

# Evaluation of pseudo-healthy reconstruction for anomaly detection in brain FDG PET

**Ravi Hassanaly**

RAVI.HASSANALY@ICM-INSTITUTE.ORG

**Camille Brianceau**

CAMILLE.BRIANCEAU@ICM-INSTITUTE.ORG

**Maëlys Solal**

MAELYS.SOLAL@ICM-INSTITUTE.ORG

**Olivier Colliot**

OLIVIER.COLLIOT@CNRS.FR

**Ninon Burgos**

NINON.BURGOS@CNRS.FR

*Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France*

**Editors:** Under Review for MIDL 2024

## Abstract

We propose an evaluation procedure based on the simulation of realistic abnormal images to validate pseudo-healthy reconstruction methods when no ground truth is available. We apply this framework to the reconstruction of 3D brain FDG PET using a convolutional variational autoencoder. This work has recently been published at MELBA.

**Keywords:** Deep learning, Pseudo-healthy reconstruction, Unsupervised anomaly detection, Variational autoencoder, 3D PET, Alzheimer’s disease

## 1. Introduction

Unsupervised anomaly detection (UAD) using deep generative models ([Chen and Konukoglu, 2022](#); [Zhang et al., 2023](#)) is an active field of research in medical imaging, especially in neuroimaging, where it has been widely applied to the detection of tumors and white matter hyper-intensities on structural magnetic resonance images (MRI) ([Baur et al., 2021](#)). We here focus on brain  $^{18}\text{F}$ -fluorodeoxyglucose (FDG) positron emission tomography (PET), a relevant modality to study Alzheimer’s disease (AD), as it allows imaging brain areas with hypometabolism ([Herholz, 1995](#)). However, when applying UAD in this context ([Choi et al., 2019](#); [Baydargil et al., 2021](#)), we do not have access to ground truth masks of the anomalies to evaluate the models.

We introduce a framework for the evaluation of pseudo-healthy reconstruction approaches in the absence of ground truth. It consists in simulating anomalies on images of healthy subjects to generate pairs of pathology-free and pathological (e.g., mimicking dementia-like lesions) images. We complement the framework by defining a new healthiness metric that measures whether the reconstructed image is of healthy appearance. We finally evaluate a VAE ([Kingma and Welling, 2014](#)) trained on 3D brain PET using the framework.

## 2. Methods

In addition to evaluating the quality of the generated images using reconstruction metrics that are common in the image synthesis literature ([Nečasová et al., 2022](#)), there is a need to evaluate if the reconstructions are looking healthy and allow detecting anomalies.

In most of the cases, datasets include lesion masks used as ground truth. However, for neurodegenerative disease studies using FDG PET, these masks are not available.

We propose to simulate hypometabolism on healthy images to have pairs of healthy (considered as ground truth) and abnormal images. We designed a mask corresponding to regions associated with AD (parietal and temporal lobes) (Landau et al., 2012). We then reduced the intensity of the PET signal within the region defined by the mask by different factors to simulate various degrees of hypometabolism as illustrated in Figure 1.

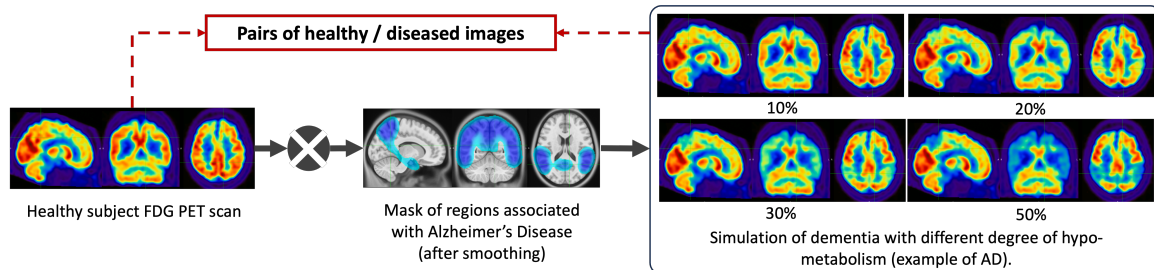


Figure 1: Hypometabolism simulation pipeline. The intensity of the image from a healthy subject is reduced by a chosen factor in a region associated with a dementia.

Furthermore, we define the “healthiness” metric  $H = \frac{\mu_M}{\mu_{\bar{M}}}$  to measure if the model is able to reconstruct images that are looking healthy.  $\mu_M$  is the average uptake in the region of the mask  $M$  used to simulate the anomaly, and  $\mu_{\bar{M}}$  is the average uptake of voxels in the brain excluding the mask  $M$ .

### 3. Experiments and results

FDG PET scans used in this study were obtained from the ADNI database (Jagust et al., 2010, 2015). After curating the dataset, we have 739 images from 378 cognitively normal (CN) subjects. We use the 60 baseline sessions from 60 CN subjects used for the test set to build new test sets by using our simulation method with different hypometabolism intensity degrees, resulting in a total of 8 simulated test sets.

We use a 3D convolutional VAE as VAEs have already shown their efficacy for UAD in medical imaging (Baur et al., 2021; Chen and Konukoglu, 2022).

We observe in Figure 2 that the input and output images of the CN subject are quite similar, both the shape of the brain and the uptake distribution look alike. The differences are due to the model imperfect reconstruction and correspond to the minimal error that it can achieve. When feeding the simulated hypometabolic image  $\mathbf{x}'$  to the model, we observe that the reconstructed image  $\hat{\mathbf{x}}'$  looks healthier than the input image. The areas highlighted in blue in the residual map correspond to the regions where hypometabolism was simulated.

We can see in Figure 3 that the healthiness of simulated images  $\mathbf{x}'$  is lower than for PET scans from CN subjects (the ground truth). However, the healthiness score of the reconstruction  $\hat{\mathbf{x}}'$  is always superior to the one of the simulated image  $\mathbf{x}'$ . We can see that it is even really close to the healthiness of the original image  $\mathbf{x}$ .

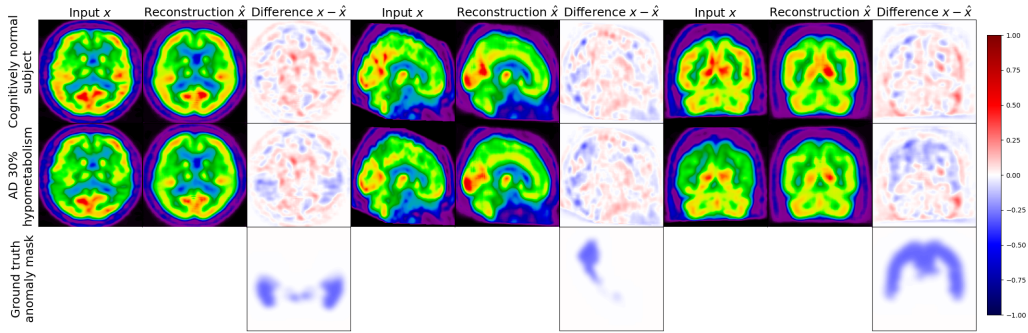


Figure 2: Results obtained from a real image of a CN subject (top row) and an image simulating AD hypometabolism based on the same CN subject (bottom row).

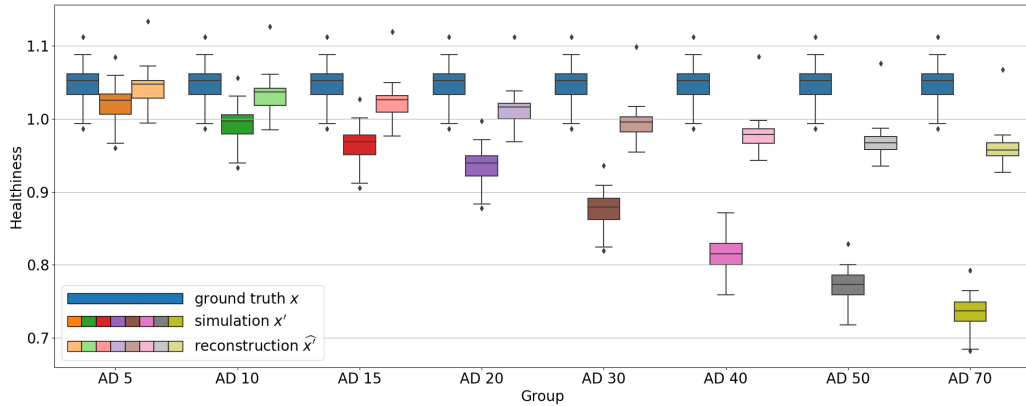


Figure 3: Distribution of the healthiness metric for the ground truth healthy images, their corresponding simulated images and their pseudo-healthy reconstructions when increasing the percentage of AD-like simulated hypometabolism.

#### 4. Discussion and conclusion

To overcome the absence of ground truth for the evaluation of the model, we introduced a framework to simulate dementias related diseases from images of healthy subjects. This framework has been applied to a 3D VAE that is suited to detect anomalies due to dementia on brain FDG PET. Thanks to simulated test sets and the defined healthiness metric, we could evaluate the model ability to reconstruct healthy looking images. This work has recently been published at MELBA ([Hassanally et al., 2024](#)).

#### Acknowledgments

Funding was received from ANR (ANR-19-P3IA-0001, ANR-10-IAIHU-06). This work was granted access to to IDRIS HPC resources (AD011011648) made by GENCI.

## References

- Christoph Baur, Stefan Denner, Benedikt Wiestler, Nassir Navab, and Shadi Albarqouni. Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study. *Medical Image Analysis*, 69:101952, 2021. doi: 10.1016/j.media.2020.101952.
- Husnu Baris Baydargil, Jang-Sik Park, and Do-Young Kang. Anomaly analysis of alzheimer’s disease in PET images using an unsupervised adversarial deep learning model. *Applied Sciences*, 11(5):2187, 2021. doi: 10.3390/app11052187.
- Xiaoran Chen and Ender Konukoglu. *Unsupervised abnormality detection in medical images with deep generative methods*, pages 303–324. Elsevier, 2022.
- Hongyoon Choi, Seunggyun Ha, Hyejin Kang, Hyekyoung Lee, Dong Soo Lee, and Alzheimer’s Disease Neuroimaging Initiative. Deep learning only by normal brain PET identify unheralded brain anomalies. *EBioMedicine*, 43:447–453, 2019. doi: 10.1016/j.ebiom.2019.04.022.
- Ravi Hassanaly, Camille Brianceau, Maëlys Solal, Olivier Colliot, and Ninon Burgos. Evaluation of pseudo-healthy image reconstruction for anomaly detection with deep generative models: Application to brain FDG PET. *Machine Learning for Biomedical Imaging*, 2: 611–656, 2024. doi: 10.59275/j.melba.2024-b87a.
- Karl Herholz. FDG PET and differential diagnosis of dementia. *Alzheimer Disease and Associated Disorders*, 9(1):6–16, 1995. doi: 10.1097/00002093-199505000-00004.
- William J. Jagust, Dan Bandy, Kewei Chen, Norman L. Foster, Susan M. Landau, Chester A. Mathis, Julie C. Price, Eric M. Reiman, Daniel Skovronsky, and Robert A. Koeppe. The Alzheimer’s Disease Neuroimaging Initiative positron emission tomography core. *Alzheimer’s & Dementia*, 6(3):221–229, 2010. doi: 10.1016/j.jalz.2010.03.003.
- William J. Jagust, Susan M. Landau, Robert A. Koeppe, Eric M. Reiman, Kewei Chen, Chester A. Mathis, Julie C. Price, Norman L. Foster, and Angela Y. Wang. The Alzheimer’s Disease Neuroimaging Initiative 2 PET Core: 2015. *Alzheimer’s & Dementia*, 11(7):757–771, 2015. doi: 10.1016/j.jalz.2015.05.001.
- Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2014.
- Susan M. Landau, Mark A. Mintun, Abhinay D. Joshi, Robert A. Koeppe, Ronald C. Petersen, Paul S. Aisen, Michael W. Weiner, and William J. Jagust. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Annals of Neurology*, 72(4):578–586, 2012. doi: 10.1002/ana.23650.
- Tereza Nečasová, Ninon Burgos, and David Svoboda. Validation and evaluation metrics for medical and biomedical image synthesis. In Ninon Burgos and David Svoboda, editors, *Biomedical Image Synthesis and Simulation*, The MICCAI Society Book Series, pages 573–600. Academic Press, 2022.
- Chuyan Zhang, Hao Zheng, and Yun Gu. Dive into the details of self-supervised learning for medical image analysis. *Medical Image Analysis*, page 102879, 2023.