Deep learning models to study the early stages of Parkinson's disease

Verónica Muñoz Ramírez, Virgilio Kmetzsch, Florence Forbes, Michel Dojat

DEEP LEARNING MODELS TO STUDY THE EARLY STAGES OF PARKINSON'S DISEASE

Verónica Muñoz Ramírez[⋆†] *Virgilio Kmetzsch*[†⋆] *Florence Forbes*[†] *Michel Dojat*[⋆]

[⋆] Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, GIN, 38000 Grenoble, France
[†] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

ABSTRACT

Current physio-pathological data suggest that Parkinson's Disease (PD) symptoms are related to important alterations in subcortical brain structures. However, structural changes in these small regions remain difficult to detect for neuroradiologists, in particular, at the early stages of the disease (*de novo* PD patients). The absence of a reliable ground truth at the voxel level prevents the application of traditional supervised deep learning techniques. In this work, we consider instead an anomaly detection approach and show that auto-encoders (AE) could provide an efficient anomaly scoring to discriminate *de novo* PD patients using quantitative Magnetic Resonance Imaging (MRI) data.

Index Terms— Brain, Anomaly detection, Autoencoder, Diffusion Imaging, MRI

1. INTRODUCTION

Today, there is a pressing need for objective and reliable biomarkers that allow the detection of Parkinson's Disease (PD) from its early stages. MRI has played a vital role in the characterization of multiple neurological diseases like Multiple Sclerosis and brain cancer [1]. However, structural MR images appear to be insufficient to detect the subtle changes caused by PD, especially in the sub-cortical structures of the brain [2]. This motivates the study of PD through quantitative MRI techniques such as DTI (Diffusion Tensor Imaging) that measures the displacement of water molecules in the brain. Indeed, DTI has been useful in the study of Alzheimer's disease [1], making it very attractive for the study of PD.

In this work, we propose to implement an anomaly detection framework to uncover alterations in the diffusion MR images of newly diagnosed (i.e. de novo) patients.

While there exists numerous techniques for anomaly detection [3], deep learning models have achieved remarkable results in recent brain lesion classification challenges such as BRATS, BrainLes and ISLES at MICCAI conferences (see for instance [4]). Nonetheless, these deep learning approaches for anomaly detection are often supervised, meaning that the networks are trained with representatives of both healthy and diseased voxels so as to learn their characteristics. Seeing that the only information available for PD MR scans is an indication of the global presence or absence of the disease, we designed a semi-supervised framework employing three different types of fully convolutional auto-encoders.

More specifically, we trained the auto-encoders to reconstruct healthy diffusion MR scans with a healthy training dataset. Since the network is learned only from healthy subjects, there is no guarantee that it provides good reconstructions outside this population. We therefore compared the scan reconstruction errors of a healthy test data set with those of a pathological data set to identify unusual patterns in the sub-cortical structures of *de novo* PD patients.

We follow the work of [5], which demonstrated the advantages of employing full MR slices as input for the network architectures instead of down sampling and patch division that result in important loss of spatial information.

2. INPUT DATA

We pooled our data from the PPMI (Parkinson Progression Markers Initiative) database. The PPMI is a longitudinal study that follows *de novo* PD patients of 35 centers for five years. The database is openly available for researchers and contains, among other clinical test results, structural and diffusion MR images. To eliminate any additional sources of bias, we only selected scans acquired with the same acquisition parameters, notably magnetic field and scanner manufacturer. As a result, we pooled 129 *de novo* patients (age: 62 ± 9; sex: 48 F, 80 M) and 57 healthy controls (age: 61 ± 10; sex: 23 F, 34 M).

From this data, two features were extracted: mean diffusivity (MD), accounting for the overall water displacement, and fractional anisotropy (FA), an indication of diffusion orientation. Values of FA and MD were normalized into the range $[0, 1]$. Each volume was composed of $116 \times 116 \times 72$ voxels. Forty coronal slices (116×72) were extracted from the center of the brain encompassing the sub-cortical structures under study.

To avoid data leakage, the control dataset was divided into

VMR is supported by a grant from NeuroCoG IDEX UGA in the framework of the "Investissements dávenir" program (ANR-15-IDEX-02).

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database. For up-to-date information on the study, visit www.ppmi-info.org.

a training dataset (42 MR volumes) and a testing dataset (15 MR volumes). Special care was taken to maintain similar age and sex profiles in both datasets. In order to choose the best network architecture and tune the corresponding hyper-parameters, all models were trained and assessed with 7-fold cross-validation, with 36 MR volumes used for training and 6 MR volumes for validation in each fold. Once the final models were defined, they were retrained with the full control training set (1680 slices) and evaluated with the PD dataset (5160 slices) and the healthy control test set (600 slices).

3. ARCHITECTURE DESIGN

Three autoencoder-based models were developed and evaluated: a spatial autoencoder (sAE), a spatial variational autoencoder (sVAE) and a dense variational autoencoder (dVAE).

As seen in Figure 1-A, all autoencoders come with two parts, an encoder f_ϕ and a decoder g_θ. The encoder maps the input images x to a lower dimensional latent representation z, then the decoder maps the latent vector z to the reconstructed output images $\hat{x} \in \mathbb{R}^{H\times W\times C}$:

$$z = f_\phi(x) \qquad\qquad \hat{x} = g_\theta(z) \tag{1}$$

We extended the work of [5] by using multiple quantitative MRI measures simultaneously as input. Thus, every input can be expressed as $x \in \mathbb{R}^{H\times W\times C}$, where H is the height, W the width and C the number of channels (FA and MD).

Depending on the network architecture, the latent code may be a simple vector ($z \in \mathbb{R}^d$) or a third-order tensor ($z \in \mathbb{R}^{h\times w\times c}$). The former is referred as a dense bottleneck and the latter as a spatial bottleneck.

All models were implemented using Python 3.6.8, PyTorch 1.0.1, CUDA 10.0.130 and trained on a NVIDIA GeForce RTX 2080 Ti GPU with batches of 32 images. After each convolutional layer, batch normalization [6] was applied for its regularization properties. The nonlinear activation function in each layer was the rectified linear unit (ReLU), except for the last layer where a sigmoid was employed in order to have output pixels normalized between $[0, 1]$. The loss functions were optimized using Adam [7], a popular optimization algorithm for training deep neural networks.

3.1. Spatial autoencoder (sAE)

The spatial autoencoder model is fully convolutional, 4 convolutional layers go from input to bottleneck and 4 transposed convolutional layers from bottleneck to output. As depicted in Figure 1-B, the output of the encoder network is directly the latent vector z and loss function is simply the reconstruction error:

$$\mathcal{L} = \|x - \hat{x}\|_1 \tag{2}$$

This model was trained for 160 epochs, with a learning rate of 10^{-3}. 5×5 kernels were convolved using padding of 1 pixel and a stride of (2, 2), and there were no pooling layers.

3.2. Spatial variational autoencoder (sVAE)

Our spatial variational autoencoder model is shown in Figure 1-C. Similar to sAE, the model is fully convolutional, however, the encoder generated the parameters of the approximate posterior of the latent variable given the input, constrained to follow a multivariate normal distribution. A sampling operation was needed to obtain an actual value for z.

Training lasted for 200 epochs using a learning rate of 0.3×10^{-3}. A 5×5 kernel was chosen as filter, along with a padding of 1 and stride of (2, 2). No pooling layers were used. The loss function was computed as follows:

$$\mathcal{L} = \lambda\|x - \hat{x}\|_1 + (1-\lambda)\Big[-\tfrac{1}{2}\sum_{j=1}^{J}(1 + \log((\sigma_j)^2) - (\mu_j)^2 - (\sigma_j)^2)\Big] \tag{3}$$

where μ and σ denote the mean and the variance of the approximate posterior, J is the number of dimensions of the latent space and λ controls the proportions between the two terms. The first term is the reconstruction error and the second term is the Kullback-Leibler (KL) divergence between the approximate posterior and the prior of the latent variable, for the Gaussian case [8]. To favor good reconstructions over a Gaussian-like distribution of the latent variables, we put more weight (90%) in the reconstruction term and less weight (10%) in the KL divergence term.

3.3. Dense variational autoencoder (dVAE)

The main difference of the dense variational autoencoder when compared to the sVAE is its dense bottleneck. Encoder and decoder also have fully connected layers in addition to the convolutional layers, as shown in Figure 1-D.

For regularization purposes, the dropout [9] technique was used to turn off 30% of the units in fully-connected layers during training. This model was trained for 100 epochs with a learning rate of 0.3×10^{-3}. There were no pooling layers. Kernels for all convolutional layers were 5×5, and convolutions were performed with a padding of 1 and stride of (2, 2). The dVAE shared the same loss function as the sVAE and we kept the same 90/10 proportion between the reconstruction term and the KL divergence term.

4. PD ANOMALY DETECTION

A reference model is learned from healthy MR images using the autoencoder-based models presented in Section 3. During the training process the model parameters are tuned so as to minimize a loss function that favors good reconstructions. Since the network was solely trained on healthy subjects, we hypothesized that the MR scans of the PD population would

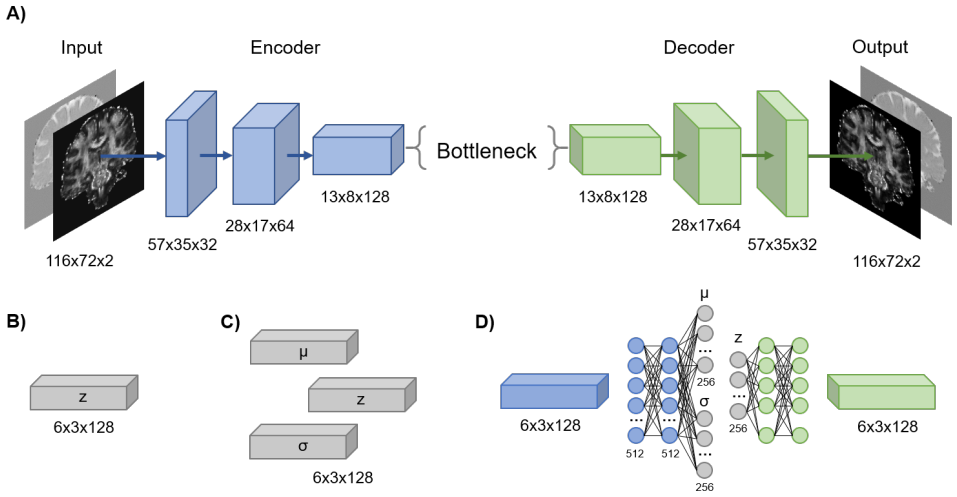

Fig. 1. A) The general architecture of the implemented autoencoders with an unspecified bottleneck; B) sAE spatial bottleneck; C) sVAE spatial bottleneck and D) fourth convolutional layer of the dVAE along with its fully connected layers and dense bottleneck. μ and σ describe the approximate posterior of the latent variable, z is obtained by a sampling operation.

have greater reconstruction errors in some regions. The idea is therefore to use the reconstruction error as an anomaly score.

The voxel-wise reconstruction errors in one image can be computed as $|x_i - \hat{x}_i|$. Seeing that our decoders output two images ($\hat{\text{FA}}$, $\hat{\text{MD}}$), we defined the joint reconstruction error of every voxel as:

$$\sqrt{(\text{FA}_i - \hat{\text{FA}}_i)^2 + (\text{MD}_i - \hat{\text{MD}}_i)^2} \tag{4}$$

We identified four sources of reconstruction errors : 1) noise in the input data, 2) loss of information due to dimension reduction in the latent space, 3) variability of healthy controls not captured by the model and 4) finally real anomalies caused by PD. Because we were only interested in the latter, the best way to evaluate and compare the models is by measuring their ability to discriminate between controls and PD patients, based on the intensity and localization of the reconstruction errors.

To help evaluate the localization of anomalies, we employed the MNI PD25 atlas [10], specifically designed for PD patient exploration. It contains 8 regions: substantia nigra (SN), red nucleus (RN), subthalamic nucleus (STN), globus pallidus interna and externa (GPi, GPe), thalamus, putamen and caudate nucleus. In addition, we considered the superior colliculus (SC) and the inferior colliculus (IC), where we recently found functional deficits [11].

Regarding comparisons of reconstruction error intensities, we investigated extreme reconstruction errors with the idea that PD patients should exhibit very abnormal voxels in larger quantities. Accordingly, we considered an extreme quantile (eg. the 95% quantile) of the distribution of errors in the control population as possible threshold value to decide whether or not a given voxel was considered as abnormal. For each control or PD subject, we counted the number of extreme abnormalities detected in every structure. The idea being to classify a subject as PD or healthy when this number was above a certain value. The critical choice of this value was investigated using a ROC curve of sensitivity and specificity to account for class imbalance. We chose the AUC as our principal indicator of discrimination performance.

5. RESULTS: RECONSTRUCTION ASSESSMENT

We used the reconstruction errors of every structure in every subject to classify patients from controls as explained in Section 4. Amongst the three models tested, we noticed that dVAE detected the biggest reconstruction errors, as an example the mean reconstruction error of the control's subcortical structures is of 0.075, 0.086 and 0.106 for the sAE, sVAE and dVAE respectively. However, the absence of ground truth at a voxel level prevented us from determining which model was the most accurate relative to PD abnormalities.

The number of voxels over the 95% quantile abnormality threshold were generally superior in a patient than in a healthy control. We employed the ROC curve as a performance measurement for our classification problem, we were able to measure the AUC to have an indication on the ability of the model to distinguish between patients and healthy controls. The structures with the highest AUC were the Substantia Nigra, the Red Nucleus, the Thalamus and the combination of all subcortical structures as seen in Table 1. This was in accordance with the literature. The total white matter present in the reconstructed slices obtained even better results with an AUC of 0.83, 0.80 and 0.74 for the sAE, sVAE and dVAE respectively.

	sAE	sVAE	dVAE
Red Nucleus	0.75	0.76	0.65
Substantia Nigra	0.74	0.73	0.72
Sub-thalamic Nucleus	0.59	0.64	0.53
Caudates	0.70	0.64	0.61
Putamen	0.72	0.74	0.63
Globus Pallidus ext.	0.65	0.69	0.69
Globus Pallidus int.	0.69	0.69	0.71
Thalamus	0.71	0.73	0.72
Superior Colliculus	0.51	0.59	0.54
Inferior Colliculus	0.54	0.56	0.49
All subcortical structures	0.76	0.77	0.73
White Matter	0.83	0.80	0.74

Table 1. AUC values obtained in different structures when counting the number of voxels above the 95% percentile to discriminate between patients and controls

When assessing the reconstructions of FA and MD separately, FA obtained the highest ROC AUC for the caudates and the putamen, whereas for the white matter and the red nucleus MD was better. The substantia nigra and the ensemble of all subcortical structures benefited from the joint measures of FA and MD.

6. DISCUSSION AND CONCLUSION

Although preliminary, these results offer compelling evidence that deep learning-based models are useful to identify subtle anomalies in *de novo* PD patients, even when trained with a moderate number of images and only two parametric maps as input. The good discriminative performances of the sub-cortical structures are in accordance to our physiopathological knowledge of PD. The dopaminergic neuron deficit in Substantia Nigra is known as critical in the development of PD. What is more, the absence of motor symptoms in the pre-clinical stages of PD may be the result of compensation mechanisms involving the structures in the motor coordination pathways of the brain such as the Thalamus and the Red Nucleus [12].

We have shown that no structural changes, including in the White Matter (WM), can be robustly observed from T1-weighted images to automatically distinguish between controls and *de novo* PD patients [2]. To explain the good performances of autoencoders based on white matter voxels, we may hypothesize that 1) diffusion parameters (FA an MD) are more informative than grey levels from T1-weighted imaging, or that 2) they dispose of more voxels, compared to subcortical structures, to build a model that captures the variability of healthy controls.

Undeniably, experiments on a larger cohort are necessary to confirm our results. The control group available for our experiments contained only 57 MRI volumes, with gender and age imbalance. Although we were able to discriminate between healthy controls and individuals affected by PD with good performances, we cannot rule out that other possible causes of variability in brain properties, such as age and gender, and other hidden parameters, might have influenced our classification performance.

Similar anomaly detection frameworks could be of interest for studying other neurological disorders where small lesions are suspected and difficult to localize for a human observer. In future work, it could be beneficial to include other quantitative MR measures, such as perfusion and relaxation times, to encode a more complete picture of the pathophysiology of the disease.

7. REFERENCES

[1] P. Tofts, *Quantitative MRI of the brain*, Wiley, 1st edition, 2005.

[2] V. Munoz Ramirez & al., "No structural differences are revealed by vbm in de novo parkinsonian patients," in *MEDINFO*, 2019.

[3] M. Pimentel & al., "Review: A review of novelty detection," *Signal Process.*, vol. 99, pp. 215–249, June 2014.

[4] "MICCAI 2018 Proceedings," in *MICCAI*, Granada, 2018, Springer, Cham.

[5] C. Baur & al., "Deep autoencoding models for unsupervised anomaly segmentation in brain mr images," in *Brainlesion*. 2019, pp. 161–169, Springer Intl.

[6] S. Ioffe & C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *ICML*, 2015, pp. 448–456.

[7] D. Kingma & J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv e-prints*, Dec 2014.

[8] D. Kingma & M. Welling, "Auto-encoding variational bayes," *CoRR*, 2013.

[9] N. Srivastava & al., "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[10] Y. Xiao & al., "Multi-contrast unbiased MRI atlas of a Parkinson's disease population," *Int. J. Comput. Assist. Radiol. Surg*, vol. 10, no. 3, pp. 329–341, 2015.

[11] E. Bellot & al., "First steps of visual processing are altered in de novo Parkinson's disease patients," in *IAPRD Conf, Lyon, FR*, 2018.

[12] I. Philippens & al., "Involvement of the Red Nucleus in the Compensation of Parkinsonism may Explain why Primates can develop Stable Parkinson 's Disease," *Sci. Rep.*, pp. 1–9, 2019.