Generating PET-derived maps of myelin content from clinical MRI using curricular discriminator training in generative adversarial networks

Théodore Soulier, Mariem Hamzaoui, Milena Sales Pitombeira, Daniele De Paula Faria, Arya Yazdan-Panah, Matteo Tonietto, Claire Leroy, Michel Bottlaender, Benedetta Bodini, Ninon Burgos, et al.

Generating PET-derived maps of myelin content from clinical MRI using curricular discriminator training in generative adversarial networks

Théodore Soulier*[a], Mariem Hamzaoui*[a], Milena Sales Pitombeira[b], Daniele De Paula Faria[c], Arya Yazdan-Panah[a,d], Matteo Tonietto[e], Claire Leroy[e], Michel Bottlaender[e], Benedetta Bodini[a,f], Ninon Burgos[d], Nicholas Ayache[g], Olivier Colliot[d], and Bruno Stankoff[a,f]

[a]Sorbonne Université, Paris Brain Institute - ICM, CNRS, Inserm, AP-HP, Paris, France
[b]Universidade de São Paulo, Department of Neurology, São Paulo, Brazil
[c]Universidade de São Paulo, Laboratory of Nuclear Medicine (LIM43), Department of Radiology and Oncology, Faculdade de Medicina FMUSP, São Paulo, Brazil
[d]Sorbonne Université, Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Paris, France
[e]Comissariat à l'Energie Atomique, Service Hospitalier Frédéric Joliot, Orsay, France
[f]AP-HP, Hôpital Saint-Antoine, Department of Neurology, DMU Neurosciences, Paris, France
[g]Inria, Epione Project-Team, Sophia-Antipolis, France

ABSTRACT

Multiple sclerosis (MS) is a demyenalinating inflammatory neurological disease. In vivo biomarkers of myelin content are of major importance for patient care and clinical trials. Positron Emission Tomography (PET) with Pittsburgh Compound B (PiB) provides a specific myelin marker. However, it is not available in clinical routine. In this paper, we propose a method to generate myelin maps by synthesizing PiB PET from clinical routine MRI sequences (T1-weighted and FLAIR). To that purpose, we introduce a new curriculum learning strategy for training generative adversarial networks (GAN). Specifically, we design a curricular approach for training the discriminator: training starts with only lesion patches and random patches (from anywhere in the white matter) are progressively introduced. We relied on two distinct cohorts of MS patients acquired each on a different scanner and in a different country. One cohort was used for training/validation and the other one for testing. We found that the synthetic PiB PET was strongly correlated to the ground-truth both at the lesion level ($r = 0.70$, $p < 10^{-5}$) and the patient level ($r = 0.74$, $p < 10^{-5}$). Moreover, the correlations were stronger when using the curricular learning strategy compared to starting the discriminator training from random patches. Our results demonstrate the interest of this new curriculum learning strategy for PET image synthesis. Even though further evaluations are needed, our approach has the potential to provide a useful biomarker for clinical routine follow-up of patients with MS.

Keywords: Multiple Sclerosis, Magnetic Resonance Imaging, Positron Emission Tomography, Myelin, Generative Adversarial Network, Curriculum Learning, Image synthesis

1. INTRODUCTION

Multiple Sclerosis (MS) is the leading cause of disability among young adults in North America and Europe; it's a chronic demyelinating inflammatory disease of the central nervous system, characterized by lesions visible on Magnetic Resonance Imaging (MRI).[1] While MRI in a clinical setting provides an accessible non invasive tool for patient's routine follow-up, it doesn't give access to specific information about the biological mechanisms underlying the disease such as the myelin content.

* contributed equally to this work
Further author information: (Send correspondence to B.S.)
B.S.: E-mail: bruno.stankoff@aphp.fr

Positron Emission Tomography (PET) with Pittsburgh Compound B (PiB) provides a specific myelin marker.[2] However, it is expensive, invasive, irradiating and not easily accessible for patients. There is still a great need for an accessible and non-invasive MRI approach able to specifically quantify myelin content. Aiming to reach this goal, previous works[3,4] used conditional generative adversarial networks (cGANs) for MRI-to-PET image synthesis. However, these works rely on MRI sequences (magnetization transfer ratio, diffusion tensor imaging) which are mainly used for research purposes and are not performed as part of clinical routine. Therefore, it would be of considerable usefulness to synthesize the myelin information contained in PIB-PET from sequences available in clinical routine.

In this paper, we propose, to the best of our knowledge for the first time, a new approach for PiB-PET synthesis to obtain myelin cartography from clinical routine MRI sequences, namely T1-weighted (T1w) and FLuid Attenuated Inversion Recovery (FLAIR). The method is based on cGANs. The originality of our approach is to introduce a new curriculum learning strategy for the discriminator which drives the learning to focus on pathological features of the disease by the use of a progressive anatomical sampling of the discriminator patches.

2. MATERIALS

2.1 Datasets

We used two independent cohorts: one for training/validation and one for testing. The two datasets were acquired on different scanners and in different countries (France and Brazil). In both cohorts, the patients underwent T1w and FLAIR MRI and [^{11}C] PiB PET quantified by Distribution Volume Ratios (DVR). The study was approved by the respective Institutional Ethical Committees and patients gave written informed consent.

Training/validation cohort (France) The training/validation cohort comprised 17 MS patients. MRI was acquired on a Siemens 3T machine and PET on a Siemens High Resolution Research Tomograph (HRRT). Subjects characteristics and PET quantification procedures have been described in a previous paper.[2] Two patients had only one usable visit: they constituted our validation set. For the 15 remaining patients, we used their two visits for the training set which was thus constituted of 30 visits.

Testing cohort (Brazil) The independent test set was composed of 44 MS patients who underwent MRI and PET scans on a General Electric PET-MR system. Subjects characteristics, acquisition details and PET quantification have been described in a previous publication.[5]

2.2 Preprocessing

MS lesions were manually segmented on MRI by trained raters on T2 sequences. White matter (WM) segmentation and brain mask extraction was done using Freesurfer after lesion filling. Normal appearing white matter (NAWM) was defined as the white matter mask after removing MS lesions. T1w and FLAIR images were linearly coregistered onto the PET space for each patient using Advanced Normalization Tools (ANTS)*.[6] All images were cropped to focus on the signal within the brain, and re-sized to a final dimension of $128 \times 128 \times 128$ voxels. Each subject was processed in its own PET space. T1w and FLAIR images were rescaled between $[-1, 1]$ excluding the 5% of extreme values in their native intensities histograms. PiB-DVR values were not altered nor adjusted in any way.

3. METHODS

3.1 Proposed approach

Our approach is based on the introduction of a new curricular strategy for discriminator training in a cGAN.[7] Fig. 1 summarizes our approach. We use a 3D UNet[8] for the generator, and a patch-based convolutional neural network (CNN) classifier for the discriminator. The discriminator uses patches of size $16 \times 16 \times 16$. A patch can be either centered on the center of gravity of a randomly chosen lesion (what we call *lesion patch*) or centered

*http://stnava.github.io/ANTs/

on a random point within the WM mask (that could be lesional or not, what we call *random patch*). We propose a curricular strategy for training the discriminator: at the first epoch only *lesion patches* are considered for training, and the proportion of *random patches* linearly increases throughout epochs until it reaches 100% at epoch 150. We compared this approach to an anticurricular training strategy, which is the exact opposite, starting with *random patches* and progressively giving only *lesion patches*. The generator loss is the sum of the pixel loss (L1) and the adversarial loss (Mean Squared Error (MSE) between the prediction of the discriminator for generated patches and the true label).

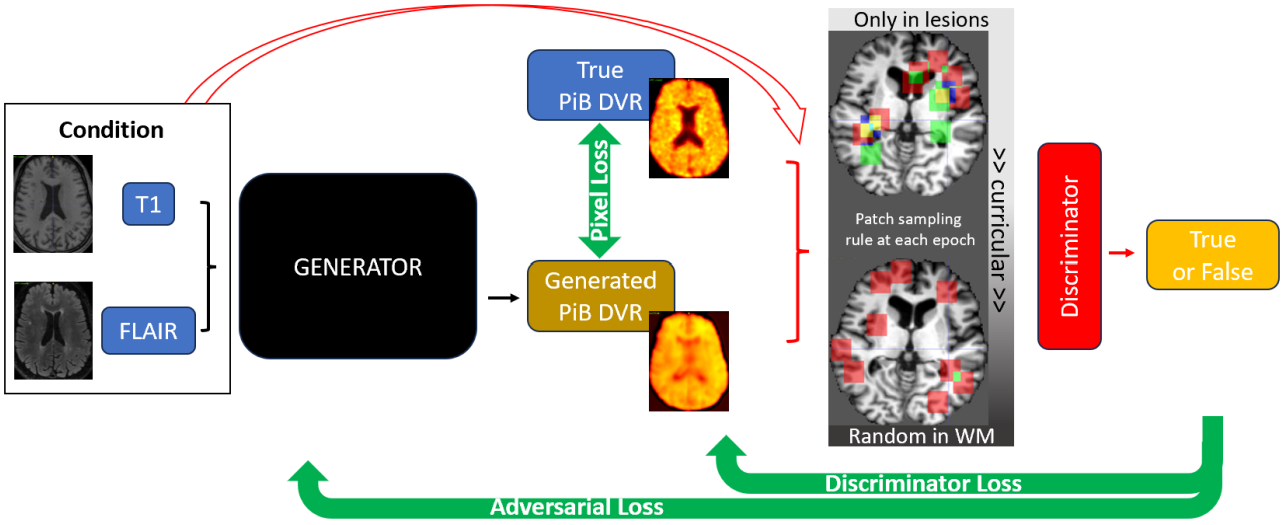

Figure 1. Summary of the proposed approach.

3.2 Implementation

The generator architecture is the following (resulting in 16,471,077 trainable parameters):

$$B_E(32) \downarrow\rightarrow B_E(64) \downarrow\rightarrow B_E(128) \downarrow\rightarrow B_E(256) \downarrow\rightarrow B_E(512) \rightarrow\uparrow B_D(256) \rightarrow\uparrow B_D(128) \rightarrow\uparrow B_D(64) \rightarrow\uparrow B_D(32) \rightarrow 3D\ Convolution(1)$$

with $B_{Block}(\mathbf{n}) := 2 \times \{3D\ Convolution(\mathbf{n}) \rightarrow Group\ Normalization \rightarrow ReLU\}$ and skip connections between Encoder blocks $B_E(n)$ and Decoder blocks $B_D(n)$.
The discriminator architecture is the following (resulting in 4,653,441 trainable parameters):

$$B'(64) \rightarrow B'(128) \rightarrow B'(256) \rightarrow B'(512) \rightarrow L(1)$$

with $B'(\mathbf{n}) := \{3D\ Convolution(\mathbf{n}),\ striding = 3 \rightarrow Batch\ Normalization \rightarrow LeakyReLU\}$ and $L(\mathbf{n}) := \{Flatten,\ Linear(\mathbf{n}),\ Sigmoid\}$
In the above, $\mathbf{n}$ represents the number of filters for convolutive blocks (B_{Block} and B') or the number of output neurons for Linear blocks (L), E and D for Encoder and Decoder block respectively, $\downarrow$ indicates max pooling, $\uparrow$ indicates trilinear upsampling.

To facilitate training, we smoothed labels with a Gaussian filter with a mean of 0.8 (for labels equal to 1) or 0.2 (otherwise) and standard deviation of 0.3. The batch size was 1 (the whole volume) for the generator and 48 patches for the discriminator. We trained for 150 epochs with Adam optimizer. We used a variable learning rate in order to prevent situations where either the generator or the discriminator loss becomes uninformative.

4. EXPERIMENTS AND RESULTS

We computed Pearson correlations between synthetic and true PiB signals at two levels: lesion level and patient level. At the lesion level, we pooled all lesions of all subjects of the testing cohort (suppressing those with size below $50mm^3$ to maintain the integrity of PET resolution, resulting in 528 lesions above $50mm^3$). At the patient level, we averaged the mean difference between actual and synthetic PiB DVR signal of all lesions for a given patient, weighted by lesion's volume. At the lesion level, results of the curricular training showed a higher correlation (r = 0.70, $p < 10^{-5}$) between the mean true PiB DVR and the mean synthetic PiB DVR compared to the anti-curricular training (r = 0.62, $p < 10^{-5}$). At the patient level, we also found a stronger correlation (r = 0.74, $p < 10^{-5}$) for curricular training than for anticurricular training (r = 0.61, $p < 10^{-5}$). An example of result in an individual patient from the testing set is shown on Fig. 2.

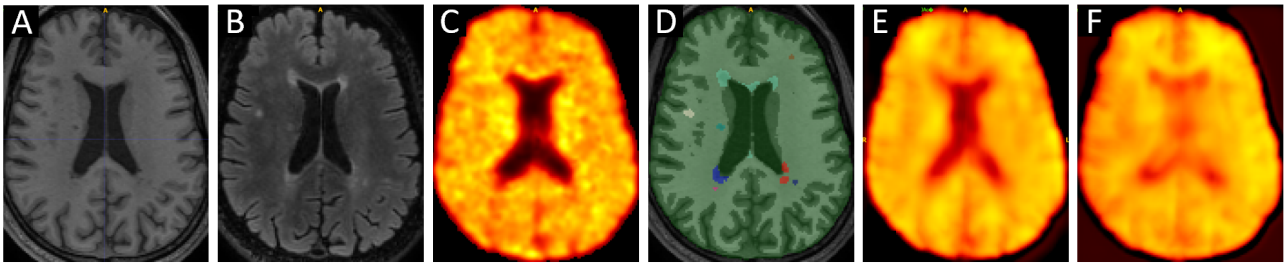

Figure 2. On the same testing patient, A: T1w, B: FLAIR, C: True PiB DVR, D: different Regions of Interest (brain mask in green, white matter in white, lesions in different colors), E: synthetic PiB DVR from curricular training, F: synthetic PiB DVR from anti-curricular training.

5. DISCUSSION

In this paper, we proposed a new approach based on curricular training of the discriminator in a conditional GAN to synthetize PET-derived myelin content, from clinical routine MRI sequences (T1w and FLAIR). On our independent external testing cohort, curricular training improved myelin content synthesis, with a strong correlation between synthetic and true lesional PiB DVR both at lesional level and at patient level. To the best of our knowledge, this is the first method that can synthesize myelin content from clinical routine MRI sequences.

A way to understand the superiority of curricular training is that lesions display more variability in their signal, so that to fool the discriminator, the generator has to start by providing credible outputs to the discriminator in the most variable regions. On the contrary, starting from randomly sampled patches in the WM (anti-curricular training) pushes the generator towards learning mainly the general anatomical relationship between T1w/FLAIR images and PiB DVR at the expense of pathological regions of interest like the demyelinated lesions.

Curriculum learning[9] has already been used in medical imaging, e.g. for classification of bone fractures[10] and mammograms,[11] or segmentation of cardiac MRI.[12] As for image synthesis, curriculum learning has been introduced in GANs[13,14] on 2D non medical images. Our results highlight, in a clinical context of 3D medical images, the interest of the curricular training process for focusing on relevant lesional areas in image synthesis.

Although our evaluation remains preliminary, our method has the potential to ultimately provide a clinically useful biomarker. A strength of our study is that the independent testing set was acquired on a different machine and in a different country, which is promising for generalizability. Moreover, the strong correlation in lesional myelin content at the patient level opens the perspective to apply our method to large-scale clinical MS cohorts (in which PET is not available, unlike T1w and FLAIR MRI) to study the correlation between the synthetic PiB DVR and clinical data such as disability scores. If such clinical correlations were to be demonstrated, it would increase the potential of our method as a disease biomarker usable for the routine follow-up of patients routine.

Our work has the following limitations. First, we did not assess here the impact of fixed oversampling strategies between lesion patches and white matter patches. Also, we have not yet performed a comparison between our approach and other strategies to focus the synthesis such as attention maps.[4] This will be done in future work.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10- IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo- Universitaire-6), and from ICM under the Big Brain Theory program (project IMAGIN-DEAL in MS). This work was supported by the Fondation pour la Recherche Médicale, grant number FDM202006011247 to Théodore Soulier and by the Fondation Sorbonne Université to Mariem Hamzaoui. Nicholas Ayache's research is also partially funded by the 3IA Côte d'Azur (reference number ANR-19-P3IA-0002). PET MRI acquisitions from Brazil were funded by General Electrics Healthcare, No.12496139131.

REFERENCES

[1] Filippi, M., Preziosa, P., Banwell, B., Barkhof, F., Ciccarelli, O., De Stefano, N., Geurts, J. J. G., Paul, F., Reich, D. S., Toosy, A. T., Traboulsee, A., Wattjes, M. P., Yousry, T. A., Gass, A., Lubetzki, C., Weinshenker, B. G., and Rocca, M. R., "Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines," *Brain* **142(7)**, 1858-75 (2019).

[2] Bodini, B., Veronese, M., García-Lorenzo, D., Battaglini, M., Poirion, E., Chardain, A., Freeman, L., Louapre, C., Tchikviladze, M., Papeix, C., Dollé, F., Zalc, B., Lubetzki, C., Bottlaender, M., Turkheimer, F., and Stankoff, B., "Dynamic imaging of individual remyelination profiles in multiple sclerosis," *Annals of Neurology* **79(5)**, 726–38 (2016).

[3] Wei, W., Poirion, E., Bodini, B., Durrleman, S., Ayache, N., Stankoff, B., and Colliot, O., "Predicting PET-derived demyelination from multimodal MRI using sketcher-refiner adversarial training for multiple sclerosis," *Medical image analysis* **58**, 101546 (2019).

[4] Wei, W., Poirion, E., Bodini, B., Tonietto, M., Durrleman, S., Colliot, O., Stankoff, B., and Ayache, N., "Predicting PET-derived myelin content from multisequence MRI for individual longitudinal analysis in multiple sclerosis," *Neuroimage* **223**, 117308 (2020).

[5] Sales Pitombeira, M., Koole, M., Campanholo, K. R., Souza, A. M., Duran, F. L. S., Fontoura Solla, D. J., Mendes, M. F., Apóstolos Pereira, S. L., Rimkus, M. C., Filho Busatto, G., Callegaro, D., Buchpiguel, C. A., and De Paula Faria, D., "Innate immune cells and myelin profile in multiple sclerosis: a multi-tracer PET/MR study," *Eur J Nucl Med Mol Imaging* **49(13)**, 4551–66 (2022).

[6] Avants, B. B., Tustison, N., and Song, G., "Advanced normalization tools (ants)," *The Insight Journal* **2**(365), 1–35 (2009).

[7] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A., "Image-to-image translation with conditional adversarial networks," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 1125–1134 (2017).

[8] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O., "3d u-net: learning dense volumetric segmentation from sparse annotation," in [*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016*], 424–432, Springer (2016).

[9] Soviany, P., Ionescu, R. T., Rota, P., and Sebe, N., "Curriculum learning: A survey," *International Journal of Computer Vision* **130**(6), 1526–1565 (2022).

[10] Jiménez-Sánchez, A., Mateus, D., Kirchhoff, S., Kirchhoff, C., Biberthaler, P., Navab, N., Ballester, M. A. G., and Piella, G., "Medical-based deep curriculum learning for improved fracture classification," in [*Lecture Notes in Computer Science*], 694–702, Springer International Publishing (2019).

[11] Lotter, W., Sorensen, G., and Cox, D., "A multi-scale CNN and curriculum learning strategy for mammogram classification," in [*Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*], 169–177, Springer International Publishing, Cham (2017).

[12] Liu, Z., Manh, V., Yang, X., Huang, X., Lekadir, K., Campello, V., Ravikumar, N., Frangi, A. F., and Ni, D., "Style curriculum learning for robust medical image segmentation," in [*Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*], 451–460, Springer (2021).

[13] Soviany, P., Ardei, C., Ionescu, R. T., and Leordeanu, M., "Image difficulty curriculum for generative adversarial networks (CuGAN)," in [*2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*], 3452–3461 (2020).

[14] Doan, T., Monteiro, J., Albuquerque, I., Mazoure, B., Durand, A., Pineau, J., and Hjelm, R. D., "On-line adaptative curriculum learning for GANs," in [*Proceedings of the AAAI Conference on Artificial Intelligence*], **33**(01), 3470–3477 (2019).