



**HAL**  
open science

## Contrast-enhanced to non-contrast-enhanced image translation to exploit a clinical data warehouse of T1-weighted brain MRI

Simona Bottani, Elina Thibeau-Sutre, Aurélien Maire, Sebastian Ströer, Didier Dormont, Olivier Colliot, Ninon Burgos

### ► To cite this version:

Simona Bottani, Elina Thibeau-Sutre, Aurélien Maire, Sebastian Ströer, Didier Dormont, et al.. Contrast-enhanced to non-contrast-enhanced image translation to exploit a clinical data warehouse of T1-weighted brain MRI. BMC Medical Imaging, 2024, 24 (1), pp.67. 10.1186/s12880-024-01242-3 . hal-03497645v1

**HAL Id: hal-03497645**

**<https://hal.science/hal-03497645v1>**

Submitted on 20 Dec 2021 (v1), last revised 21 Mar 2024 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Homogenization of brain MRI from a clinical data warehouse using contrast-enhanced to non-contrast-enhanced image translation

Simona Bottani<sup>a,b,c,d,e,f</sup>, Elina Thibeau-Sutre<sup>b,c,d,e,f,a</sup>, Aurélien Maire<sup>g</sup>, Sebastian Ströer<sup>h</sup>,  
Didier Dormont<sup>h,b,c,d,e,f,a</sup>, Olivier Colliot<sup>b,c,d,e,f,a</sup>, Ninon Burgos<sup>b,c,d,e,f,a,\*</sup>, the APPRIMAGE  
Study Group<sup>†</sup>

<sup>a</sup>Inria, Aramis project-team, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>b</sup>Sorbonne Université, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>c</sup>Institut du Cerveau - Paris Brain Institute, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>d</sup>Inserm, U 1127, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>e</sup>CNRS, UMR 7225, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>f</sup>AP-HP, Hôpital Pitié-Salpêtrière, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>g</sup>AP-HP, WIND department, 33 boulevard Picpus, 75012 Paris, France

<sup>h</sup>AP-HP, Hôpital Pitié Salpêtrière, Department of Neuroradiology, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>†</sup>Members of the study group are listed in section 5

## Abstract.

**Purpose** Clinical data warehouses (CDW) provide access to massive amounts of medical images, but these images are often heterogeneous. They can for instance include images acquired both with or without the injection of a gadolinium-based contrast agent. Harmonizing such data sets is thus fundamental to guarantee unbiased results, for example when performing differential diagnosis. Furthermore, classical neuroimaging software tools for feature extraction are typically applied only to images without gadolinium. The objective of this work is to homogenize images from a CDW and enable the extraction of consistent features from brain MR images, no matter the initial presence or absence of gadolinium.

**Approach** We propose and compare different 3D U-Net and conditional GAN models to convert contrast-enhanced T1-weighted (T1w-ce) into non-contrast-enhanced (T1w-nce) brain MRI. These models were trained using 230 image pairs and tested on 77 image pairs from the CDW of the Greater Paris area.

**Results** Validation using standard image similarity measures demonstrated that the similarity between real and synthetic T1w-nce images was higher than between real T1w-nce and T1w-ce images for all the models compared. The best performing models were further validated on a segmentation task. We showed that tissue volumes extracted from synthetic T1w-nce images were closer to those of real T1w-nce images than volumes extracted from T1-ce images.

**Conclusions** We showed that deep learning models could synthesize T1w-nce from T1w-ce images and that reliable features could be extracted from the synthetic images, thus demonstrating the ability of such methods to homogenize a data set coming from a CDW.

**Keywords:** Image translation, Clinical data warehouse, Gadolinium injection, Anatomical MRI, Brain, Deep learning.

\*Corresponding author: Ninon Burgos ([ninon.burgos@cnrs.fr](mailto:ninon.burgos@cnrs.fr))

## 37 **1 Introduction**

38 Clinical data warehouses, gathering hundreds of thousands of medical images from numerous  
39 hospitals, offer unprecedented opportunities for research. They can for example be used to develop  
40 and validate machine learning and deep learning algorithms for the computer-aided diagnosis of  
41 neurological diseases. However, they also pose important challenges, a major challenge being their  
42 heterogeneity. Neurological diseases can result in a variety of brain lesions that are each studied  
43 with specific magnetic resonance imaging (MRI) sequences. For example, T1-weighted (T1w)  
44 brain MR images enhanced with a gadolinium-based contrast agent are used to study lesions such  
45 as tumors, and T1w images without gadolinium are used to study neurodegenerative diseases.

46 To perform differential diagnosis using classification algorithms, homogeneous features must  
47 be extracted from the images, no matter the disease, otherwise a link could be established between  
48 MRI sequence and pathology, which would create bias. This is critical as differential diagnosis in  
49 a clinical setting can be more challenging than in a research setting as different diseases may co-  
50 exist. Software tools such as SPM,<sup>1</sup> ANTs<sup>2</sup> or FSL<sup>3</sup> have been widely used for feature extraction  
51 but they were largely validated using structural T1w MRI without gadolinium, to the best of our  
52 knowledge, and their good performance on images with gadolinium is thus not guaranteed. A  
53 solution could then be to convert contrast-enhanced T1w (T1w-ce) into non-contrast-enhanced  
54 T1w (T1w-nce) brain MRI before using such tools.

55 Deep learning has been widely used in the image translation domain. The U-Net and condi-  
56 tional generative adversarial networks (GANs) appear as the two most popular options. The U-Net  
57 was originally proposed for image segmentation:<sup>4</sup> an encoder with convolutional and downsam-  
58 pling blocks is followed by a decoder with upsampling and convolutional layers. The skip con-

59 nections linking the encoder and decoder blocks at the same level enable the reconstruction of  
60 fine-grained details, explaining the popularity of this architecture for image translation.<sup>5-12</sup> Con-  
61 ditional GANs consist of a generator, which may adopt the U-Net architecture, followed by a  
62 discriminator in charge of distinguishing synthetic from real images and challenging the generator  
63 so that it improves the quality of the generated images. The good results obtained with conditional  
64 GANs explain their wide use for image translation.<sup>13-22</sup>

65 Both U-Net like models and conditional GANs have been proposed for diverse applications.  
66 Some aim to enhance the quality of the input images, for example by reducing noise in MRI<sup>23-25</sup> or  
67 positron emission tomography<sup>26</sup> images or by performing super-resolution.<sup>13,15,27-29</sup> Other works  
68 aim to translate an image of a particular modality into another modality, such as an MRI into  
69 an X-ray computed tomography (CT)<sup>7,8,12,17,18,30</sup> or a particular MRI sequence into another se-  
70 quence.<sup>19-22</sup> The U-Net architecture has also been used for the data harmonization: Dewey et al.<sup>31</sup>  
71 built Deep-Harmony that aims to homogenize the contrast between images coming from different  
72 sites.

73 Closer to our application, various deep learning models have been developed for the synthesis  
74 of images with gadolinium from images without gadolinium: they include reinforcement learning  
75 for liver MRI,<sup>32</sup> or Gaussian mixture modeling for CT images.<sup>33</sup> As for the other image translation  
76 tasks, 3D U-Net like models have also been used to convert T1w-nce into T1w-ce images.<sup>34-36</sup> In  
77 two studies,<sup>34,35</sup> multimodal MRI sequences were used as input of the 3D U-Net that was trained  
78 and tested on patients with brain cancers. More specifically, the 3D U-Net proposed in the first  
79 study<sup>35</sup> predicts patches of T1w-ce, while the one in the second study<sup>34</sup> directly predicts the full 3D  
80 T1w-ce image. The residual attention U-Net described in the last work<sup>36</sup> outputs synthetic T1w-ce  
81 that are used for the evaluation of cerebral blood volume in mice, instead of the real T1w-ce.

82 Our objective in this work was to obtain a homogeneous data set of T1w-nce images from very  
83 heterogeneous images coming from a clinical data warehouse. This homogenization step should  
84 enable a consistent extraction of features that would later be used for computer-aided diagnosis in  
85 a clinical setting. We thus developed and compared different deep learning models that rely on  
86 typical architectures used in the medical image translation domain to convert T1w-ce into T1w-  
87 nce images. In particular, we implemented 3D U-Net like models with the addition of residual  
88 connections, attention modules or transformer layers. We also used these 3D U-Net like models  
89 in a conditional GAN setting. We trained and tested our models using 307 pairs of T1w-nce and  
90 T1w-ce images coming from a very large clinical data warehouse (39 different hospitals of the  
91 Greater Paris area). We first assessed synthesis accuracy by comparing real and synthetic T1w-nce  
92 images using standard metrics. We tested our models both on images of good or medium quality  
93 and on images of bad quality to ensure that deep learning models could generate accurate T1w-  
94 nce images no matter the quality of the input T1w-ce images. We then compared the volumes of  
95 gray matter, white matter and cerebrospinal fluid obtained by segmenting the real T1w-nce, real  
96 T1w-ce and synthetic T1w-nce images using SPM<sup>37</sup> in order to verify that features extracted from  
97 synthetic T1w-nce were reliable. Preliminary work is accepted for publication in the proceedings  
98 of the SPIE Medical Imaging 2022 conference.<sup>38</sup> Contributions specific to this paper include the  
99 development of additional models (a 3D U-Net like model with the addition of transformer layers,  
100 and three conditional GAN models using 3D U-Net like models as generators and a patch-based  
101 discriminator) and an extended validation of the segmentation task with a deeper analysis the tissue  
102 volume differences.

## 103 2 MATERIALS AND METHODS

### 104 2.1 Data set description

105 This work relies on a large clinical data set containing all the T1w brain MR images of adult  
106 patients scanned in one of the 39 hospitals of the Greater Paris area (Assistance Publique-Hôpitaux  
107 de Paris [AP-HP]). The data were made available by the AP-HP data warehouse and the study  
108 was approved by the Ethical and Scientific Board of the AP-HP. According to French regulation,  
109 consent was waived as these images were acquired as part of the routine clinical care of the patients.

110 Images were acquired as part of the routine clinical care in the different hospital sites and gath-  
111 ered in a central hospital PACS. Images relevant to the research project were copied to the research  
112 PACS and pseudonymized. They always remain within the hospital network that we accessed re-  
113 motely. Images from this clinical data warehouse are very heterogeneous:<sup>39</sup> they include images  
114 of patients with a wide range of ages (from 18 to more than 90 years old) and diseases, acquired  
115 with different scanners (more than 30 different models) from 1980 up to now.

116 In a previous work,<sup>39</sup> we developed a quality control framework to identify images that are not  
117 proper T1w brain MRIs, to identify acquisitions for which gadolinium was injected, and to rate  
118 the overall image quality defined based on three characteristics: motion, contrast and noise. We  
119 did so by manually annotating 5500 images (out of a batch of 9941 images that were available) to  
120 train and test convolutional neural network (CNN) classifiers. The graphical interface used to man-  
121 ually annotate the images is publicly available ([https://github.com/SimonaBottani/  
122 Quality\\_Control\\_Interface](https://github.com/SimonaBottani/Quality_Control_Interface)).

123 The data set used in this work is composed of 307 pairs of T1w-ce and T1w-nce images that  
124 were extracted from the batch of 9941 images made available by the AP-HP data warehouse. We

125 first selected all the images of low, medium and good quality, excluding images that were not  
126 proper T1w brain MRI,<sup>39</sup> resulting in 7397 images. This selection was based on manual quality  
127 control for 5500 images and on automatic quality control for the remaining 4441 images.<sup>39</sup> In the  
128 same way, the presence or absence of gadolinium-based contrast agent was manually noted for  
129 5500 images, while it was obtained through the application of a CNN classifier for the remaining  
130 4441 images. We then considered only patients having both a T1w-ce and a T1w-nce image at the  
131 same session, with a T1w-nce image of medium or good quality. Finally, to limit heterogeneity  
132 in the training data set, we visually checked all the images and excluded 52 image pairs that were  
133 potential outliers because of extremely large lesions. Among the selected images, 256 image pairs  
134 were of medium and good quality, and 51 image pairs had a T1w-ce of low quality and a T1w-nce  
135 of good or medium quality. In total the data set comprises 614 images: 534 images were acquired  
136 at 3 T and 80 at 1.5 T, 556 images were acquired with a Siemens machine (with seven different  
137 models) and 58 with a GE Healthcare machine (with five different models).

## 138 *2.2 Image preprocessing*

139 All the images were organised using the Brain Imaging Data Structure (BIDS).<sup>40</sup> We applied the  
140 following pre-processing using the ‘t1-linear’ pipeline of Clinica,<sup>41</sup> which is a wrapper of the ANTs  
141 software.<sup>2</sup> Bias field correction was applied using the N4ITK method.<sup>42</sup> An affine registration to  
142 MNI space was performed using the SyN algorithm.<sup>43</sup> The registered images were further rescaled  
143 based on the min and max intensity values, and cropped to remove background resulting in images  
144 of size  $169 \times 208 \times 179$ , with 1 mm isotropic voxels.<sup>44</sup> Finally all the images were resampled to  
145 have a size of  $128 \times 128 \times 128$  using trilinear interpolation in Pytorch.

## 146 2.3 Network architecture

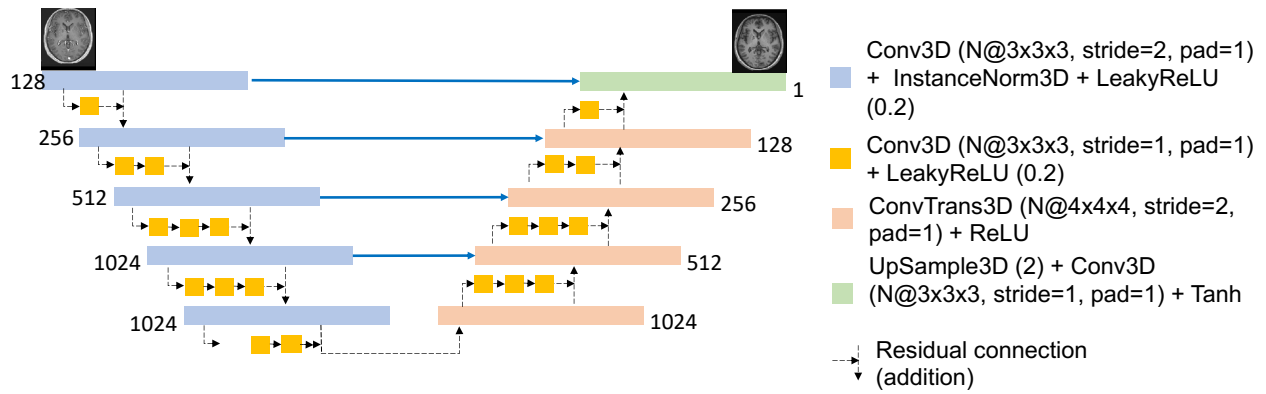
147 To generate T1w-nce from T1w-ce images, both 3D U-Net like models and conditional GANs were  
148 developed and compared. The code used to implement all the architectures and perform the experi-  
149 ments is openly available ([https://github.com/SimonaBottani/image\\_synthesis](https://github.com/SimonaBottani/image_synthesis)).

### 150 2.3.1 3D U-Net like structures

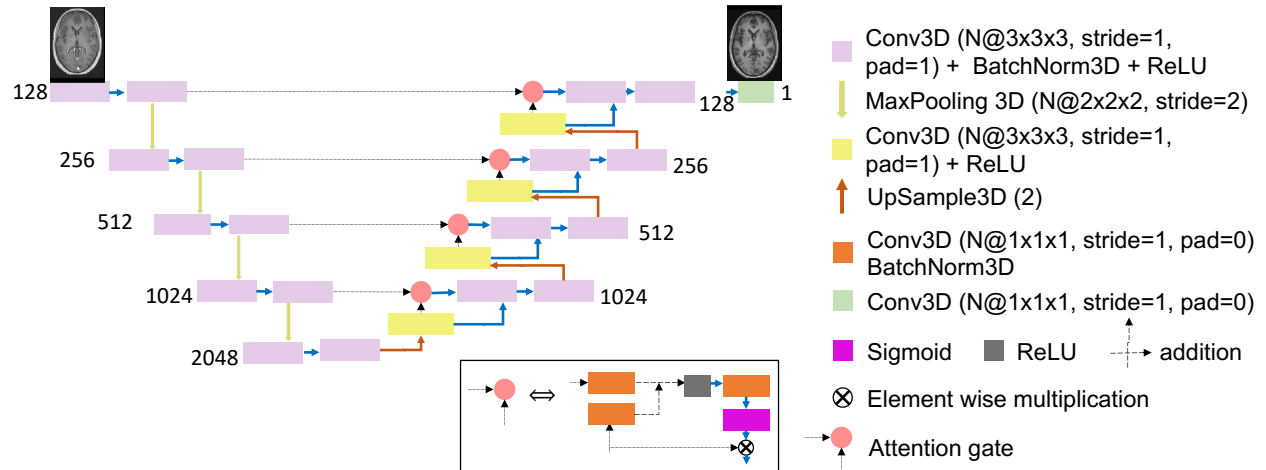
151 We implemented three models derived from the 3D U-Net:<sup>4</sup> a 3D U-Net with the addition of  
152 residual connections (called *Res-U-Net*), a 3D U-Net with the addition of attention mechanisms  
153 (called *Att-U-Net*), a 3D U-Net with both transformer and convolutional layers (called *Trans-U-*  
154 *net*). The U-Net structure allows preserving the details present in the original images thanks to the  
155 skip connections<sup>4</sup> and has shown good performance for image-to-image translation.<sup>5-12</sup> Here we  
156 detail the three architectures, which are also shown in Figure 1.

157 **Res-U-Net** The *Res-U-Net* we implemented is based on the architecture first proposed by Milletari  
158 et al.<sup>45</sup> and later used in Bône et al.<sup>34</sup> The five descending blocks are composed of 3D con-  
159 volutional layers followed by an instance normalization block and a LeakyReLU (negative  
160 slope coefficient  $\alpha = 0.2$ ). The four ascending blocks are composed of transposed con-  
161 volutional layers followed by a ReLU. The final layer is composed of an upsample module  
162 (factor of 2), a 3D convolutional block and a hyperbolic tangent module. Each descending or  
163 ascending block is followed by a residual module, which can vary from one to three blocks  
164 composed of a 3D convolutional layer and a LeakyReLU ( $\alpha = 0.2$ ). Residual blocks were  
165 introduced to avoid the problem of the vanishing gradients in the training of deep neural  
166 network:<sup>46</sup> they ease the training since they improve the flow of the information within the  
167 network.

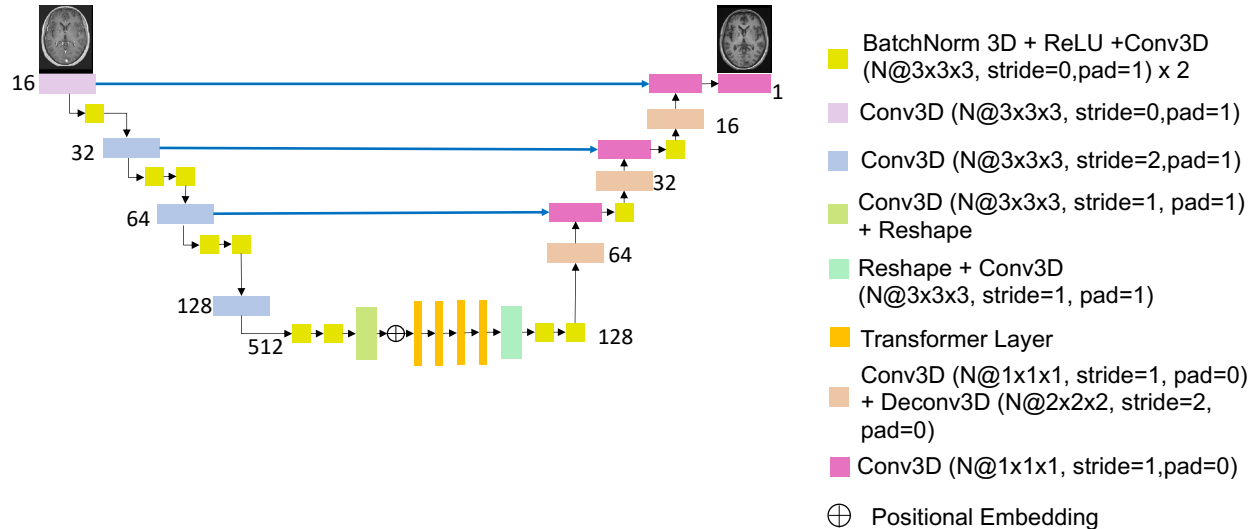




(a) *Res-U-Net*



(b) *Att-U-Net*



(c) *Trans-U-Net*

Fig 1: Architectures of the proposed 3D U-Net like models. The models take as input a real T1w-nce image of size  $128 \times 128 \times 128$  and generate a synthetic T1w-nce of size  $128 \times 128 \times 128$ . *Res-U-Net*: images pass through five descending blocks, each one followed by a residual module, and then through four ascending blocks and one final layer. *Att-U-Net*: images pass through five descending blocks and then through four ascending blocks and one final layer. One of the input of each ascending block is the result of the attention gate. *Trans-U-Net*: images pass through four descending blocks, four transformer layers and four ascending layers. All the parameters such as kernel size, stride, padding, size of each feature map (N) are reported.

168 **Att-U-Net** We implemented the *Att-U-Net* relying on the work of Oktay et al.<sup>47</sup> In this architec-  
169 ture, the five descending blocks are composed of two blocks with a 3D convolutional layer  
170 followed by a batch normalization layer and a ReLU. They are followed by four ascend-  
171 ing blocks. Each ascending block is composed of an upsample module (factor of 2), a 3D  
172 convolutional layer followed by a ReLU, an attention gate and two 3D convolutional lay-  
173 ers followed by a ReLU. The attention gate is composed of two 3D convolutional layers,  
174 a ReLU, a convolutional layer and a sigmoid layer. Its objective is to identify only salient  
175 image regions: the input of the attention gate is multiplied (element-wise multiplication) by  
176 a factor (in the range 0–1) resulting from the training of all the blocks of the networks. In  
177 this way it discards parts of the images that are not relevant to the task at hand.

178 **Trans-U-Net** The *Trans-U-Net* was implemented by Wang et al.<sup>48</sup> (who called the model *Trans-*  
179 *BTS*). They proposed a 3D U-Net like structure composed of both a CNN and a transformer.  
180 The CNN is used to produce an embedding of the input images in order not to lose local  
181 information across depth and space. The features extracted by the CNN are the input of the  
182 transformer whose aim is to model the global features. The descending blocks are composed  
183 of four different blocks, each being composed of a 3D convolutional layer and one, two or  
184 three blocks composed of a batch normalization layer, a ReLU and another 3D convolutional  
185 layer. The model is then composed of four transformer layers, after a linear projection of  
186 the features. Each transformer layer is itself composed of a multi-head attention block and a  
187 feed forward network. The four ascending blocks are composed of a 3D convolutional layer  
188 and one or two blocks with a batch normalization layer, a ReLU, a 3D convolutional layer  
189 followed by a 3D deconvolutional layer. The final layer is composed of a 3D convolutional

190 layer and a soft-max layer.

191 For the three 3D U-Net like models we used the same training parameters. We used the Adam  
192 optimizer, the L1 loss, a batch size of 2 and trained during 300 epochs. The model with the best  
193 loss, determined using the training set, was saved as final model. We relied on Pytorch for the  
194 implementation.

### 195 2.3.2 Conditional GANs

196 Generative adversarial networks (GANs) were firstly introduced by Goodfellow et al.<sup>49</sup> They are  
197 generative deep learning models composed of two elements: a generator for synthesizing new  
198 examples and a discriminator for classifying whether examples are real, i.e. the original ones, or  
199 fake, i.e. synthesized by the generator. Conditional GANs (cGANs)<sup>50</sup> are a variant of GANs where  
200 the generator and the discriminator are conditioned by the true samples. They can only be used  
201 with paired data sets.

202 We propose three different cGAN models that differ in the architecture of the generators, which  
203 correspond to the three architectures presented above. The discriminator is the same for all the  
204 cGANs: it is a 3D patch CNN, first proposed by Isola et al.<sup>51</sup> and used in the medical image  
205 translation domain.<sup>52,53</sup> Its aim is to classify if each pair of patches contains two real images or a  
206 real and a fake image. The advantages of working with patches is that the discriminator focuses on  
207 the details of the images and the generator must improve them to fool the discriminator.

208 Our discriminator is composed of four blocks: the first three blocks are composed of a 3D  
209 convolutional layer followed by a LeakyReLU (negative slope coefficient  $\alpha = 0.2$ ), and the last  
210 block is composed of a 3D convolutional layer and a 3D average pooling layer. From images of  
211 size  $128 \times 128 \times 128$ , we created eight patches of size  $64 \times 64 \times 64$  with a stride of 50.

212 For the training of the discriminator we used the least-square-loss as proposed in 54 in order to  
213 increase the stability, thus avoiding the problem of vanishing gradients that occurs with the usual  
214 cross-entropy loss. Stability of the training was also improved using soft labels: random numbers  
215 between 0 and 0.3 represented real images and random numbers between 0.7 and 1 represented  
216 fake images.

217 The total loss of the cGANs combines

- 218 • the loss of the generator composed of the sum of the L1 loss (i.e. pixel-wise absolute er-  
219 ror) computed between the generated and true images, and the least-square loss computed  
220 between the predicted probabilities of the generated images and positive labels.
- 221 • the loss of the discriminator composed of the mean of the least-square loss computed be-  
222 tween the predicted probabilities of the true images and positive labels and the least-square  
223 loss computed between the predicted probabilities of the generated images and negative la-  
224 bels.

225 At first, both the generators and discriminators were pretrained separately. Regarding each gen-  
226 erator, we reused the best model obtained previously. The discriminators were pretrained for the  
227 recognition of real and fake patches (fake images were obtained from each pretrained generator).  
228 The generators and discriminators were then trained together. The generator models with the best  
229 loss, determined using the training set, were saved as final models. Note that the batch size was set  
230 to 1 due to limited computing resources.

## 231 2.4 Experiments and validation measures

232 The experiments relied on 307 pairs of T1w-ce and T1w-nce images. We randomly selected 10%  
233 of the 256 image pairs of medium and good quality for testing (data set called  $\text{Test}_{\text{good}}$ ), the other  
234 230 image pairs being used for training. Only images of good and medium quality were used for  
235 training to ensure that the model focuses on the differences related to the presence or absence of  
236 gadolinium, and not to other factors. The remaining 51 image pairs with a T1w-ce of low quality  
237 and a T1w-nce of good or medium quality were used only for testing (data set called  $\text{Test}_{\text{low}}$ ).

### 238 2.4.1 Synthesis accuracy

239 Image similarity was evaluated using the mean absolute error (MAE), peak signal-to-noise ratio  
240 (PNSR) and structural similarity (SSIM).<sup>55</sup> The MAE is the mean of each absolute value of the  
241 difference between the true pixel and the generated pixel and PSNR is a function of the mean  
242 squared error: these two metrics allows a direct comparison between the synthetic image and the  
243 real one. The SSIM aims to measure quality by capturing the similarity of images, it is a weighted  
244 combination of the luminance, contrast and structure. For the MAE, the minimum value is 0 (the  
245 lower, the better), for PSNR the maximum value is infinite (the higher, the better) and for SSIM  
246 the maximum value is 1 (the higher, the better). We calculated these metrics both between the real  
247 and synthetic T1w-nce images and between the real T1w-nce and T1w-ce images (as reference).  
248 These metrics were calculated within the brain region. A brain mask was obtained for each subject  
249 by skull-stripping the T1w-nce and T1w-ce images using HD-BET<sup>56</sup> and computing the union of  
250 the two resulting brain masks.

## 251 2.4.2 Segmentation fidelity

252 Our goal is to obtain gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) segmen-  
253 tations from T1w-ce images using widely-used software tools that are consistent with segmenta-  
254 tions obtained from T1w-nce images. We thus assessed segmentation consistency by analyzing  
255 the tissue volumes resulting from the segmentations, which are important features when studying  
256 atrophy in the context of neurodegenerative diseases.

The volumes of the different tissues were obtained as follows. At first, synthetic T1w-nce im-  
ages were resampled back to a size of  $169 \times 208 \times 179$  using trilinear interpolation in Pytorch so  
that real and synthetic images have the same grid size. We processed the images using the ‘t1-  
volume-tissue-segmentation’ pipeline of Clinica.<sup>41,57</sup> This wrapper of the Unified Segmentation  
procedure implemented in SPM<sup>37</sup> simultaneously performs tissue segmentation, bias correction  
and spatial normalization. Once the probability maps were obtained for each tissue, we computed  
the maximum probability to generate binary masks and we multiplied the number of voxels by the  
voxel dimension to obtain the volume of each tissue. We calculated both the relative absolute dif-  
ference (rAD) and the relative difference (rD) for each tissue between the real T1w-ce or synthetic  
T1w-nce and the real T1w-nce as follows:

$$\text{rAD} = \frac{|V_t^I - V_t^J|}{TIV^I} \times TIV, \quad (1a)$$

$$\text{rD} = \frac{V_t^I - V_t^J}{TIV^I} \times TIV, \quad (1b)$$

257 where  $V_t^I$  is the volume of tissue  $t$  extracted from the real T1w-nce image  $I$ ,  $V_t^J$  is the volume  
258 of tissue  $t$  extracted from image  $J$ ,  $J$  being the synthetic T1w-nce or real T1w-ce image.  $TIV^I$   
259 corresponds to the total intracranial volume obtained from the real T1w-nce image  $I$  and  $TIV$

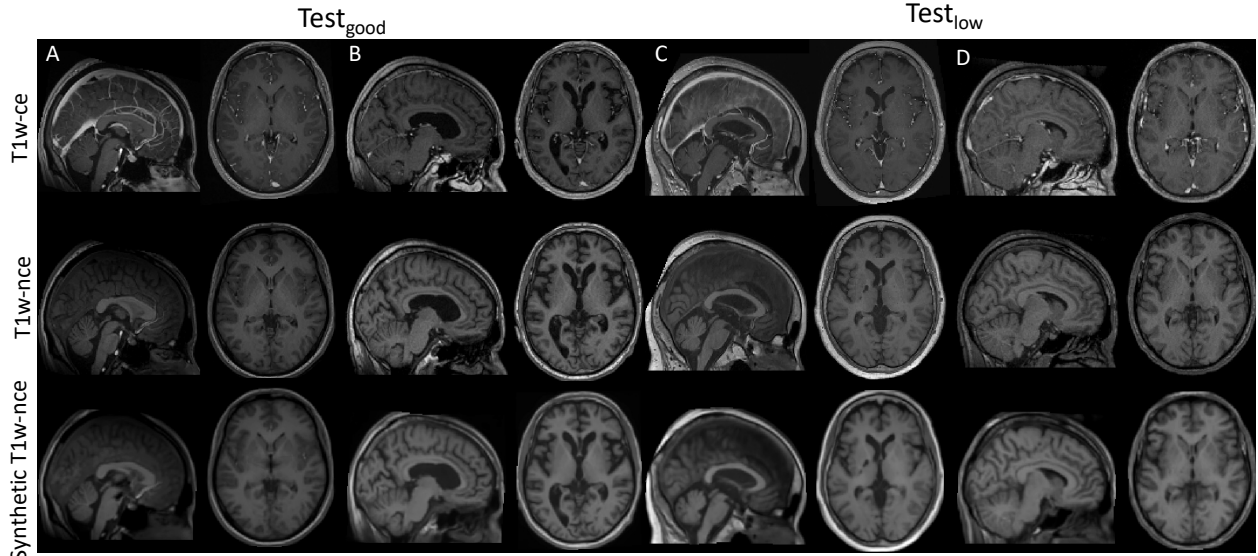


Fig 2: Examples of real T1w-ce (top), real T1w-nce (middle) and synthetic T1w-nce obtained with the *cGAN Att-U-Net* model (bottom) images in the sagittal and axial planes. Images of patients A and B belong to  $\text{Test}_{\text{good}}$  (left) while images of patients C and D belong to  $\text{Test}_{\text{low}}$  (right).

260 corresponds to the average total intracranial volume computed across the two test sets. The multi-  
 261 plication by the average total intracranial volume (TIV) aims at obtaining volumes (in  $\text{cm}^3$ ) rather  
 262 than fractions of the TIV of each subject, which is easier to interpret. Since this is a multiplication  
 263 by a constant, it has not impact on the results. To assess whether the tissue volumes presented  
 264 a statistically significant difference in terms of rAD depending on the images they were obtained  
 265 from, we performed paired t-tests using Bonferroni correction for multiple comparisons.

266 In addition, we compared the binary tissue maps extracted from the real T1w-ce or synthetic  
 267 T1w-nce image to those extracted from the real T1w-nce using the Dice score.

### 268 3 RESULTS

269 We report results for the proposed 3D U-Net like models and cGANs trained on 230 image pairs  
 270 of good and medium quality, and tested on  $\text{Test}_{\text{good}}$  and  $\text{Test}_{\text{low}}$  obtained from a clinical data set.

271 Examples of synthetic T1w-nce images obtained with the *cGAN Att-U-Net* model together  
 272 with the real T1w-ce and T1w-nce images are displayed in Figure 2. Images of patients A and

273 B belong to  $\text{Test}_{\text{good}}$  while images of patients C and D belong to  $\text{Test}_{\text{low}}$ . We note the absence of  
274 contrast agent in the synthetic T1w-nce, while it is clearly visible in the sagittal slice of the T1w-ce  
275 (particularly visible for patients A and C) and that the anatomical structures are preserved between  
276 the synthetic and real T1w-nce, even in the case of a disease (as for patient B). We also note that  
277 contrast between gray and white matter is preserved in the synthetic T1w-nce (particularly visible  
278 for patients B and D). For  $\text{Test}_{\text{low}}$ , the contrast seems improved in the synthetic compared with the  
279 real T1w-ce image (especially for patient D).

### 280 3.1 Synthesis accuracy

281 Table 1 reports the image similarity metrics obtained for the two test sets within the brain region.  
282 We computed these metrics to assess the similarity between real and synthetic T1w-nce images, but  
283 also between T1w-nce and T1w-ce images to set a baseline. We observe that, for all models, the  
284 similarity is higher between real and synthetic T1w-nce images than between T1w-nce and T1w-  
285 ce images according to all three metrics on both test sets. The differences observed in terms of  
286 MAE, PSNR and SSIM between the baseline and each image translation approach are statistically  
287 significant (corrected p-value  $< 0.05$  according to a paired t-test corrected for multiple comparisons  
288 using the Bonferroni correction).

289 Among the generators composed of 3D U-Net like models, the *Att-U-Net* performed slightly  
290 better than the others, both for  $\text{Test}_{\text{good}}$  (mean MAE: 2.73%, PSNR: 29.07 dB, SSIM: 0.96) and  
291  $\text{Test}_{\text{low}}$  (mean MAE: 2.89%, PSNR: 27.18 dB, SSIM: 0.95). The performance of the cGANs  
292 were comparable to their counterparts composed only of the generator. *cGAN Att-U-Net* had a  
293 lower MAE for both test sets (mean MAE: 2.69% for  $\text{Test}_{\text{good}}$  and mean MAE: 2.86% for  $\text{Test}_{\text{low}}$ ).  
294 There was no statistically significant difference observed, no matter the synthesis accuracy mea-



Table 1: MAE, PSNR and SSIM obtained on the two independent test sets with various image quality. For each metric, we report the average and standard deviation across the corresponding test set. We compute the metrics for both T1w-ce and synthetic T1w-nce in relation to the real T1w-nce, and so within the brain region.

Test set	Compared images	Model	MAE (%)	PSNR (dB)	SSIM
Test <sub>good</sub>	T1w-nce / T1w-ce	-	4.14 ± 1.59	23.03 ± 2.83	0.90 ± 0.05
	T1w-nce / Synthetic T1w-nce	<i>Res-U-Net</i>	3.06 ± 1.50	26.89 ± 4.30	0.95 ± 0.04
		<i>Att-U-Net</i>	2.73 ± 1.69	29.07 ± 4.53	0.96 ± 0.05
		<i>Trans-U-Net</i>	2.80 ± 1.42	28.00 ± 4.13	0.96 ± 0.04
		<i>cGAN Res-U-Net</i>	3.47 ± 1.59	23.89 ± 4.30	0.95 ± 0.04
		<i>cGAN Att-U-Net</i>	2.69 ± 1.68	28.89 ± 4.44	0.97 ± 0.05
		<i>cGAN Trans-U-Net</i>	2.86 ± 1.59	28.00 ± 4.32	0.96 ± 0.04
Test <sub>low</sub>	T1w-nce / T1w-ce	-	3.71 ± 1.99	24.20 ± 3.85	0.91 ± 0.06
	T1w-nce / Synthetic T1w-nce	<i>Res-U-Net</i>	2.93 ± 1.77	26.71 ± 4.32	0.95 ± 0.05
		<i>Att-U-Net</i>	2.89 ± 1.85	27.15 ± 4.57	0.95 ± 0.05
		<i>Trans-U-Net</i>	2.98 ± 1.89	26.71 ± 4.38	0.94 ± 0.05
		<i>cGAN Res-U-Net</i>	3.20 ± 1.96	26.20 ± 4.42	0.93 ± 0.05
		<i>cGAN Att-U-Net</i>	2.86 ± 1.83	27.12 ± 4.50	0.95 ± 0.05
		<i>cGAN Trans-U-Net</i>	2.97 ± 1.83	26.68 ± 4.40	0.94 ± 0.05

295 sure, between *cGAN Att-U-Net*, the best performing model according to the MAE, and the other  
296 approaches for both test sets (corrected p-value > 0.05). For further validation we kept only *Att-*  
297 *U-Net* and *cGAN Att-U-Net*.

### 298 3.2 Segmentation fidelity

299 Absolute volume differences (rAD) obtained between T1w-nce and T1w-ce images and between  
300 T1w-nce and synthetic T1w-nce images (obtained with the *Att-U-Net* model and the *cGAN Att-*  
301 *U-Net*) for GM, WM and CSF are reported in Table 2. For both test sets and all tissues, the  
302 absolute volume differences are smaller between T1w-nce and synthetic T1w-nce images than be-  
303 tween T1w-nce and T1w-ce images for the two models. Using the *Att-U-Net* on Test<sub>good</sub>, absolute  
304 volume differences of GM and CSF between T1w-nce/T1w-ce and T1w-nce/Synthetic T1w-nce are  
305 statistically significantly different (corrected p-value < 0.01 according to a paired t-test corrected  
306 for multiple comparisons using the Bonferroni correction), while on Test<sub>low</sub> absolute volume dif-

Table 2: Absolute volume difference (mean  $\pm$  standard deviation in  $\text{cm}^3$ ) between T1w-nce and T1w-ce images and between T1w-nce and synthetic T1w-nce images (obtained with the *Att-U-Net* and *cGAN Att-U-Net* models) for the gray matter, white matter and cerebrospinal fluid (CSF). \* indicates that the absolute volume difference between T1w-nce and synthetic T1w-nce images is statistically significantly different from that of the baseline (corrected p-value  $<0.01$ ) according to a paired t-test corrected for multiple comparisons using the Bonferroni correction.

	Compared images	Model	Test <sub>good</sub> [ $\text{cm}^3$ ]	Test <sub>low</sub> [ $\text{cm}^3$ ]
Gray matter	T1w-nce / T1w-ce	-	$26.68 \pm 15.92$	$49.63 \pm 49.38$
	T1w-nce / Synthetic T1w-ce	<i>Att-U-Net</i> <i>cGAN Att-U-Net</i>	$10.36 \pm 6.98$ * $9.24 \pm 6.10$ *	$19.61 \pm 29.54$ * $19.67 \pm 28.32$ *
White matter	T1w-nce / T1w-ce	-	$10.81 \pm 3.71$	$25.36 \pm 27.73$
	T1w-nce / Synthetic T1w-ce	<i>Att-U-Net</i> <i>cGAN Att-U-Net</i>	$7.79 \pm 5.87$ $6.40 \pm 4.43$ *	$13.95 \pm 24.74$ * $14.49 \pm 21.06$ *
CSF	T1w-nce / T1w-ce	-	$61.62 \pm 34.61$	$69.55 \pm 37.77$
	T1w-nce / Synthetic T1w-ce	<i>Att-U-Net</i> <i>cGAN Att-U-Net</i>	$13.37 \pm 10.18$ * $18.27 \pm 17.20$ *	$12.25 \pm 7.72$ * $17.10 \pm 18.45$ *

307 ferences of all the tissues are statistically significantly different (corrected p-value  $<0.01$ ). Using  
308 the *cGAN Att-U-Net* model, absolute volume differences of all the tissues are statistically signifi-  
309 cantly different (corrected p-value  $<0.01$ ) for both test sets. This means that there is an advantage  
310 in using synthetic T1w-nce images rather than T1w-ce images, no matter the model used for the  
311 synthesis: segmentation of GM, CSF and WM is more reliable since closer to the segmentation of  
312 the tissues in the real T1w-nce.

313 Volume differences (rD) computed between T1w-nce and T1w-ce images and between T1w-  
314 nce and synthetic T1w-nce images (obtained with the *Att-U-Net* and *cGAN Att-U-Net*) for GM,  
315 WM and CSF are reported in Figure 3. We observe that volumes extracted from T1w-ce images  
316 tend to be over-estimated (GM) or under-estimated (CSF) and that most of these biases disappear  
317 when tissues are extracted from synthetic T1w-nce images (mean rD closer to 0).

318 The Dice scores obtained when comparing the GM, WM and CSF segmentations between

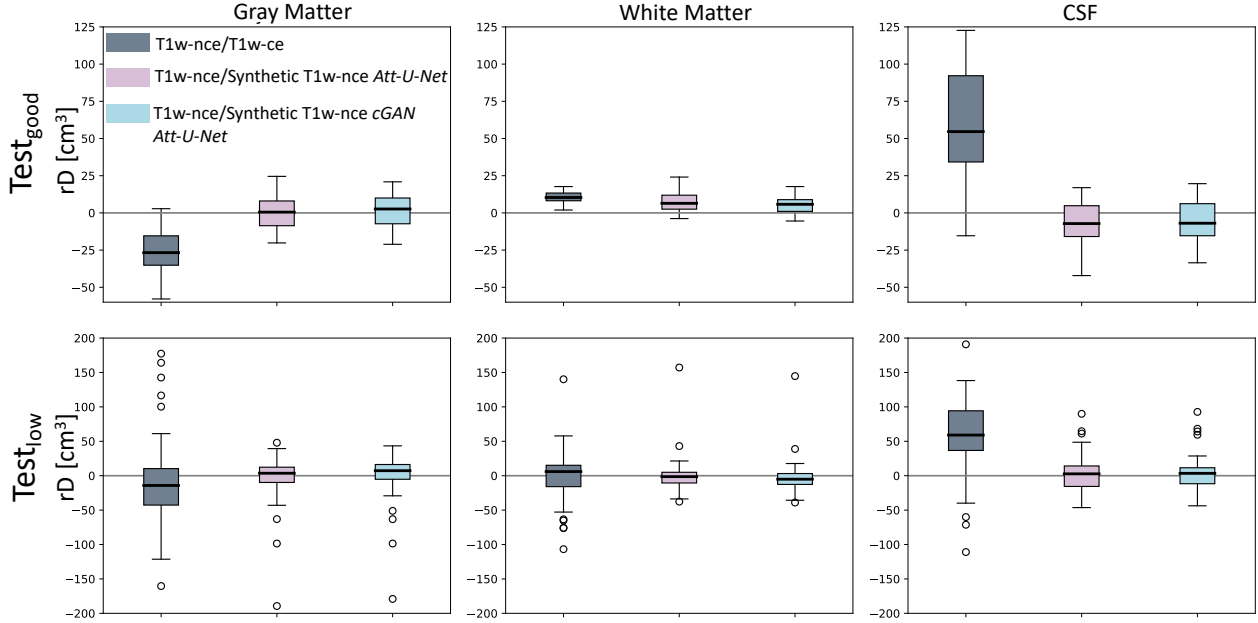


Fig 3: Volume differences (rD) in  $\text{cm}^3$  between T1w-nce and T1w-ce images and between T1w-nce and synthetic T1w-nce images (obtained with the *Att-U-Net* and the *cGAN Att-U-Net* models) for gray matter (left), white matter (middle) and cerebrospinal fluid (CSF, right) for both Test<sub>good</sub> (top) and Test<sub>low</sub> (bottom).

Table 3: Dice scores obtained when comparing the gray matter, white matter and cerebrospinal fluid (CSF) segmentations between T1w-nce and T1w-ce images and between T1w-nce and synthetic T1w-nce images (obtained with the *Att-U-Net* and the *cGAN Att-U-Net*)

	Compared images	Model	Test <sub>good</sub>	Test <sub>low</sub>
Gray matter	T1w-nce / T1w-ce	-	$0.88 \pm 0.02$	$0.77 \pm 0.12$
	T1w-nce / Synthetic T1w-ce	<i>Att-U-Net</i>	$0.87 \pm 0.02$	$0.81 \pm 0.07$
		<i>cGAN Att-U-Net</i>	$0.87 \pm 0.02$	$0.81 \pm 0.07$
White matter	T1w-nce / T1w-ce	-	$0.93 \pm 0.01$	$0.85 \pm 0.10$
	T1w-nce / Synthetic T1w-ce	<i>Att-U-Net</i>	$0.90 \pm 0.02$	$0.86 \pm 0.04$
		<i>cGAN Att-U-Net</i>	$0.91 \pm 0.02$	$0.86 \pm 0.03$
CSF	T1w-nce / T1w-ce	-	$0.63 \pm 0.10$	$0.62 \pm 0.10$
	T1w-nce / Synthetic T1w-ce	<i>Att-U-Net</i>	$0.80 \pm 0.05$	$0.78 \pm 0.07$
		<i>cGAN Att-U-Net</i>	$0.80 \pm 0.05$	$0.78 \pm 0.07$

319 T1w-nce and T1w-ce images and between T1w-nce and synthetic T1w-nce images (obtained with  
320 the *Att-U-Net* and the *cGAN Att-U-Net*) are displayed in Table 3. We observe that for both gray

321 and white matter, the Dice scores are similar between T1w-nce and T1w-ce or synthetic T1w-nce  
322 images, while for CSF higher Dice scores are obtained using synthetic T1w-nce images.

## 323 **4 DISCUSSION**

324 The use of clinical images for the validation of computer-aided diagnosis (CAD) systems is still  
325 largely unexplored. One of the obstacles lies in the heterogeneity of the data acquired in the  
326 context of routine clinical practice. Post-acquisition homogenization is crucial because, contrary  
327 to research data, no strict acquisition protocols, that would ensure a certain homogeneity among the  
328 images, exist for clinical data. Heterogeneity originates from the fact that images are acquired with  
329 different scanners at different field strengths during a large period of time and because patients may  
330 suffer from a large variety of diseases. Homogenization of clinical data sets of 3D T1w brain MRI,  
331 and consequently of the features extracted from them, is an important step for the development of  
332 reliable CAD systems. Indeed, when training a CAD system, the algorithms must not be affected  
333 by the data set variations even though clinical images may greatly vary.

334 A source of heterogeneity among clinical data sets is the fact that they contain a mix of images  
335 acquired with and without gadolinium-based contrast agent. In our case, among the 7397 proper  
336 T1w brain images made available by the AP-HP data warehouse out of a batch of 9941 images,  
337 59% of the images were contrast-enhanced.<sup>39</sup> To homogenize this data set, we thus proposed a  
338 framework to convert T1w-ce images into T1w-nce images using deep learning models. The choice  
339 to synthesize T1w-nce images from T1w-nce images was constrained by the fact that software tools  
340 for feature extraction in the neuroimaging community were developed for T1w-nce MRI. To the  
341 best of our knowledge, none of these tools has largely been applied to the extraction of features  
342 from T1w-ce MRI data and their performance in this scenario is thus mostly unknown.

343 The contribution of our work consists in the development and validation of deep learning mod-  
344 els (U-Net models and conditional GANs) for the translation of T1w-ce to T1w-nce images coming  
345 from a clinical data warehouse. We compared three 3D U-net models differentiated by the addi-  
346 tion of residual modules, of attention modules or of transformer layers, used as simple generators  
347 and also within a conditional GAN setting with the addition of a patch-based discriminator. These  
348 models have widely been used for the image translation of medical images,<sup>58,59</sup> but their applica-  
349 tion to clinical data has not been proven yet. The proposed models were trained using 230 image  
350 pairs and tested on two different test sets: 26 image pairs had both a T1w-nce and T1w-ce of good  
351 or medium quality and 51 image pairs had a T1w-nce of good or medium quality and a T1w-ce  
352 of bad quality. Having two test sets of different qualities is a key point since we are dealing with  
353 a real clinical heterogeneous data set where images of low quality, corresponding in majority to  
354 T1w-ce images with a low contrast, may represent 30% of the data.<sup>39</sup>

355 We first assessed the similarity between real and synthetic T1w-nce images and between real  
356 T1w-nce and T1w-ce images using three similarity metrics, MAE, PSNR and SSIM. We showed  
357 that the similarity between real and synthetic T1w-nce images was higher than the similarity be-  
358 tween real T1w-nce and T1w-ce images according to all the metrics, no matter the models used nor  
359 the quality of the input image. The synthesis accuracy obtained with the models evaluated was of  
360 the same order as the one reached in recent works on non-contrast-enhanced to contrast-enhanced  
361 image translation.<sup>34,35</sup> The performance of all the models was equivalent (no statistically signifi-  
362 cant difference observed), meaning that all were able to synthesize T1w-nce images. Slightly better  
363 performance was reached with the addition of attention modules (*Att-U-Net* and *cGAN Att-U-Net*  
364 models), these models were thus further evaluated.

365 In the second step of the validation, we assessed the similarity of features extracted from the

366 different images available using a widely adopted segmentation framework, SPM.<sup>1</sup> We showed  
367 that the absolute volume differences of GM, WM and CSF were larger between real T1w-nce and  
368 Tw-ce images than between real and synthetic T1w-nce images (statistically significant difference  
369 most of the times). This confirms the hypothesis that gadolinium-based contrast agent may alter  
370 the contrast between the different brain tissues, making features extracted from such images with  
371 standard segmentation tools, here SPM,<sup>1</sup> unreliable. At the same time, we validated the suitability  
372 of the synthetic images since their segmentation was consistent with those obtained from real T1w-  
373 nce images as the volume differences were small. In particular we see that for both test sets, volume  
374 differences are statistically significantly different (corrected p-value<0.01 according to a paired t-  
375 test corrected for multiple comparisons using the Bonferroni correction) for GM which is the main  
376 feature when studying atrophy in neurodegenerative diseases. The fact that the relative differences  
377 between the volumes extracted from the real and synthetic T1w-nce images are relatively close to  
378 zero show that the tissue volumes are not systematically under- or over-estimated when extracted  
379 from the synthetic images.

380 Even though the synthetic T1w-nce images enable the extraction of reliable features, their qual-  
381 ity could still be improved. Many constraints exist when working with data from a clinical data  
382 warehouse. One is the fact that these data are accessible only through a closed environment pro-  
383 vided by the IT department of the AP-HP as described in.<sup>60</sup> Limitations in computational resources  
384 and storage space make training deep learning models difficult and thus limits the experiments that  
385 can be performed to find the optimal model. The proposed models could be improved by bet-  
386 ter optimizing the hyperparameters (such as the learning rate or the size of the kernels), adding  
387 a perceptual loss when training the conditional GANs<sup>61</sup> or adding more layers in the patch-based  
388 discriminator. Other architectures could also be explored. We have restricted our work to condi-

389 tional GANs, which need paired data to be trained, but we could exploit more data working with  
390 cycle GANs<sup>62</sup> as they can deal with unpaired data.

391 Several steps remain to be performed before using synthetic T1w-nce images for the differential  
392 diagnosis of neurological diseases. First, the performance of CAD systems trained with a mix of  
393 real T1w-nce and T1w-ce images should be compared with the performance of CAD systems  
394 trained with a mix of real and synthetic T1w-nce images. To prevent introducing a correlation  
395 between image properties (e.g. smoothness) and pathology, which would bias the classification  
396 performance, it may be necessary to also feed the real T1w-nce images to the neural network and  
397 use the resulting images as inputs of the CAD system, as suggested in 31.

## 398 **5 CONCLUSION**

399 Clinical data warehouses offer fantastic opportunities for computer-aided diagnosis of neurolog-  
400 ical diseases but their heterogeneity must be reduced to avoid biases. In this work we proposed  
401 to homogenize such a large clinical data set by converting images acquired after the injection of  
402 gadolinium into non-contrast-enhanced images using 3D U-Net models and conditional GANs.  
403 Validation using standard image similarity measures demonstrated that the similarity between real  
404 and synthetic T1w-nce images was higher than between real T1w-nce and T1w-ce images for all  
405 the models compared. We also showed that features extracted from the synthetic images (GM,  
406 WM, CSF volumes) were closer to those obtained from the T1w-nce brain MR images (consid-  
407 ered as reference) than the original T1w-ce images. These results demonstrate the ability of deep  
408 learning methods to homogenize a data set coming from a clinical data warehouse.

409 **DISCLOSURES**

410 Competing financial interests related to the present article: none to disclose for all authors.

411 Competing financial interests unrelated to the present article: OC reports having received consult-  
412 ing fees from AskBio (2020), having received fees for writing a lay audience short paper from  
413 Expression Santé (2019). Members from his laboratory have co-supervised a PhD thesis with my-  
414 BrainTechnologies (2016-2019) and with Qynapse (2017-present). OC's spouse is an employee  
415 and holds stock-options of myBrainTechnologies (2015-present). O.C. holds a patent registered at  
416 the International Bureau of the World Intellectual Property Organization (PCT/IB2016/0526993,  
417 Schiratti J-B, Allasonniere S, Colliot O, Durrleman S, A method for determining the temporal  
418 progression of a biological phenomenon and associated methods and devices) (2017).

419 **ACKNOWLEDGMENTS**

420 The research was done using the Clinical Data Warehouse of the Greater Paris University Hos-  
421 pitals. The authors are grateful to the members of the AP-HP WIND and URC teams, and in  
422 particular Stéphane Bréant, Florence Tubach, Jacques Ropers, Antoine Rozès, Camille Nevoret,  
423 Christel Daniel, Martin Hilka, Yannick Jacob, Cyrina Saussol, Julien Dubiel, Faerber Philippe,  
424 Mmaka Ibrahim and Gozlan Rafel. They would also like to thank the “Collégiale de Radiologie of  
425 AP-HP” as well as, more generally, all the radiology departments from AP-HP hospitals.

426 The research leading to these results has received funding from the Abeona Foundation (project  
427 Brain@Scale), from the French government under management of Agence Nationale de la Recherche  
428 as part of the “Investissements d’avenir” program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA  
429 Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut  
430 Hospitalo-Universitaire-6).



431 **APPRIMAGE Study Group**

432 Olivier Colliot, Ninon Burgos, Simona Bottani <sup>1</sup>

433 Didier Dormont <sup>1,2</sup>, Samia Si Smail Belkacem, Sebastian Ströer <sup>2</sup>

434 Nathalie Boddaert <sup>3</sup>

435 Farida Benoudiba, Ghaida Nasser, Claire Ancelet, Laurent Spelle <sup>4</sup>

436 Hubert Ducou-Le-Pointe<sup>5</sup>

437 Catherine Adamsbaum<sup>6</sup>

438 Marianne Alison<sup>7</sup>

439 Emmanuel Houdart<sup>8</sup>

440 Robert Carlier <sup>9,17</sup>

441 Myriam Edjlali<sup>9</sup>

442 Betty Marro<sup>10,11</sup>

443 Lionel Arrive<sup>10</sup>

444 Alain Luciani<sup>12</sup>

445 Antoine Khalil<sup>13</sup>

446 Elisabeth Dion<sup>14</sup>

447 Laurence Rocher<sup>15</sup>

448 Pierre-Yves Brillet<sup>16</sup>

449 Paul Legmann, Jean-Luc Drape <sup>18</sup>

450 Aurélien Maire, Stéphane Bréant, Christel Daniel, Martin Hilka, Yannick Jacob, Julien Dubiel,

451 Cyrina Saussol <sup>19</sup>

452 Florence Tubach, Jacques Ropers, Antoine Rozès, Camille Nevoret <sup>20</sup>

453

454 <sup>1</sup> Sorbonne Université, Institut du Cerveau - Paris Brain Institute, Inserm, CNRS, AP-HP, Hôpital de la Pitié

455 Salpêtrière, Inria, Aramis project-team, F-75013, Paris, France

456 <sup>2</sup> AP-HP, Hôpital de la Pitié Salpêtrière, Department of Neuroradiology, F-75013, Paris, France

457 <sup>3</sup> AP-HP, Hôpital Necker, Department of Radiology, F-75015, Paris, France

458 <sup>4</sup> AP-HP, Hôpital Bicêtre, Department of Radiology, F-94270, Le Kremlin-Bicêtre, France

459 <sup>5</sup> AP-HP, Hôpital Armand-Trousseau, Department of Radiology, F-75012, Paris, France

460 <sup>6</sup> AP-HP, Hôpital Bicêtre, Department of Pediatric Radiology, F-94270, Le Kremlin-Bicêtre, France

461 <sup>7</sup> AP-HP, Hôpital Robert-Debré, Department of Radiology, F-75019, Paris, France

462 <sup>8</sup> AP-HP, Hôpital Lariboisière, Department of Neuroradiology, F-75010, Paris, France

463 <sup>9</sup> AP-HP, Hôpital Raymond-Poincaré, Department of Radiology, F-92380, Garches, France

464 <sup>10</sup> AP-HP, Hôpital Saint-Antoine, Department of Radiology, F-75012, Paris, France

465 <sup>11</sup> AP-HP, Hôpital Tenon, Department of Radiology, F-75020, Paris, France

466 <sup>12</sup> AP-HP, Hôpital Henri-Mondor, Department of Radiology, F-94000, Créteil, France

467 <sup>13</sup> AP-HP, Hôpital Bichat, Department of Radiology, F-75018, Paris, France

468 <sup>14</sup> AP-HP, Hôpital Hôtel-Dieu, Department of Radiology, F-75004, Paris, France

469 <sup>15</sup> AP-HP, Hôpital Antoine-Béclère, Department of Radiology, F-92140, Clamart, France

470 <sup>16</sup> AP-HP, Hôpital Avicenne, Department of Radiology, F-93000, Bobigny, France

471 <sup>17</sup> AP-HP, Hôpital Ambroise Paré, Department of Radiology, F-92100 104, Boulogne-Billancourt, France

472 <sup>18</sup> AP-HP, Hôpital Cochin, Department of Radiology, F-75014, Paris, France

473 <sup>19</sup> AP-HP, WIND department, F-75012, Paris, France

474 <sup>20</sup> AP-HP, Unité de Recherche Clinique, Hôpital de la Pitié Salpêtrière, Department of Neuroradiology, F-

475 75013, Paris, France

477 *References*

- 478 1 W. D. Penny, K. J. Friston, J. T. Ashburner, *et al.*, *Statistical parametric mapping: the analy-*  
479 *sis of functional brain images*, Elsevier (2011).
- 480 2 B. B. Avants, N. J. Tustison, M. Stauffer, *et al.*, “The Insight ToolKit image registration  
481 framework,” *Frontiers in Neuroinformatics* **8**, 44 (2014).
- 482 3 J. Mark, F. B. Christian, E. B. Timothy, *et al.*, “FSL,” *NeuroImage* **62**(2), 782–790 (2012).
- 483 4 O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical im-  
484 age segmentation,” in *International Conference on Medical image computing and computer-*  
485 *assisted intervention*, 234–241, Springer (2015).
- 486 5 X. Han, “MR-based synthetic CT generation using a deep convolutional neural network  
487 method,” *Medical Physics* **44**(4), 1408–1419 (2017).
- 488 6 I. Shiri, P. Ghafarian, P. Geramifar, *et al.*, “Direct attenuation correction of brain PET images  
489 using only emission data via a deep convolutional encoder-decoder (Deep-DAC),” *European*  
490 *radiology* **29**(12), 6867–6879 (2019).
- 491 7 K. Gong, J. Yang, K. Kim, *et al.*, “Attenuation correction for brain PET imaging using deep  
492 neural network based on Dixon and ZTE MR images,” *Physics in Medicine & Biology* **63**(12),  
493 125011 (2018).
- 494 8 C. N. Ladefoged, L. Marner, A. Hindsholm, *et al.*, “Deep learning based attenuation correc-  
495 tion of PET/MRI in pediatric brain tumor patients: evaluation in a clinical setting,” *Frontiers*  
496 *in neuroscience* **12**, 1005 (2019).

- 497 9 K. D. Spuhler, J. Gardus, Y. Gao, *et al.*, “Synthesis of patient-specific transmission data  
498 for PET attenuation correction for PET/MRI neuroimaging using a convolutional neural net-  
499 work,” *Journal of nuclear medicine* **60**(4), 555–560 (2019).
- 500 10 J. Yang, D. Park, G. T. Gullberg, *et al.*, “Joint correction of attenuation and scatter in image  
501 space using deep convolutional neural networks for dedicated brain 18F-FDG PET,” *Physics*  
502 *in medicine & biology* **64**(7), 075019 (2019).
- 503 11 S. Nepl, G. Landry, C. Kurz, *et al.*, “Evaluation of proton and photon dose distributions  
504 recalculated on 2D and 3D Unet-generated pseudoCTs from T1-weighted MR head scans,”  
505 *Acta Oncologica* **58**(10), 1429–1434 (2019).
- 506 12 J. M. Wolterink, A. M. Dinkla, M. H. Savenije, *et al.*, “Deep MR to CT synthesis using  
507 unpaired data,” in *International workshop on simulation and synthesis in medical imaging*,  
508 14–23, Springer (2017).
- 509 13 Y. Chen, F. Shi, A. G. Christodoulou, *et al.*, “Efficient and accurate MRI super-resolution  
510 using a generative adversarial network and 3D multi-level densely connected network,” in  
511 *International Conference on Medical Image Computing and Computer-Assisted Intervention*,  
512 91–99, Springer (2018).
- 513 14 J. Gu, Z. Li, Y. Wang, *et al.*, “Deep generative adversarial networks for thin-section infant  
514 MR image reconstruction,” *IEEE Access* **7**, 68290–68304 (2019).
- 515 15 K. H. Kim, W.-J. Do, and S.-H. Park, “Improving resolution of MR images with an adversar-  
516 ial network incorporating images with different contrast,” *Medical physics* **45**(7), 3120–3131  
517 (2018).
- 518 16 A. M. Dinkla, J. M. Wolterink, M. Maspero, *et al.*, “MR-only brain radiation therapy: dosi-

- 519 metric evaluation of synthetic CTs generated by a dilated convolutional neural network,”  
520 *International Journal of Radiation Oncology\* Biology\* Physics* **102**(4), 801–812 (2018).
- 521 17 H. Emami, M. Dong, S. P. Nejad-Davarani, *et al.*, “Generating synthetic CTs from magnetic  
522 resonance images using generative adversarial networks,” *Medical physics* **45**(8), 3627–3636  
523 (2018).
- 524 18 D. Nie, R. Trullo, J. Lian, *et al.*, “Medical image synthesis with deep convolutional adversar-  
525 ial networks,” *IEEE Transactions on Biomedical Engineering* **65**(12), 2720–2730 (2018).
- 526 19 S. U. Dar, M. Yurt, L. Karacan, *et al.*, “Image synthesis in multi-contrast MRI with con-  
527 ditional generative adversarial networks,” *IEEE Transactions on Medical Imaging* **38**(10),  
528 2375–2388 (2019).
- 529 20 B. Yu, L. Zhou, L. Wang, *et al.*, “Ea-GANs: edge-aware generative adversarial networks for  
530 cross-modality MR image synthesis,” *IEEE transactions on medical imaging* **38**(7), 1750–  
531 1762 (2019).
- 532 21 H. Li, J. C. Paetzold, A. Sekuboyina, *et al.*, “DiamondGAN: unified multi-modal generative  
533 adversarial networks for MRI sequences synthesis,” in *International Conference on Medical  
534 Image Computing and Computer-Assisted Intervention*, 795–803, Springer (2019).
- 535 22 A. Sharma and G. Hamarneh, “Missing MRI pulse sequence synthesis using multi-modal  
536 generative adversarial network,” *IEEE transactions on medical imaging* **39**(4), 1170–1183  
537 (2019).
- 538 23 A. Benou, R. Veksler, A. Friedman, *et al.*, “Ensemble of expert deep neural networks for  
539 spatio-temporal denoising of contrast-enhanced MRI sequences,” *Medical Image Analysis*  
540 **42**, 145–159 (2017).

- 541 24 D. Jiang, W. Dou, L. Vosters, *et al.*, “Denoising of 3D magnetic resonance images with multi-  
542 channel residual learning of convolutional neural network,” *Japanese journal of radiology*  
543 **36**(9), 566–574 (2018).
- 544 25 M. Ran, J. Hu, Y. Chen, *et al.*, “Denoising of 3D magnetic resonance images using a residual  
545 encoder–decoder Wasserstein generative adversarial network,” *Medical image analysis* **55**,  
546 165–180 (2019).
- 547 26 F. Hashimoto, H. Ohba, K. Ote, *et al.*, “Dynamic PET image denoising using deep convolu-  
548 tional neural networks without prior training datasets,” *IEEE Access* **7**, 96594–96603 (2019).
- 549 27 J. Du, L. Wang, Y. Liu, *et al.*, “Brain mri super-resolution using 3d dilated convolutional  
550 encoder–decoder network,” *IEEE Access* **8**, 18938–18950 (2020).
- 551 28 C.-H. Pham, A. Ducournau, R. Fablet, *et al.*, “Brain MRI super-resolution using deep 3D  
552 convolutional networks,” in *2017 IEEE ISBI*, 197–200 (2017).
- 553 29 K. Zeng, H. Zheng, C. Cai, *et al.*, “Simultaneous single-and multi-contrast super-resolution  
554 for brain MRI images based on a convolutional neural network,” *Computers in Biology and*  
555 *Medicine* **99**, 133–141 (2018).
- 556 30 X. Han, “MR-based synthetic CT generation using a deep convolutional neural network  
557 method,” *Medical physics* **44**(4), 1408–1419 (2017).
- 558 31 B. E. Dewey, C. Zhao, J. C. Reinhold, *et al.*, “DeepHarmony: a deep learning approach to  
559 contrast harmonization across scanner changes,” *Magnetic Resonance Imaging* **64**, 160–170  
560 (2019).
- 561 32 C. Xu, D. Zhang, J. Chong, *et al.*, “Synthesis of gadolinium-enhanced liver tumors on nonen-

- 562 hanced liver mr images using pixel-level graph reinforcement learning,” *Medical Image Anal-*  
563 *ysis* **69**, 101976 (2021).
- 564 33 M. Seo, D. Kim, K. Lee, *et al.*, “Neural contrast enhancement of ct image,” in *Proceedings of*  
565 *the IEEE/CVF Winter Conference on Applications of Computer Vision*, 3973–3982 (2021).
- 566 34 A. Bône, S. Ammari, J.-P. Lamarque, *et al.*, “Contrast-enhanced brain MRI synthesis with  
567 deep learning: key input modalities and asymptotic performance,” in *2021 IEEE ISBI*, (2021).
- 568 35 J. Kleesiek, J. N. Morshuis, F. Isensee, *et al.*, “Can virtual contrast enhancement in brain MRI  
569 replace gadolinium?: a feasibility study,” *Investigative Radiology* **54**(10), 653–660 (2019).
- 570 36 H. Sun, X. Liu, X. Feng, *et al.*, “Substituting Gadolinium in Brain MRI Using DeepContrast,”  
571 in *2020 IEEE ISBI*, 908–912 (2020).
- 572 37 J. Ashburner and K. J. Friston, “Unified segmentation,” *NeuroImage* **26**(3), 839–851 (2005).
- 573 38 S. Bottani, E. Thibeau-Sutre, A. Maire, *et al.*, “Homogenization of brain MRI from a clinical  
574 data warehouse using contrast-enhanced to non-contrast-enhanced image translation with U-  
575 Net derived models,” in *SPIE Medical Imaging 2022*, (2022).
- 576 39 S. Bottani, N. Burgos, A. Maire, *et al.*, “Automatic quality control of brain T1-weighted mag-  
577 netic resonance images for a clinical data warehouse,” *Medical Image Analysis* **75**, 102219  
578 (2022).
- 579 40 K. J. Gorgolewski, T. Auer, V. D. Calhoun, *et al.*, “The brain imaging data structure, a format  
580 for organizing and describing outputs of neuroimaging experiments,” *Scientific data* **3**(1), 1–9  
581 (2016).
- 582 41 A. Routier, N. Burgos, M. Díaz, *et al.*, “Clinica: An Open-Source Software Platform for  
583 Reproducible Clinical Neuroscience Studies,” *Frontiers in Neuroinformatics* **15**, 39 (2021).

- 584 42 N. J. Tustison, B. B. Avants, P. A. Cook, *et al.*, “N4ITK: improved N3 bias correction,” *IEEE*  
585 *Transactions on Medical Imaging* **29**(6), 1310–1320 (2010).
- 586 43 B. B. Avants, C. L. Epstein, M. Grossman, *et al.*, “Symmetric diffeomorphic image registra-  
587 tion with cross-correlation: evaluating automated labeling of elderly and neurodegenerative  
588 brain,” *Medical Image Analysis* **12**(1), 26–41 (2008).
- 589 44 J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, *et al.*, “Convolutional Neural Networks for Clas-  
590 sification of Alzheimer’s Disease: Overview and Reproducible Evaluation,” *Medical Image*  
591 *Analysis* **63**, 101694 (2020).
- 592 45 F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for  
593 volumetric medical image segmentation,” in *2016 fourth international conference on 3D vi-*  
594 *sion (3DV)*, 565–571, IEEE (2016).
- 595 46 K. He, X. Zhang, S. Ren, *et al.*, “Identity mappings in deep residual networks,” in *European*  
596 *conference on computer vision*, 630–645, Springer (2016).
- 597 47 O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, “Attention U-Net: Learning Where to Look for  
598 the Pancreas,” in *Medical Imaging with Deep Learning - MIDL 2018*, (2018).
- 599 48 W. Wang, C. Chen, M. Ding, *et al.*, “Transbts: Multimodal brain tumor segmentation us-  
600 ing transformer,” in *International Conference on Medical Image Computing and Computer-*  
601 *Assisted Intervention*, 109–119, Springer (2021).
- 602 49 I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, “Generative adversarial nets,” *Advances in*  
603 *neural information processing systems* **27** (2014).
- 604 50 M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv:1411.1784*  
605 (2014).



- 606 51 P. Isola, J.-Y. Zhu, T. Zhou, *et al.*, “Image-to-image translation with conditional adversarial  
607 networks,” in *Proceedings of the IEEE conference on computer vision and pattern recogni-  
608 tion*, 1125–1134 (2017).
- 609 52 W. Wei, E. Poirion, B. Bodini, *et al.*, “Predicting PET-derived demyelination from multi-  
610 modal MRI using sketcher-refiner adversarial training for multiple sclerosis,” *Medical image  
611 analysis* **58**, 101546 (2019).
- 612 53 H. Choi and D. S. Lee, “Generation of structural MR images from amyloid PET: application  
613 to MR-less quantification,” *Journal of Nuclear Medicine* **59**(7), 1111–1117 (2018).
- 614 54 X. Mao, Q. Li, H. Xie, *et al.*, “Least squares generative adversarial networks,” in *Proceedings  
615 of the IEEE international conference on computer vision*, 2794–2802 (2017).
- 616 55 Z. Wang, A. C. Bovik, H. R. Sheikh, *et al.*, “Image quality assessment: from error visibility  
617 to structural similarity,” *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).
- 618 56 F. Isensee, M. Schell, I. Pflueger, *et al.*, “Automated brain extraction of multisequence MRI  
619 using artificial neural networks,” *Human Brain Mapping* **40**(17), 4952–4964 (2019).
- 620 57 J. Samper-González, N. Burgos, S. Bottani, *et al.*, “Reproducible evaluation of classifica-  
621 tion methods in Alzheimer’s disease: Framework and application to MRI and PET data,”  
622 *NeuroImage* **183**, 504–521 (2018).
- 623 58 X. Yi, E. Walia, and P. Babyn, “Generative adversarial network in medical imaging: A re-  
624 view,” *Medical image analysis* **58**, 101552 (2019).
- 625 59 N. Burgos, S. Bottani, J. Faouzi, *et al.*, “Deep learning for brain disorders: from data pro-  
626 cessing to disease treatment,” *Briefings in Bioinformatics* **22**(2), 1560–1576 (2021).

- 627 60 C. Daniel and E. Salamanca, “Hospital Databases,” in *Healthcare and Artificial Intelligence*,  
628 B. Nordlinger, C. Villani, and D. Rus, Eds., 57–67, Springer (2020).
- 629 61 H. Zhao, O. Gallo, I. Frosio, *et al.*, “Loss functions for image restoration with neural net-  
630 works,” *IEEE Transactions on computational imaging* **3**(1), 47–57 (2016).
- 631 62 J.-Y. Zhu, T. Park, P. Isola, *et al.*, “Unpaired image-to-image translation using cycle-  
632 consistent adversarial networks,” in *Proceedings of the IEEE international conference on*  
633 *computer vision*, 2223–2232 (2017).

634 **Simona Bottani** is a PhD student in the ARAMIS Lab, a joint laboratory between CNRS, Inria,  
635 Inserm and Sorbonne University within the Paris Brain Institute. She received a master degree in  
636 Biomedical Engineering at Politecnico di Torino in 2016. Her research focuses on deep learning,  
637 clinical data warehouse and neurodegenerative diseases.

638 **Elina Thibeau-Sutre** was a PhD student in the ARAMIS Lab, a joint laboratory between CNRS,  
639 Inria, Inserm and Sorbonne University within the Paris Brain Institute. Her research interest in-  
640 cludes deep learning application to neuroimaging data, its interpretability and reproducibility. She  
641 received master degrees from Ecole Nationale des Mines de Paris and Ecole supérieure de physique  
642 et de chimie industrielles (Paris, France).

643 **Aurélien Maire** is the head of the imaging group of the Assistance Publique-Hôpitaux de Paris  
644 (AP-HP) Megadata Platform, which exploits the AP-HP data warehouse.

645 **Sebastian Ströer** is a neuroradiologist working in the neuroradiology department of the Pitié  
646 Salpêtrière hospital (Assistance Publique-Hôpitaux de Paris).

647 **Didier Dormont** is Professor of Neuroradiology at the Pitié Salpêtrière hospital (Assistance Publique-  
648 Hôpitaux de Paris) and affiliated to Sorbonne Université. He is the Medical Director of the DIA-  
649 MENT Department (Radiology, Nuclear Medicine and Pathologic Anatomy), AP-HP/Sorbonne  
650 Université.

651 **Olivier Colliot** is a Research Director at CNRS and the co-head of the ARAMIS Lab, a joint  
652 laboratory between CNRS, Inria, Inserm and Sorbonne University within the Paris Brain Institute  
653 (ICM), France. He received his PhD in Signal Processing from Telecom ParisTech in 2003 and  
654 the Habilitation degree from University Paris-Sud in 2011. His research interests include machine

655 learning, medical image analysis and their applications to neurological disorders.

656 **Ninon Burgos** is a CNRS researcher in the ARAMIS Lab, a joint laboratory between CNRS,  
657 Inria, Inserm and Sorbonne University within the Paris Brain Institute, France. She completed her  
658 PhD at University College London, UK, in 2016. Her research focuses on the development of  
659 computational imaging tools to improve the understanding and diagnosis of dementia.

## 660 List of Figures

- 661 1 Architectures of the proposed 3D U-Net like models. The models take as input a  
662 real T1w-nce image of size  $128 \times 128 \times 128$  and generate a synthetic T1w-nce of  
663 size  $128 \times 128 \times 128$ . *Res-U-Net*: images pass through five descending blocks, each  
664 one followed by a residual module, and then through four ascending blocks and  
665 one final layer. *Att-U-Net*: images pass through five descending blocks and then  
666 through four ascending blocks and one final layer. One of the input of each as-  
667 cending block is the result of the attention gate. *Trans-U-Net*: images pass through  
668 four descending blocks, four transformer layers and four ascending layers. All the  
669 parameters such as kernel size, stride, padding, size of each feature map (N) are  
670 reported.
- 671 2 Examples of real T1w-ce (top), real T1w-nce (middle) and synthetic T1w-nce ob-  
672 tained with the *cGAN Att-U-Net* model (bottom) images in the sagittal and axial  
673 planes. Images of patients A and B belong to  $\text{Test}_{\text{good}}$  (left) while images of pa-  
674 tients C and D belong to  $\text{Test}_{\text{low}}$  (right).
- 675 3 Volume differences ( $\text{rD}$ ) in  $\text{cm}^3$  between T1w-nce and T1w-ce images and be-  
676 tween T1w-nce and synthetic T1w-nce images (obtained with the *Att-U-Net* and  
677 the *cGAN Att-U-Net* models) for gray matter (left), white matter (middle) and cere-  
678 brospinal fluid (CSF, right) for both  $\text{Test}_{\text{good}}$  (top) and  $\text{Test}_{\text{low}}$  (bottom).

## List of Tables

- 1 MAE, PSNR and SSIM obtained on the two independent test sets with various image quality. For each metric, we report the average and standard deviation across the corresponding test set. We compute the metrics for both T1w-ce and synthetic T1w-nce in relation to the real T1w-nce, and so within the brain region.
- 2 Absolute volume difference (mean  $\pm$  standard deviation in  $\text{cm}^3$ ) between T1w-nce and T1w-ce images and between T1w-nce and synthetic T1w-nce images (obtained with the *Att-U-Net* and *cGAN Att-U-Net* models) for the gray matter, white matter and cerebrospinal fluid (CSF). \* indicates that the absolute volume difference between T1w-nce and synthetic T1w-nce images is statistically significantly different from that of the baseline (corrected p-value  $<0.01$ ) according to a paired t-test corrected for multiple comparisons using the Bonferroni correction.
- 3 Dice scores obtained when comparing the gray matter, white matter and cerebrospinal fluid (CSF) segmentations between T1w-nce and T1w-ce images and between T1w-nce and synthetic T1w-nce images (obtained with the *Att-U-Net* and the *cGAN Att-U-Net*)