Discriminant Analysis for Multiway Data

Gisela Lechuga, Laurent Le Brusquet, Vincent Perlbarg, Louis Puybasset, Damien Galanaud, Arthur Tenenhaus

Discriminant Analysis for Multiway Data

Lechuga G., Le Brusquet L., Perlbarg V., Puybasset L., Galanaud D. and Tenenhaus A.

Abstract A multiway Fisher Discriminant Analysis (MFDA) formulation is presented in this paper. The core of MFDA relies on the structural constraint imposed to the discriminant vectors in order to account for the multiway structure of the data. This results in a more parsimonious model than that of Fisher Discriminant Analysis (FDA) performed on the unfolded data table. Moreover, computational and overfitting issues that occur with high dimensional data are better controlled. MFDA is applied to predict the long term recovery of patients after traumatic brain injury from multi-modal brain Magnetic Resonance Imaging. As compared to FDA, MFDA clearly tracks down the discrimination areas within the white matter region of the brain and provides a ranking of the contribution of the neuroimaging modalities. Based on cross validation, the accuracy of MFDA is equal to 77% against 75% for FDA.

Lechuga G.
Supélec Sciences des Systèmes - EA4454 (E3S), Gif-sur-Yvette, France
e-mail: gisela.lechuga@supelec.fr

Le Brusquet L.
Supélec Sciences des Systèmes - EA4454 (E3S), Gif-sur-Yvette, France
e-mail: laurent.lebrusquet@supelec.fr

Perlbarg V.
Bioinformatics/Biostatistics Platform IHU-A-ICM, Brain and Spine Institute 47-83, Bd de l'hôpital; Paris, France e-mail: vperlbar@imed.jussieu.fr

Puybasset L.
AP-HP, Pitié-Salpêtrière Hospital, Surgical Neuro-Intensive Care Unit, Paris, France
e-mail: louis.puybasset@psl.aphp.fr

Galanaud D.
AP-HP, Pitié-Salpêtrière Hospital, Department of Neuroradiology, Paris, France
e-mail: galanaud@gmail.com

Tenenhaus A.
EA4454 (E3S) and Bioinformatics/Biostatistics Platform IHU-A-ICM, Paris, France
e-mail: arthur.tenenhaus@supelec.fr

1 Introduction

In standard multivariate data analysis, an individuals $\times$ variables data table is usually considered. However, from a practical viewpoint this simple data structure appears to be somehow restricted. An example is found in multi-modal brain Magnetic Resonance Imaging (MRI) where K neuroimaging modalities (each characterized by J voxels), are collected on a set of I patients. In that context, an individuals $\times$ voxels $\times$ modalities data table can be considered and yields a three-way dataset (or tensor). A three-way dataset can be considered in terms of a stack of matrices as illustrated in Figure 1. Most data analysis methods in their primary definition do not take into account this natural three-way structure. Indeed, such structure is lost by considering a $I \times JK$ unfolded version leading potentially (i) to a procedure that destroys the integrity of the structure of the data and (ii) to a very large parameter vector to estimate. These two aspects can yield a lack of relevant interpretations of the resulting model and additional structural constraints are required.

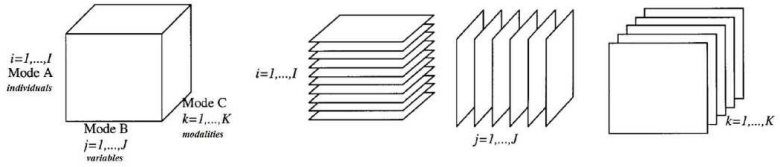


Fig. 1 Structure of the three-way dataset

Many two-way data analysis methods have been extended to the multiway configuration. For instance PARAFAC proposed by Harshman [4] is a generalization of Principal Component Analysis. PARAFAC relies on the maximization of a variance criterion but explicitly takes into account the multiway structure of the input data by imposing a special Kronecker structure on the weight vectors. A second approach is the Multi-linear Partial Least Squares Regression (N-PLS) proposed by Bro [1] which is a generalization of the classic PLS regression method to multiway data. N-PLS relies on the maximization of a covariance criterion but has the same PARAFAC structural constraint on the weight vectors. N-PLS relies on SVD decomposition and is particularly well suited to the high dimensional setting. In this paper, a multiway formulation of Fisher Discriminant Analysis (MFDA) is presented. The structural constraint that is imposed to N-PLS and PARAFAC weight vectors constitutes the starting point of MFDA.

This paper is organized as follows: Fisher Discriminant Analysis (FDA) and its multiway counterpart (MFDA) are presented in section 2. In section 3, MFDA is illustrated on a multi-modal Magnetic Resonance Brain Imaging (MRI) dataset in order to predict the long-term recovery, of patients that suffered from traumatic brain injury. A comparison between MFDA and FDA is discussed in section 4.

2 Multiway Fisher Discriminant Analysis

Let $\mathbf{X}$ be the individuals $\times$ variables $\times$ modalities tensor and $\mathbf{X}^{\mathrm{u}}$ the associated unfolded matrix where all the $I \times J$ two-way matrices are collected next to each other in an $I \times JK$ matrix. In addition, let $\mathbf{y}$ be the qualitative variable that encodes the class membership of each individual. Let $\mathbf{Y}$ be the matrix of dummy variables indicating the group memberships.

Regularized Fisher Discriminant Analysis. FDA consists in finding a projection vector $\mathbf{w}$ such that the between class variance is maximized relative to the within-class variance. Regularized FDA is defined by the optimization problem:

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^\top \mathbf{S}_B \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_T \mathbf{w} + \lambda \mathbf{w}^\top \mathbf{R} \mathbf{w}}, \tag{1}$$

where $\mathbf{S}_B = (\mathbf{X}^{\mathrm{u}})^\top \mathbf{Y}(\mathbf{Y}^\top \mathbf{Y})^{-1} \mathbf{Y}^\top \mathbf{X}^{\mathrm{u}} = (\mathbf{X}^{\mathrm{u}})^\top \mathbf{H}_B \mathbf{X}^{\mathrm{u}}$ is the between covariance matrix and $\mathbf{S}_T = (\mathbf{X}^{\mathrm{u}})^\top \mathbf{X}^{\mathrm{u}}$ is the total covariance matrix. A regularization term $\lambda \mathbf{w}^\top \mathbf{R} \mathbf{w}$ is added to improve the numerical stability when computing the inverse of $\mathbf{S}_T$ in high dimensional setting ($I \ll JK$). $\mathbf{R}$ is usually equal to the identity.
$\mathbf{w}^*$ is obtained as the first eigenvector of $(\mathbf{S}_T + \lambda \mathbf{R})^{-1} \mathbf{S}_B$.

Additional structural constraints should be added to the optimization problem (1) in order to account for the three-way structure of the data.

MFDA criterion. Structural constraints are imposed in such way that the weight vector $\mathbf{w}$ will be decomposed in two vectors as $\mathbf{w} = \mathbf{w}_K \otimes \mathbf{w}_J$. $\mathbf{w}_K$ is a weight vector associated with the K modalities while $\mathbf{w}_J$ is the weight vector related to the J variables. This structural constraint results in a more parsimonious model ($J + K$ instead of $J \times K$ parameters to estimate), and allows to study separately the effects of the variables and the modalities. A possible reformulation of FDA that takes into account the three-way structure of the data is introduced through the optimization problem:

$$\arg\max_{\mathbf{w}} \frac{(\mathbf{w}_K \otimes \mathbf{w}_J)^\top \mathbf{S}_B (\mathbf{w}_K \otimes \mathbf{w}_J)}{(\mathbf{w}_K \otimes \mathbf{w}_J)^\top \mathbf{S}_T (\mathbf{w}_K \otimes \mathbf{w}_J) + \lambda (\mathbf{w}_K \otimes \mathbf{w}_J)^\top \mathbf{R} (\mathbf{w}_K \otimes \mathbf{w}_J)}. \tag{2}$$

where the matrix $\mathbf{R} = \mathbf{R}_K \otimes \mathbf{R}_J$ is introduced to avoid numerical issues as in (1). $\mathbf{R}_K$ is of dimension $K \times K$ and $\mathbf{R}_J$ is of dimension $J \times J$. The choices of $\mathbf{R}$ and λ are illustrated in section 3.

MFDA algorithm. The optimization problem (2) is solved by considering the following identities:

$$\mathbf{w}^\top \mathbf{S_B} \mathbf{w} = \mathbf{w}_J^\top \left(\sum_{k=1}^{K} (\mathbf{w}_K)_k \mathbf{X}_{..k} \right)^\top \mathbf{H}_B \left(\sum_{k=1}^{K} (\mathbf{w}_K)_k \mathbf{X}_{..k} \right) \mathbf{w}_J \tag{3}$$

$$= \mathbf{w}_K^\top \left(\sum_{j=1}^{J} (\mathbf{w}_J)_j \mathbf{X}_{.j.} \right)^\top \mathbf{H}_B \left(\sum_{j=1}^{J} (\mathbf{w}_J)_j \mathbf{X}_{.j.} \right) \mathbf{w}_K \tag{4}$$

$$\mathbf{w}^\top \mathbf{S_T} \mathbf{w} = \mathbf{w}_J^\top \left(\sum_{k=1}^{K} (\mathbf{w}_K)_k \mathbf{X}_{..k} \right)^\top \left(\sum_{k=1}^{K} (\mathbf{w}_K)_k \mathbf{X}_{..k} \right) \mathbf{w}_J \tag{5}$$

$$= \mathbf{w}_K^\top \left(\sum_{j=1}^{J} (\mathbf{w}_J)_j \mathbf{X}_{.j.} \right)^\top \left(\sum_{j=1}^{J} (\mathbf{w}_J)_j \mathbf{X}_{.j.} \right) \mathbf{w}_K \tag{6}$$

and

$$(\mathbf{w}_K \otimes \mathbf{w}_J)^\top \mathbf{R} (\mathbf{w}_K \otimes \mathbf{w}_J) = (\mathbf{w}_J^\top \mathbf{R}_J \mathbf{w}_J)(\mathbf{w}_K^\top \mathbf{R}_K \mathbf{w}_K) \tag{7}$$

Solving the optimization problem (2) with respect to $\mathbf{w}_J$ while maintaining $\mathbf{w}_K$ fixed, is achieved with a joint use of equations (3) and (5). Similarly, solving the optimization problem (2) with respect to $\mathbf{w}_K$ while maintaining $\mathbf{w}_J$ fixed, is achieved with a joint use of equations (4) and (6). The MFDA algorithm that solves optimization problem (2) is described in Algorithm 1. This algorithm starts by assigning random

Algorithm 1 Computation of the first Multi-way FDA axis

Require: $\varepsilon > 0$, $\mathbf{w}_K^{(0)}$
  $q \leftarrow 0$
  repeat
    $\mathbf{X}_K = \sum_{k=1}^{K} \left(\mathbf{w}_K^{(q)} \right)_k \mathbf{X}_{..k}$, $\lambda_K = \left(\mathbf{w}_K^{(q)} \right)^\top \mathbf{R}_K \mathbf{w}_K^{(q)}$
    $\mathbf{w}_J^{(q+1)} \leftarrow \underset{\mathbf{w}_J, \|\mathbf{w}_J\|=1}{\operatorname{argmax}} \frac{\mathbf{w}_J^\top \mathbf{X}_K^\top \mathbf{H}_B \mathbf{X}_K \mathbf{w}_J}{\mathbf{w}_J^\top \mathbf{X}_K^\top \mathbf{X}_K \mathbf{w}_J + \lambda_K \mathbf{w}_J^\top \mathbf{R}_J \mathbf{w}_J} \leftarrow \mathrm{FDA}(\mathbf{y}, \mathbf{X}_K, \lambda_K)$
    $\mathbf{X}_J = \sum_{j=1}^{J} \left(\mathbf{w}_J^{(q+1)} \right)_j \mathbf{X}_{.j.}$, $\lambda_J = \left(\mathbf{w}_J^{(q+1)} \right)^\top \mathbf{R}_J \mathbf{w}_J^{(q+1)}$
    $\mathbf{w}_K^{(q+1)} \leftarrow \underset{\mathbf{w}_K, \|\mathbf{w}_K\|=1}{\operatorname{argmax}} \frac{\mathbf{w}_K^\top \mathbf{X}_J^\top \mathbf{H}_B \mathbf{X}_J \mathbf{w}_K}{\mathbf{w}_K^\top \mathbf{X}_J^\top \mathbf{X}_J \mathbf{w}_K + \lambda_J \mathbf{w}_K^\top \mathbf{R}_K \mathbf{w}_K} \leftarrow \mathrm{FDA}(\mathbf{y}, \mathbf{X}_J, \lambda_J)$
    $q \leftarrow q+1$
  until $\|\mathbf{w}_K^{(q-1)} - \mathbf{w}_K^{(q)}\| < \varepsilon$
  return $(\mathbf{w}_K^{(q)}, \mathbf{w}_J^{(q)})$

initial values for $\mathbf{w}_J$ or $\mathbf{w}_K$ and then iterates a sequence of FDA problems. More specifically, each update boils down to perform FDA on either $\mathbf{X}_J = \sum_{k=1}^{K} (\mathbf{w}_K)_k \mathbf{X}_{..k}$ or $\mathbf{X}_K = \sum_{J=1}^{J} (\mathbf{w}_J)_j \mathbf{X}_{.j.}$. From the expressions of $\mathbf{X}_J$ and $\mathbf{X}_K$, it becomes clear that

$(\mathbf{w}_J)_j$ reflects the influence of the the jth variable while $(\mathbf{w}_K)_k$ the influence of the kth modality. Notice that $\mathbf{X}_J$ (resp. $\mathbf{X}_K$) is a $I \times J$ (resp. $I \times K$) matrix as compared to the $I \times JK$ unfolded matrix $\mathbf{X}^u$.
Algorithm 1 yields $\mathbf{w}^1 = \mathbf{w}_K^1 \otimes \mathbf{w}_J^1$ corresponding to the first discriminant axis. Subsequent discriminant axes can be determined by imposing orthogonality constraints as detailed hereinafter.

Additional constraints. At the end of Algorithm 1, one discriminant vector $\mathbf{w}^1 = \mathbf{w}_K^1 \otimes \mathbf{w}_J^1$ is obtained. The following $C-1$ axes (where C is the number of classes): $\mathbf{w}_J^s$, $\mathbf{w}_K^s$, $s = 2, \ldots, C-1$, are obtained subject to orthogonality constraints expressed as follows:

$$\begin{aligned}(\mathbf{w}^s)^\top [\mathbf{w}^1 \ldots \mathbf{w}^{s-1}] = \mathbf{0} \iff & (\mathbf{w}_K^s \otimes \mathbf{w}_J^s)^\top (\mathbf{w}_K^c \otimes \mathbf{w}_J^c) = 0 \quad \forall c \in [1, \ldots, s-1] \\ & (\mathbf{w}_K^{s\top} \mathbf{w}_K^c)(\mathbf{w}_J^{s\top} \mathbf{w}_J^c) = 0 \quad \forall c \in [1, \ldots, s-1] \end{aligned} \tag{8}$$

From equation (8), orthogonality can be obtained by either imposing $\mathbf{w}_K^{s\top} \mathbf{w}_K^c = 0$ or $\mathbf{w}_J^{s\top} \mathbf{w}_J^c = 0$. The construction of the next discriminant axes is derived below for the constraint $\mathbf{w}_J^{s\top} \mathbf{w}_J^c = 0$.

Second discriminant axis. Considering $\mathbf{H} = \text{span}\left\{\mathbf{w}_J^1\right\}$ and $\mathbf{P}_{\mathbf{H}^\perp}$ the projection matrix over $\mathbf{H}^\perp$. The orthogonality condition is equivalent to say that there exists a non unique $\mathbf{v} \in \mathbb{R}^J$ such that $\mathbf{w}_J^2 = \dfrac{\mathbf{P}_{\mathbf{H}^\perp}\mathbf{v}}{\|\mathbf{P}_{\mathbf{H}^\perp}\mathbf{v}\|}$. The orthogonality constraint on $\mathbf{w}_J^2$ yields the optimization problem:

$$\max_{\mathbf{w}_K, \mathbf{v}} \frac{(\mathbf{w}_K \otimes (\mathbf{P}_{\mathbf{H}^\perp}\mathbf{v}))^\top \mathbf{S}_B (\mathbf{w}_K \otimes (\mathbf{P}_{\mathbf{H}^\perp}\mathbf{v}))}{(\mathbf{w}_K \otimes (\mathbf{P}_{\mathbf{H}^\perp}\mathbf{v}))^\top \mathbf{S}_T (\mathbf{w}_K \otimes (\mathbf{P}_{\mathbf{H}^\perp}\mathbf{v})) + \lambda (\mathbf{w}_K \otimes (\mathbf{P}_{\mathbf{H}^\perp}\mathbf{v}))^\top \mathbf{R} (\mathbf{w}_K \otimes (\mathbf{P}_{\mathbf{H}^\perp}\mathbf{v}))}, \tag{9}$$

subject to $\|\mathbf{w}_K\| = 1$ and $\|\mathbf{v}\| = 1$ which is also a MFDA problem due to the following identities:

$$\begin{aligned}\mathbf{w}^\top \mathbf{S}_B \mathbf{w} &= (\mathbf{w}_K \otimes (\mathbf{P}_{\mathbf{H}^\perp}\mathbf{v}))^\top \mathbf{X}^\top \mathbf{H}_B \mathbf{X} (\mathbf{w}_K \otimes (\mathbf{P}_{\mathbf{H}^\perp}\mathbf{v})) \\ &= \left(\sum_{k=1}^K \left(\mathbf{w}_K^2\right)_k (\mathbf{X}_{..k} \mathbf{P}_{\mathbf{H}^\perp}) \mathbf{v} \right)^\top \mathbf{H}_B \left(\sum_{k=1}^K \left(\mathbf{w}_K^2\right)_k (\mathbf{X}_{..k} \mathbf{P}_{\mathbf{H}^\perp}) \mathbf{v} \right) \\ &= \left(\sum_{j=1}^K \left(\mathbf{w}_J^2\right)_j (\mathbf{X}_{.j.} \mathbf{P}_{\mathbf{H}^\perp}) \mathbf{v} \right)^\top \mathbf{H}_B \left(\sum_{j=1}^J \left(\mathbf{w}_J^2\right)_j (\mathbf{X}_{.j.} \mathbf{P}_{\mathbf{H}^\perp}) \mathbf{v} \right)\end{aligned}$$

We emphasize that $\mathbf{P}_{\mathbf{H}^\perp}$ is of rank $J-1$ but does not pose any computational issues because $\mathbf{P}_{\mathbf{H}^\perp} = \mathbf{I} - \mathbf{P}_{\mathbf{H}}$ with $\mathbf{P}_{\mathbf{H}} = \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top = \mathbf{w}_J^1 (\mathbf{w}_J^1)^\top$. It comes that the projection is now advantageously replaced by a deflation:

$$\mathbf{X}_{..k} \mathbf{P}_{\mathbf{H}^\perp} = \mathbf{X}_{.k} - (\mathbf{X}_{..k} \mathbf{w}_J^1)(\mathbf{w}_J^1)^\top.$$

Subsequent discriminant axis. The s^{th} discriminant axis is obtained using the same deflation strategy considering the vectorial space $\mathbf{H} = \text{span}\left\{\mathbf{w}_J^1, \mathbf{w}_J^2, \ldots, \mathbf{w}_J^{s-1}\right\}$. Let $\mathbf{X}'$ be the three-way data obtained from the previous step. Since $\mathbf{X}'_{..k}$ has already been projected over $\text{span}\left\{\mathbf{w}_J^1, \mathbf{w}_J^2, \ldots, \mathbf{w}_J^{s-2}\right\}^{\perp}$. The vector $\mathbf{w}_J^{s-1}$ is thus obtained using Algorithm 1 on the deflated version of $\mathbf{X}'$ which is obtained from the following deflation:

$$\mathbf{X}'_{..k}(= \mathbf{X}_{..k}\mathbf{P}_{\mathbf{H}^{\perp}}) \leftarrow \mathbf{X}'_{..k} - (\mathbf{X}'_{..k}\mathbf{w}_J^{s-1})(\mathbf{w}_J^{s-1})^{\top}.$$

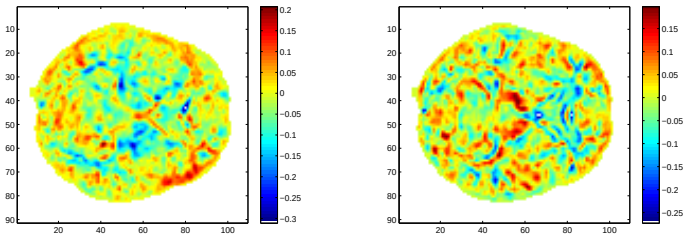
3 Application to traumatic brain injury

Traumatic brain injury is one of the leading causes of death and disability in the industrialized world, generally requiring prolonged rehabilitation [3].

In the scope of this paper MFDA is applied to a multi-modal brain MRI data set in order to predict, in the long term, the recovery of patients that suffered from traumatic brain injury. The I horizontal slices characterize the patients $i = 1, ..., I$, the J lateral slices are related to the voxels $j = 1, ..., J$ and the K frontal slices correspond to the different modalities $k = 1, ..., K$. From this the data can be structured into the tensor $\mathbf{X} = \left\{\mathbf{X}_{ijk}\right\}_{1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K}$ of order 3. Due to the high dimensionality of the dataset, a kernel version of FDA is used [5]. The optimal value for the regularization parameter λ is tailored through a leave-one-out cross-validation procedure. The $\mathbf{R}$ matrix is set to be the identity.

Data description. The multi-modal MRI diffusion images are acquired on individuals divided into 3 classes: 39 controls, 65 coma patients with a positive outcome and 39 coma patients with a negative outcome ($I = 143$). Four diffusion images ($K = 4$), namely fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (L1) and radial diffusivity (Lt), images were acquired from the entire brain of both patients and controls. Each volumetric image has 91x109x91 voxels which are then reshaped into a 1 x 902629 vector ($J = 902629$). The resulting tensor $\mathbf{X}$ considered as input for MFDA is of dimension $143 \times 902629 \times 4$, whereas the resulting unfolded tensor $\mathbf{X}^{\text{u}}$ is of dimension 143×3610516.

FDA applied to the entire brain. A linear kernel version of FDA [5] applied to $\mathbf{X}^u$ results in 8 weight matrices (4 for each eigenvector). A leave-one-out cross-validation yields the optimal regularization parameter to be $\lambda = 400$ with an accuracy of 76%. Moreover, the resulting FA weights matrix obtained by considering the segment of the eigenvectors corresponding to FA are shown in Figure 2. These images are difficult to interpret since there is no focalized region used for the discrimination. We mention that other modalities (i.e. other segments) could be visualized but do not give additional discriminative information (results not shown).



(a) FDA analysis FA, 1st eigenvector. (b) FDA analysis FA, 2nd eigenvector.

Fig. 2 Entire brain. FA segment of the FDA weights vectors ($\mathbf{w}^1_{FA}$, $\mathbf{w}^2_{FA}$) for $\lambda = 400$.

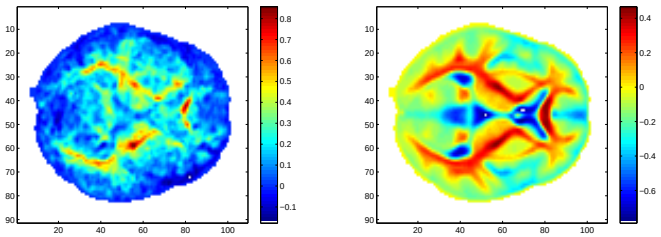
MFDA applied to the entire brain. MFDA applied to $\mathbf{X}$ results in 2 weight matrices associated with $\mathbf{w}^1_J$ and $\mathbf{w}^2_J$ which integrate all the modalities. This yields a single volumetric image that integrate the 4 modalities instead of one for each modality in FDA. After performing a leave-one-out cross-validation, the optimal regularization parameter for MFDA is found to be $\lambda = 10^4$ with an accuracy of 71%. Table 1 shows the contribution of each modality for the construction of the single volumetric image. FA has the highest weight in the discrimination for $\mathbf{w}^1_K$. For $\mathbf{w}^2_K$, all modalities but FA have been taken into account in the same proportion.

Table 1 Entire brain. MFDA weights vectors ($\mathbf{w}^1_K$, $\mathbf{w}^2_K$)

Modality	$\mathbf{w}^1_K$	$\mathbf{w}^2_K$
FA	**0.9887**	-0.0066
MD	0.0036	0.5703
L1	0.0046	0.6094
Lt	0.0031	0.5508

Figure 3 shows an example of MFDA weights $\mathbf{w}^1_J$ and $\mathbf{w}^1_J$ obtained on the entire brain (same plane as for FDA). Contrary to FDA, MFDA clearly locates the discriminative voxels in the white matter (in red). Since specific and smooth regions are selected, MFDA model is easier to interpret. In addition, $\mathbf{w}^2_J$ is reported in Table 1 and shows that all the modalities participate in the same proportion to the construction of the MFDA model. These results are consistent with the ones obtained by Sidaros *et al*. [6] and Galanaud *et al*. [2] regarding the importance of FA when assessing long-term recovery.

MFDA exhibits that the discriminating voxels are located, within the main white matter bundles, more specifically in the posterior limb of the internal capsule. Indeed, traumatic brain injury is characterized by the presence of diffuse axonal injury mainly located within deep and axial white matter bundle as found by Galanaud *et al*. [2]. For this reason, a second analysis based only on the white matter region is applied giving a $143 \times 20764 \times 4$ tensor to analyze.



(a) $\mathbf{w}_J^1$. (b) $\mathbf{w}_J^2$.

Fig. 3 Entire brain. MFDA weights vectors ($\mathbf{w}_J^1$, $\mathbf{w}_J^2$) for $\lambda = 10,000$.

FDA and MFDA applied to the white matter. In Figure 4, training and testing accuracies for FDA are reported for different values of λ. The optimal regularization parameter for FDA is equal to $\lambda = 400$ with an accuracy of 75%. The associated confusion matrix is shown in Table 2. We note that the most frequent error is done between the positive and negative outcome, and that the distinction between patients and controls is very accurate. In Figure 5, training and testing accuracies for MFDA are reported. The optimal regularization parameter is $\lambda = 100$ with an accuracy of 77%. The associated confusion matrix is shown in Table 3.

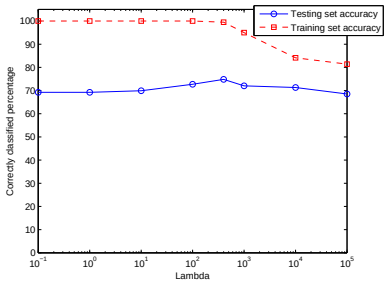


Fig. 4 FDA Leave-one-out cross validation.

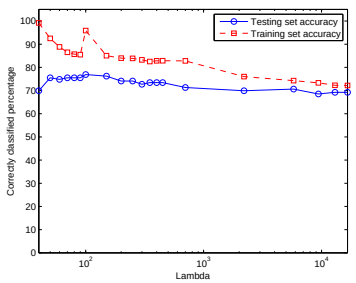


Fig. 5 MFDA Leave-one-out cross validation.

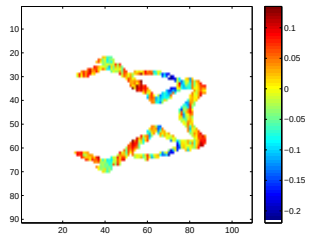
Table 2 FDA confusion matrix with $\lambda = 400$.

FDA	**Predicted**		
Observed	Controls	Positive outcomes	Negative outcomes
Controls	39	0	0
Positive outcomes	6	49	10
Negative outcomes	0	20	19

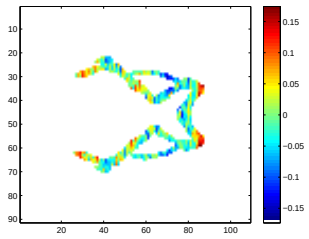
Table 3 MFDA confusion matrix with $\lambda = 100$.

MFDA	**Predicted**		
Observed	Controls	Positive outcomes	Negative outcomes
Controls	37	2	0
Positive outcomes	0	49	16
Negative outcomes	0	15	24

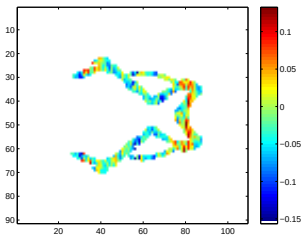
The resulting weights obtained when analyzing the white matter are presented in Figure 6, together with the corresponding w_K values in Table 4. These results are consistent with the ones obtained using the entire brain, where modality FA serves as the most discriminant modality.



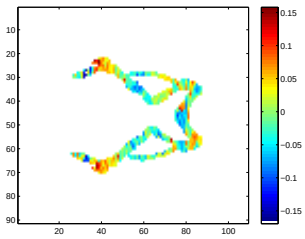
(a) FDA analysis FA, 1st eigenvector.



(b) MFDA analysis, $\mathbf{w}_J^1$.



(c) FDA analysis FA, 2nd eigenvector.



(d) MFDA analysis, $\mathbf{w}_J^2$.

Fig. 6 White matter. MFDA weights vectors ($\mathbf{w}_J^1$, $\mathbf{w}_J^2$) for $\lambda = 100$. FA segment of the FDA weights vectors ($\mathbf{w}_{FA}^1$, $\mathbf{w}_{FA}^2$) for $\lambda = 400$.

Table 4 White matter. MFDA weights vectors ($\mathbf{w}_K^1$, $\mathbf{w}_K^2$)

Modality	$\mathbf{w}_K^1$	$\mathbf{w}_K^2$
FA	**0.8017**	-0.0681
MD	-0.2224	-0.5327
L1	0.2319	-0.8072
Lt	-0.5040	-0.2448

4 Discussion

In this paper, we propose a multiway formulation of FDA that considers the intrinsic tensor structure of the data. MFDA was applied to multi-modal MRI diffusion images to predict the long term recovery of patients with traumatic brain injury, for which good accuracy rates were obtained, from 71% for MFDA to 76% for FDA,

when using the entire brain. This loss in accuracy for MFDA is compensated by an improvement in the interpretability of the obtained classifier. This improvement is due to the introduced a priori structure that has been taken into account during the modelisation process. When analyzing the white matter we obtain a 75% accuracy for FDA and 77% for MFDA. MFDA separates the influence of spatial positions and the influence of the different modalities.
Another observation is that the FDA weights give higher importance to the borders of the brain, when the majority of the discriminant information should be found in the white matter since there is evidence that damage in this region is a distinctive feature of traumatic brain injury [2] as shown in the MFDA results. The MFDA weight matrices seem to supply more information on the location of the discrimination regions, as shown in Figure 3. Moreover, FDA results in 8 weight matrices ($J \times K$ classifier), complicating the interpretability, as opposed to only 2 weight matrices ($J + K$ classifier) obtained with MFDA which integrate all the modalities.
Future perspectives include an improvement of the accuracy of the classification of the positive and negative outcomes. In order to further improve the interpretability of the classifier a sparse MFDA algorithm is under development for reducing the number of active variables in the MFDA model.

Acknowledgements This study was funded by a grant from the French Ministry of Health (Projet Hospitalier de Recherche Clinique registration #P051061 [2005]) and from departmental funds from the Assistance Publique-Hôpitaux de Paris. The research leading to these results has also received funding from the program "Investissements d'avenir" ANR-10-IAIHU-06 and LG acknowledges support from CONACYT.

References

1. Rasmus Bro. *Multi-way Analysis in the food industry: models, algorithms and applications*. PhD thesis, Royal Veterinary and Agricultural University, 1998.
2. D Galanaud, V Perlbarg, R Gupta, RD Stevens, P Sanchez, E Tollar, NM Champfleur, J Dinkel, S Faivre, G Soto-Ares, B Veber, V Cottenceau, F Masson, T Tourdias, E Andr, G Audibert, E Schmitt, D Ibarrola, F Dailler, A Vanhaudenhuyse, L Tshibanda, JF Payen, JF Le Bas, A Krainik, N Bruder, N Girard, S Laureys, H Benali, and L Puybasset. Assessment of white matter injury and outcome in severe brain trauma: a prospective multicenter cohort. *Anesthesiology*, 117(6):1300–10, December 2012.
3. A. Grubb, P. Walsh, N. Lambe, T. Murrells, and S. Robinson. Survey of british clinicians' views on management of patients in persistent vegetative state. *Lancet.*, (348):35–40, 1996.
4. R. A. Harshman. Foundations of the parafac procedure: models and conditions for an explanatory multi-mode factor analysis. *UCLA Working Papers in Phonetics*, (16):1–84, 1970.
5. Sebastian Mika, Gunnar Ratsch, Jason Weston, Bernhard Scholkopf, and Klaus-Robert Muller. Fisher discriminant analysis with kernels. In *IEEE Conference on Neural Networks for Signal Processing IX*, pages 41–48, August 1999.
6. A. Sidaros, AW. Engberg, K. Sidaros, MG. Liptrot, M. Herning, P. Petersen, OB. Paulson, TL. Jernigan, and E. Rostrup. Diffusion tensor imaging during recovery from severe traumatic brain injury and relation to clinical outcome: A longitudinal study. *Brain*, (131):559–72, 2008.