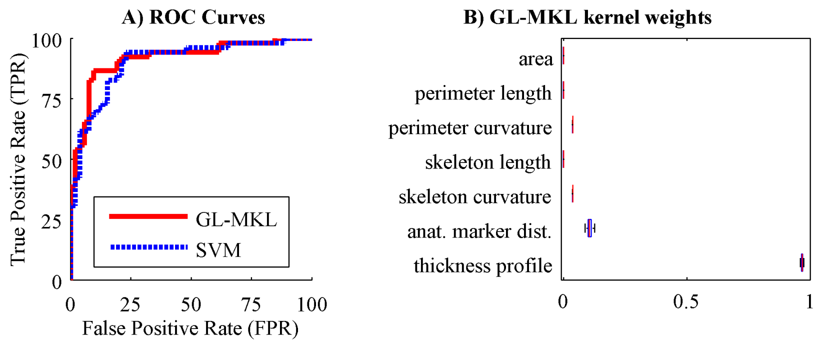

Fig. 1. A) Receiver operating characteristic of the classifiers. B) kernel weight estimates obtained from the GL-MKL classifier with the leave-one-out procedure.

two perimeter sections. The local CC thickness is computed as the minimum distance between the superior and the inferior sections of the perimeter measured on a straight line passing through a given point. The thickness profile is obtained measuring the thickness on fifty points uniformly distributed along the skeleton. At the end of the image processing the CC is characterized by the following measures, reflecting the aspects usually considered by the neuroradiologists: area, perimeter length, perimeter curvature, skeleton length, skeleton curvature, distance between anatomical markers and thickness profile. We compared different pipelines and software (eg: ANTs, FSL, SPM, CCseg, C8) for the CC extraction and characterization. We present here the results obtained with the most robust approach, especially in subjects with malformed CC.

A GL-MKL classifier is built using a different rbf kernel for each measure of the CC. The classifier is trained to separate subjects into the two groups (with or without CC malformations) identified by the neuroradiologists. A standard SVM classifier with rbf kernel is also built for comparison purpose. In this case, all features are concatenated to build a single feature table. The parameter optimization for both classifiers is performed with the grid search approach, whereas performances are evaluated using the leave-one-out procedure.

4 Results and discussion

GL-MKL performs better than the standard SVM both in terms of overall accuracy (87.5% versus 83.7%) and area under the Receiver Operating Characteristic (ROC) curve (91.9 versus 90.1), see figure 1-A.

One of the GL-MKL advantages is the possibility to tune the weights associated to the kernels, thus performing an implicit feature selection procedure. In our case, each kernel is associated to a different kind of measure, thus an analysis of the kernel weights allows to infer the discriminative power of the different measures. Figure 1-B reports the kernel weights computed using GL-MKL. The thickness profile shows the largest weight, so it contains the largest part of

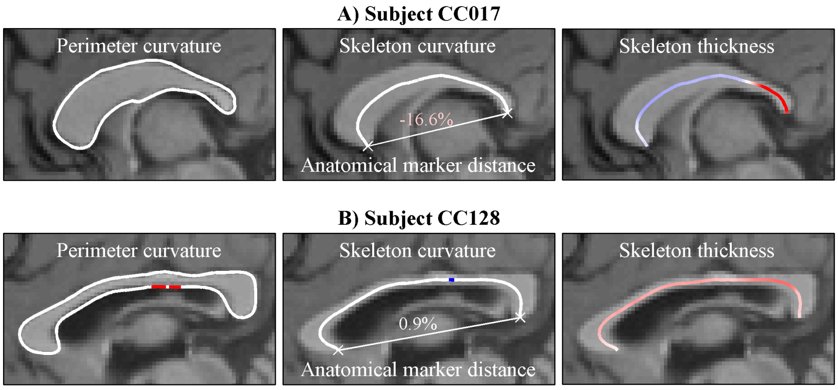

Fig. 2. Examples of the discriminative direction analysis in two subjects with abnormal CC for the four features with non-zero kernel weight. Red-blue color-code is used to indicate whether the feature is greater (blue) or smaller (red) than expected. Color intensity is proportional to the grade of the malformation. The "x" markers in the middle panels report the rostrum and the splenium of the CC.

the discriminative information. Other useful measures are the distance between the anatomical markers and the perimeter/skeleton curvatures. Conversely, area and skeleton/perimeter lengths are associated to a zero kernel weight. However, this does not mean that such measures contain no discriminative information, but that, if present, it is redundant compared to the other measures. In this experiment the best accuracy is obtained with the sparsity parameter p set to 1.5, reaching a trade off between the cooperative and the competitive approach among kernels. Therefore, a zero weight associated to a specific kernel only implies that that kernel does not impact the classifier decision function.

We perform the discriminative direction analysis to identify the specific abnormalities of the CC that guided the classification. Two examples are reported in figure 2. In the first example (subject CC017) the discriminative direction analysis highlights an irregular thickness profile and a short distance between the anatomical markers. In particular, the thickness profile is larger than expected in the anterior and middle portion of CC, whereas it is much smaller than expected in the posterior part. This example shows how the discriminative direction analysis can detect dishomogeneus abnormal aspects even inside the same kernel. Moreover, the color intensity is modulated using the integral of the discriminative direction, giving a quantitative and easily interpretable description of the abnormal features. In subject CC017 the thinning of the CC profile in the posterior region appears more severe compared to its anterior thickening or to the distance reduction between the anatomical markers. This can be seen also in the second example (subject CC128), where the color intensity shows that the thinning of the profile is more severe posteriorly than in the other parts of the CC. In this subject the perimeter and skeleton curvatures appear also focally irregular, indicating that the CC should be more arched than what it looks like. This example shows that the discriminative direction analysis can detect both widespread and localized malformations of the CC.

5 Conclusions

In this study we extend the concept of discriminative direction to the linear MKL methods and apply it to a GL-MKL classifier which discriminate between normal and malformed CC.

The GL-MKL approach allows a natural and efficient integration among different measures derived from the MRI images, reproducing the neuroradiologist approach. Moreover, the kernel weight optimization performs an implicit feature selection that discard redundant features and improve the classification accuracy.

The discriminative direction analysis highlights the discriminative features for each single subject, thus providing a qualitative and quantitative description of the key aspects of the CC abnormal shape. This is a powerful approach in comparison to the common literature studies, which study the discriminative information to infer group differences rather than focusing on the single subject.

This opens new perspective in the personalized medicine field, providing a tool for the detection and analysis of the peculiar features of the single subject.

References

1. Francis R. Bach, Gert R. G. Lanckriet, and Michael I. Jordan. Multiple kernel learning, conic duality, and the smo algorithm. In *Proceedings of the 21st International conference on Machine learning, ICML*, pages 41–48, 2004.
2. Eduardo Castro, Martínez-Rámon, Godfrey Pearlson, Jing Sui, and Vince D. Calhoun. Characterization of groups using composite kernels and multi-source fmri analysis data: application to schizophrenia. *Neuroimage*, 58:526–536, 2011.
3. Remi Cuingnet, Joan Alexis Galunes, Marie Chupin, Habib Benali, Oliver Colliot, and The Alzheimers Disease Neuroimaging Initiative. Spatial and anatomical regularization of svm a general framework for neuroimaging data. *IEEE transaction on medical imaging*, 35(3):682–696, 2013.
4. Polina Golland, W. Eric L. Girmson, Martha E. Shenton, and Ron Kikinis. Detection and analysis of statistical differencies in anatomical shape. *Medical Image Analysis*, 9:69–86, 2005.
5. Timothy J. Herron, Xiaojian Kang, and David L. Woods. Automated measurement of the human corpus callosum using mri. *Frontiers in Neuroinformatics*, 6(25):1–15, 2012.
6. João R. Sato, Carlos E. Thomaz, Ellison F. Cardoso, André Fujita, Maria da Graça Martin, and Edson Jr Amaro. Hyperplane navigation: a method to set individual scores in fmri group datasets. *Neuroimage*, 42:1473–1480, 2008.
7. Zenglin Xu, Rong Jin, Haiqin Yang, Irwin King, and Michael R. Lyu. Simple and efficient multiple kernel learning by group lasso. In *Proceedings of the 27th International conference on Machine learning, ICML*, pages 1175–1182, 2010.
8. Luping Zhou, Richard Hartley, Lei Wang, Paulette Lieby, and Nick Barnes. Identifying anatomical shape difference by regularized discriminative direction. *IEEE transaction on medical imaging*, 28(6):937–950, 2009.