Improved Parkinson's Disease Classification from Diffusion MRI Data by Fisher Vector Descriptors

Luis Salamanca[1], Nikos Vlassis[2], Nico Diederich[1,3], Florian Bernard[1,3,4], and Alexander Skupin[1]

[1] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[2] Adobe Research, San Jose, CA, USA
[3] Centre Hospitalier de Luxembourg, Luxembourg City, Luxembourg
[4] Trier University of Applied Sciences, Trier, Germany

Abstract. Due to the complex clinical picture of Parkinson's disease (PD), the reliable diagnosis of patients is still challenging. A promising approach is the structural characterization of brain areas affected in PD by diffusion magnetic resonance imaging (dMRI). Standard classification methods depend on an accurate non-linear alignment of all images to a common reference template, and are challenged by the resulting huge dimensionality of the extracted feature space. Here, we propose a novel diagnosis pipeline based on the Fisher vector algorithm. This technique allows for a precise encoding into a high-level descriptor of standard diffusion measures like the fractional anisotropy and the mean diffusivity, extracted from the regions of interest (ROIs) typically involved in PD. The obtained low dimensional, fixed-length descriptors are independent of the image alignment and boost the linear separability of the problem in the description space, leading to more efficient and accurate diagnosis. In a test cohort of 50 PD patients and 50 controls, the implemented methodology outperforms previous methods when using a logistic linear regressor for classification of each ROI independently, which are subsequently combined into a single classification decision.

Keywords: neurodegenerative diseases, diagnosis, diffusion magnetic resonance imaging, machine learning, feature extraction.

1 Introduction

Magnetic resonance imaging (MRI) has become a well-known useful and established technique for the diagnosis and progression monitoring of neurodegenerative diseases. Recent developments on MRI allow for capturing different properties of brain tissues, like the structural integrity by diffusion MRI (dMRI), or the activation and interaction between different brain regions when performing a task or in resting-state by functional MRI. To extract diagnostic parameters, traditional approaches for the statistical analysis of these images rely on

a group-level, voxel-wise comparison of the extracted measures by using univariate statistical tools such as the t-test, χ^2-test, etc. Common techniques like voxel-based morphometry [1] can provide good insights in the pathological involvement of different brain regions, but lack the ability of fully integrate the extracted features.

Currently, more advanced machine learning methodologies are being used to fully integrate the data and provide results at the level of the individual [11]. The typically small sample size and the high-dimensionality of the data pose major problems when using these methods. Although a step of feature selection may alleviate the problem, there is still a huge risk of overfitting that may obscure the results' validity [11]. Additionally, the considered features are, most presumably, non-linearly separable, what restrains the ability of linear classifiers, and even more advanced kernel-based methods, for obtaining optimal decision thresholds.

In this paper we propose a novel pipeline for the diagnosis of Parkinson's disease (PD). Our methodology utilizes the advantages of the Fisher vector (FV) algorithm [14] to regionally encode the data in high-level descriptors. This way, and in contrast to standard approaches, we get rid of the necessity of a fine alignment between different subjects' brains, since a voxel-wise comparison is not required. Therefore, the step of non-linear registration of all images to a reference template can be skipped, excluding artefacts or undesired corruption of the subject data due to under- or over-alignments. Further, thanks to representing the distribution of the data captured from all voxels of each region of interest (ROI) as a fixed-length high-level descriptor, we lead to an immense reduction of the dimensionality of the problem. Finally, since the FV algorithm is based on the Fisher Kernel (FK) principle [8], we can understand the process of computing the FV descriptor as an embedding of the extracted features into a higher-dimensional space, more suitable to linear classification. This strategy directly supports the last stage of classification and crucially improves the diagnosis results.

Although the diagnosis of PD remains as a challenging problem, especially in the early stages when there is not clear evidence of brain atrophy, there have been already promising machine learning approaches in the literature [10,13]. However, these previous studies lack a sufficient number of subjects. In the current paper, we apply the proposed methodology to the dMRI data of a test cohort of 50 PD patients and 50 healthy controls, chosen from the PPMI consortium database [12]. In our test scenario we show that our approach reaches a global accuracy of 77%, outperforming other machine learning strategies on the same data set by at least 5%.

2 Methods

In this section we provide a brief overview of the Fisher vector algorithm, and how the high-level descriptor is obtained from the raw neuroimaging data. We also describe in detail the developed pipeline, from the preprocessing of the data to the last step of classification.

2.1 Fisher Vector Algorithm

Given the measures extracted from all the subjects for a specific brain ROI, we define the sample $\mathbf{X} = \{\mathbf{x}_n : n = 1, \ldots, N\}$ of D-dimensional observations $\mathbf{x}_n \in \mathcal{X}$, with N the total number of voxels and D the number of extracted features. Let's consider a generative process defined by the probability density function (p.d.f.) u_θ, which models the generation of the elements in $\mathcal{X}$. The p.d.f. u_θ is defined by the set of parameters $\theta = \{\theta_m : m = 1, \ldots, M\}$, where $\theta \in \mathbb{R}^M$. To measure the similarity between two samples $\mathbf{X}$ and $\mathbf{Y}$, the authors in [8] use the Fisher Kernel, which is defined as follows:

$$K_{\mathrm{FK}}(\mathbf{X}, \mathbf{Y}) = G_\theta^{\mathbf{X}'} F_\theta^{-1} G_\theta^{\mathbf{Y}} = G_\theta^{\mathbf{X}'} L_\theta' L_\theta G_\theta^{\mathbf{Y}} = \mathcal{G}_\theta^{\mathbf{X}'} \mathcal{G}_\theta^{\mathbf{Y}}, \tag{1}$$

where F_θ is the Fisher information matrix (FIM), factorized into terms L_θ through Cholesky decomposition, and $G_\theta^{\mathbf{X}}$ is the score function. This function, defined as $G_\theta^{\mathbf{X}} = \nabla_\theta \log u_\theta(\mathbf{X})$, measures how the parameters θ of the generative model u_θ should be modified to better fit $\mathbf{X}$. The Cholesky decomposition of the FIM allows to re-write the FK as the dot-product of the gradient statistics $\mathcal{G}_\theta$, as we observe in Eq. (1).

In [14], the authors choose a Gaussian mixture model (GMM) as the generative model u_θ, due to it can accurately capture any continuous distribution [15]. Thus, the p.d.f. u_θ is defined by K Gaussian components as follows:

$$u_\theta(\mathbf{x}) = \sum_{k=1}^{K} w_k u_k(\mathbf{x}) \quad \text{with} \quad u_k \sim \mathcal{N}(\mu_k, \sigma_k \mathbf{I}), \tag{2}$$

where $\mu_k \in \mathbb{R}^D$ is the mean and $\sigma_k \in \mathbb{R}^D$ is the diagonal of the covariance matrix for the k-th Gaussian component. The restrictions $w_k \geq 0 \,\forall\, k$ and $\sum_{k=1}^{K} w_k = 1$ over the mixture weights w_k can be implicitly included in the auxiliary parameter $\alpha_k \in \mathbb{R}$ [14]. Therefore, the set of parameters results in $\theta = \{\alpha_k, \mu_k, \sigma_k : k = 1, \ldots, K\}$. Given $\mathbf{X}$, they can be computed by using the expectation-maximization (EM) algorithm [5].

Now, let us define $\mathbf{X}_\ell = \{\mathbf{x}_{n_\ell} : n_\ell = 1, \ldots, N_\ell\}$ as the subset of observations $\mathbf{X}_\ell \subset \mathbf{X}$ for subject ℓ. Given the set of parameters θ computed using the whole sample $\mathbf{X}$, and assuming independence between observations in $\mathbf{X}_\ell$, the gradient statistics in Eq. (1) can be obtained as [14]:

$$\mathcal{G}_\theta^{\mathbf{X}_\ell} = \sum_{n_\ell=1}^{N_\ell} L_\theta \nabla_\theta \log u_\theta(\mathbf{x}_{n_\ell}) \tag{3}$$

Thanks to considering a GMM as the generative model, and by assuming the FIM to be diagonal, analytic expressions can be derived for each parameters θ_m, as detailed in [14]. Through these expressions, we readily calculate the gradients $\mathcal{G}_{\mu_k}^{\mathbf{X}_\ell}$, $\mathcal{G}_{\sigma_k}^{\mathbf{X}_\ell}$ and $\mathcal{G}_{\alpha_k}^{\mathbf{X}_\ell}$ for each Gaussian component k, which are stacked together to form the Fisher vector (FV) for each subject ℓ, hereafter denoted as $\mathbf{F}_\ell$. The gradient statistics $\mathcal{G}_{\theta_m}^{\mathbf{X}_\ell}$ measure the deviation of the subject data $\mathbf{X}_\ell$ from the

distribution of the generative model, fitted for all data $\mathbf{X}$. This way, the FV descriptor captures the slight variations in each subject's data distribution with respect to the distribution for all samples. Now, since $\mathcal{G}_{\alpha_k}^{\mathbf{X}_\ell}$ is scalar, and $\mathcal{G}_{\mu_k}^{\mathbf{X}_\ell}$ and $\mathcal{G}_{\sigma_k}^{\mathbf{X}_\ell}$ are D-dimensional vectors, the dimension of $\mathbf{F}_\ell$ is $(2D+1)K$, therefore depending only on the parameters of the generative model, and not on the size of the ROI described. Finally, given that the Fisher kernel is equivalent to the dot product of the gradient statistics, Eq. (1), using the FV descriptor as input to a linear classifier is equivalent to directly feeding the original features to a non-linear classifier using the Fisher kernel. Therefore, the description space of the Fisher vector is a more amenable representation of the extracted measures.

2.2   Pipeline for dMRI Analysis

To conveniently illustrate the proposed methodology, we depict in Fig. 1 the flow diagram of the pipeline. The input data required is:

- The diffusion weighted image (DWI) of each subject. Through *eddy*, included in the FMRIB Software Library (FSL) [9], we first correct the eddy current and head motion. Then, the brain is extracted from the T2-weighted b0 volume, contained in the DWI tensor, using *bet*, also included in FSL.
- A reference template to align the subjects to.
- Masks for the ROIs considered, in the same space as the reference template.

Given this data, we first use FLIRT (FMRIB's Linear Image Registration Tool) to linearly align the DWI volume to the reference template, by using an affine transformation with 12 degrees of freedom and the T2-weighted b0 volume as the moving image. Then, the reference template is non-linearly aligned to each subject's space, using the symmetric image normalisation method (SyN) contained in the ANTS package [2], considering cross-correlation as the image metric. The transformations obtained are used to warp the masks of all the selected ROIs to each subject space. Thanks to this strategy of registration, already used before in the literature, the patient images are not deformed, and therefore the risk of corrupting the data is minimised, in contrast to approaches where patient images are non-linearly aligned to the reference template.

Using the warped masks of the ROIs for all the subjects, we subdivide each ROI in different subROIs in the x, y and z directions according to some given minimum and maximum voxel sizes for the resulting subdivisions. It is important to note that, although for every ROI all the subjects will have the same number of subROIs and the same scheme of division in every direction, each subROI will have a particular voxel size because the subjects' brains are not non-linearly aligned to a common template. Then, by using the *DTIfit* tool included in FSL, we apply diffusion tensor imaging (DTI) method on the DWI data to extract the diffusion tensor and related diffusion measures. With the subROIs indices obtained from the masks, we extract a matrix $\mathbf{X}_\ell^r$ of diffusion features for every subject ℓ and subROI r.

Finally, for each r, we fit the GMM to the data $\mathbf{X}^r = [\mathbf{X}_1^r; \ldots; \mathbf{X}_L^r]$ using the EM algorithm [5]. The parameters obtained, together with each subROI data
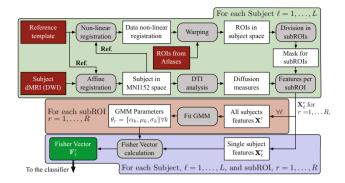
Fig. 1. Flow diagram for the proposed pipeline.

$\mathbf{X}_\ell^r$, allow to finally compute the fixed-length Fisher vector descriptor $\mathbf{F}_\ell^r$ for every subROI and subject, as described in Section 2.1.

2.3   Classification

To show the enhanced separability of the problem in the description space, we consider a logistic linear regressor as classifier, with an $L2$ regularizer. Since this classifier can easily handle thousands of features, we first perform the training and classification for all the ROIs. Additionally, we also provide the performance for each ROI separately. These individual predictions are combined to provide a single diagnosis for each subject, by using a polling procedure, i.e. the decision chosen is the one taken by the majority of the ROIs. Additionally, and for the sake of comparison, we also consider a support vector machine (SVM) classifier, with a radial-basis function (RBF) kernel. This will allow to compare the performance of using the FV descriptors as input to a linear classifier, against the performance of more powerful kernel methods using directly the diffusion measures. But the FV descriptors are not fed to the SVM because we just want to show their potential when using linear classifiers. The software packages *liblinear* and *libSVM* are used to run the experiments [3,6].

3   Experiments

The data used in the preparation of this article was obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. From the whole database, we initially chose only the baseline data of those subjects with DTI-gated images, resulting a total of 185 subjects, 131 PD patients and 54 controls. We visually inspected all the images, and excluded 10 of them due to noticeable imaging artefacts. Then, we select 50 PD patients and 50 controls by using the following criteria: we first select the 50 PD cases that minimize the deviation from the mean disease duration, from the date of diagnosis. Then, the control group is

chosen such that we match the mean age of the PD group. The ages (mean ± standard deviation) of the PD and controls groups are 60.21 ± 10.23 and 61.00 ± 10.74, respectively, and the disease duration 4.56 ± 1.73 months.

The DWI images from the PPMI database have a resolution of $1.98 \times 1.98 \times 2$mm. Thus, from the MNI-ICBM152 symmetric template [7], we obtain a T2 reference image with resolution $2 \times 2 \times 2$mm. For the analyses we initially selected 14 ROIs usually related with the disease pathology: *superior and middle cerebellar penduncle*, *internal and external capsule*, *gyrus rectus*, *caudate body*, *dentate nucleus*, *pons*, *sub-thalamic nucleus*, *red nucleus*, *globus pallidus externa and interna*, *putamen* and *substantia nigra*. From them, only the most discriminative ones were selected, as we observe in Table 1. The atlases considered to extract these ROIs are the MNI, JHU and Talairach, all included in FSL. From the set of orthogonal diffusion measures obtained through DTI, we just consider the FA and MD, as they are the most extended ones in the neuroimaging literature [4]. Any linear correlation of these measures with the age and/or gender is regressed out before computing the FV descriptors.

The algorithm implemented for dividing the ROIs tries to bound the size of each subROI between some minimum and maximum number of voxels, whenever possible. After testing different configurations, a size from 48 to 64 voxels has been chosen. This range results in a good trade-off between having enough samples to conveniently approximate the data distribution of each subROI, and enough subROIs for encoding in detail the regional information of each ROI. Then, once obtained $\mathbf{X}_\ell^r$ for all subjects and subROIs, instead of jointly considering the FA and MD, we fit a GMM for each measure in order to provide independent scores, showing the different sensibility of each ROI to the brain alterations caused by the disease. The number of components K for the GMM is set to 6, which allows to precisely fit the data distribution without excessively increasing the model complexity.

To show the performance of the proposed methodology, we compare it against an approach that directly uses the diffusion measures as inputs to a linear or an SVM-RBF classifier. For each of these schemes, we adjust separately the regularization parameter. We perform 100 experiments of 10-fold cross-validation to obtain an average accuracy score. A mean prediction is also obtained for each ROI by averaging the results of each experiment, which are then use to obtain a combined score for each subject. As we can observe in Table 1, when comparing separately each of the ROIs, the proposed methodology allows to moderately enhance the accuracy of the classifier, and remarkably in the case of the *putamen*, brain region critically involved in the disease. However, sometimes the methodology is not able to correctly capture the distribution of the data through the FV descriptor, reducing its informativeness, as with the *gyrus rectus*, which performs better in the traditional schemes. But in general, the individual scores for all the ROIs are better than those provided by the other schemes, slight improvements that are crucial when computing the combined score. For the case of the accuracy score for all the ROIs, the proposed methodology also outperform standard approaches, but with a decrease on the performance.

Table 1. Rate of correct classification for the different experiments performed.

	FA and MD			FA			MD		
Input Data	FV Desc.	Diff. Feat.	Diff. Feat.	FV Desc.	Diff. Feat.	Diff. Feat.	FV Desc.	Diff. Feat.	Diff. Feat.
Classifier	Log. Reg.	Log. Reg.	SVM-RBF	Log. Reg.	Log. Reg.	SVM-RBF	Log. Reg.	Log. Reg.	SVM-RBF
Internal Caps.	**63.98**%	56.26%	61.08%	**64.06**%	62.19%	60.98%	**61.16**%	55.34%	60.72%
External Caps.	63.41%	60.13%	61.15%	**67.30**%	51.24%	61.18%	63.15%	**65.42**%	60.78%
Middle Cer. Pend.	57.78%	57.11%	**61.20**%	52.74%	55.67%	**61.50**%	54.87%	48.04%	**61.13**%
Gyrus rectus	50.63%	60.49%	**61.62**%	52.30%	**68.61**%	61.51%	50.36%	59.01%	**61.52**%
Caudate	**61.96**%	56.66%	61.46%	**64.17**%	52.61%	60.74%	54.04%	50.13%	**61.54**%
Dentate	**65.36**%	56.26%	61.66%	58.65%	58.13%	**60.90**%	59.02%	50.42%	**61.74**%
Putamen	**77.18**%	63.45%	60.76%	**74.91**%	51.32%	60.40%	60.01%	**63.53**%	60.95%
Combined score	**77**%	66%	60%	**74**%	64%	69%	**65**%	61%	61%
All ROIs	**70.34**%	59.58%	60.66%	**72.90**%	61.58%	61.06%	**62.24**%	58.00%	59.84%

4 Discussion and Future Work

In the present paper we have proposed a novel pipeline for the diagnosis of PD through dMRI data. By using the FV algorithm, we accurately encode the regional information of the diffusion measures extracted, obtaining a more amenable descriptor for the classifier, as well as alleviating the problem of dimensionality. We have reported an accuracy of 77%, outperforming equivalent approaches when either considering all the ROIs as inputs to the classifier or each of them separately. Besides, and in contrast to voxel-based morphometry approaches, where the accuracy of the non-linear transformation from the subjects' images to a common template plays a major role, the combination of the registration strategy from the template image to the patient image and the use of high-level FV descriptors allows to reduce the influence of the non-linear registration accuracy on the outcome. Furthermore, despite we have focused on the problem of PD diagnosis through dMRI, we can equally apply the current pipeline for the diagnosis of other diseases, as well as make use of other imaging modalities. In these other cases, due to the general nature of our approach, it can be expected that the proposed methodology will outperform more naive approaches, specially considering the higher sensibility other modalities may have to the alterations caused by the disease.

Whilst there is room for improvements in our methodology, for example in the way of combining the individual ROI classification results to obtain a single score, or in tuning parameters, we were able to demonstrate the validity of using high-level description techniques for the classification of neuroimaging data. Future work will tackle these potential weaknesses by, for example: considering more involved approaches for the combination of the class probabilities in order to provide a better prediction; exploring the advantages of using a single GMM for all the extracted features; and defining optimal methods to select the number of GMM components. Together with all these potential improvements, we will lead then to a more robust and generalisable diagnostic methodology.

References

1. Ashburner, J., Friston, K.J.: Voxel-based morphometry–the methods. NeuroImage 11(6 Pt 1), 805–821 (2000)
2. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12(1), 26–41 (2008)
3. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011)
4. Cochrane, C.J., Ebmeier, K.P.: Diffusion tensor imaging in parkinsonian syndromes: a systematic review and meta-analysis. Neurology 80(9), 857–864 (2013)
5. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Society, Series B 39(1), 1–38 (1977)
6. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: A library for large linear classification. The Journal of Machine Learning Research 9, 1871–1874 (2008)
7. Fonov, V., Evans, A., McKinstry, R., Almli, C., Collins, D.: Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage 47(S1), S102 (2009)
8. Jaakkola, T.S., Haussler, D.: Exploiting generative models in discriminative classifiers. In: Proc. of the 1998 Conf. on Advances in Neural Information Processing Systems II, pp. 487–493. MIT Press, Cambridge (1999)
9. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: FSL. Neuroimage 62(2), 782–790 (2012)
10. Long, D., Wang, J., Xuan, M., Gu, Q., Xu, X., Kong, D., Zhang, M.: Automatic classification of early Parkinson's disease with multi-modal MR imaging. PLOS ONE 7(11), e47714 (2012)
11. Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G., Mechelli, A.: Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neuroscience & Biobehavioral Reviews 36(4), 1140–1152 (2012)
12. Parkinson Progression Marker Initiative: The Parkinson Progression Marker Initiative (PPMI). Progress in Neurobiology 95(4), 629–635 (2011)
13. Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., Arabia, G., Morelli, M., Gilardi, M.C., Quattrone, A.: Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and progressive supranuclear palsy. Journal of Neuroscience Methods 222, 230–237 (2014)
14. Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J.: Image classification with the Fisher vector: Theory and practice. International Journal of Computer Vision 105(3), 222–245 (2013)
15. Titterington, D., Smith, A., Makov, U.: Statistical Analysis of Finite Mixture Distributions. John Wiley (1985)