Artificial Intelligence Research and Development
T. Alsinet et al. (Eds.)

doi:10.3233/FAIA240415

Unsupervised Deep Learning Architectures for Anomaly Detection in Brain MRI Scans

Jordi MALÉ [a1], Víctor XIRAU [a], Juan FORTEA [c], Yann HEUZÉ [d], Neus MARTÍNEZ-ABADÍAS [b] and Xavier SEVILLANO [a]

[a] *HER - Human-Environment Research Group, La Salle - URL, Barcelona, Spain*
[b] *Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals (BEECA), Facultat de Biologia, Universitat de Barcelona (UB), Barcelona, Spain*
[c] *Sant Pau Memory Unit, Hospital de Sant Pau i la Santa Creu, Barcelona, Spain*
[d] *Univ. Bordeaux, CNRS, Ministère de la Culture, PACEA, UMR 5199, Pessac, France*

Abstract. Brain imaging techniques, particularly magnetic resonance imaging (MRI), play a crucial role in understanding the neurocognitive phenotype and associated challenges of many neurological disorders, providing detailed insights into the structural alterations in the brain. Despite advancements, the links between cognitive performance and brain anatomy remain unclear. The complexity of analyzing brain MRI scans requires expertise and time, prompting the exploration of artificial intelligence for automated assistance. In this context, unsupervised deep learning techniques, particularly Transformers and Autoencoders, offer a solution by learning the distribution of healthy brain anatomy and detecting alterations in unseen scans. In this work, we evaluate several unsupervised models to reconstruct healthy brain scans and detect synthetic anomalies.

Keywords. Unsupervised Deep Learning, Autoencoders, Brain MRI Scans, Anomaly detection

1. Introduction

Brain magnetic resonance imaging (MRI) is crucial for diagnosing and understanding various disorders. Expert neurologists and radiologists use their knowledge of brain anatomy to identify anomalies in MRI scans, but visual analysis can miss relevant pathologies in 5 to 10% of cases [1]. Recent machine learning advances have led to automated techniques that identify abnormalities, showing success in detecting tumors and lesions related to neurodegenerative diseases like Alzheimer's and multiple sclerosis [2].

While most techniques use supervised deep learning, the scarcity of annotated datasets has prompted a shift to Unsupervised Anomaly Detection (UAD). UAD methods mimic expert examination of MRI scans without needing annotated data, thus identifying various brain anomalies.

Recent UAD approaches, employing autoencoders and generative models, not only detect but also locate and delineate anomalies. These methods use deep representation learning to define the healthy brain's anatomical distribution, identifying anomalies as

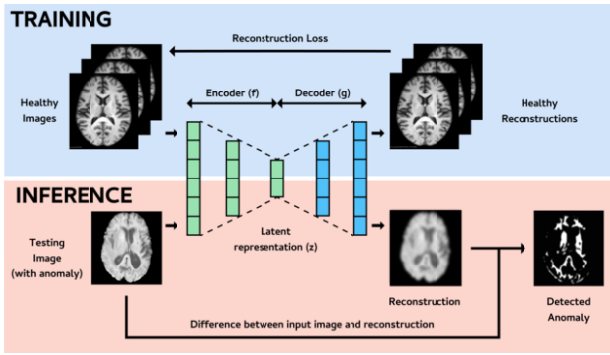

Figure 1. Unsupervised Anomaly Detection using an Autoencoder-based architecture

outliers. Autoencoders (AEs) are particularly popular, learning to compress and reconstruct MRI data of healthy anatomy. Figure 1 illustrates how AEs are used for brain UAD: trained on healthy scans, the AE learns the latent space distribution and reconstructs input scans. When fed with an anomalous scan, the AE produces a pseudo-healthy reconstruction, which is compared to the input to detect anomalies.

The goal of this paper is to implement and evaluate the performance of four unsupervised deep learning models—Vector Quantization Variational AEs (VQ-VAE), Masked AEs (Swin MAE), Reverse AEs (RAE), and VQ-VAE+Transformers—in detecting and localizing brain anomalies using structural MRI scans. We intend to use these methods to automatically detect altered regions and define diagnostic-potential biomarkers in psychotic disorders, furthering our understanding of the neurodegenerative effects of these disorders and improving diagnosis.

2. Experimental Setup

Four unsupervised deep learning models were implemented and evaluated: Vector Quantization Variational AEs (VQ-VAE[2]) [3], Masked AEs (Swin MAE[3]) [4], Reverse AEs (RAE[4]) [5], and VQ-VAE+Transformers[5] [6]. Our implementations followed either MONAI or the public official implementation, and was developed using Python 3.9 and PyTorch. We ran the experiments on NVIDIA Tesla V100 GPUs. Table 1 summarizes the main training hyperparameters of each method (in all cases, the ADAM optimizer and data augmentation were used during training).

Models were trained using 1113 structural MRI scans of healthy brains obtained from the Human Connectome Project (HCP)[6]. The dataset was split into 90% for training (1002 scans) and 10% for evaluation (111 scans).

[2]https://github.com/Project-MONAI/GenerativeModels
[3]https://github.com/Zian-Xu/Swin-MAE
[4]https://github.com/ci-ber/RA
[5]https://github.com/Project-MONAI/GenerativeModels
[6]Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Method	Batch size	Epochs	Loss function(s)	Learning rate
VQ-VAE	32	1000	L1	$5 \cdot 10^{-4}$
Swin MAE	32	1000	L1	$10^{-3} \rightarrow 10^{-6}$
Reverse AE	8	1000	MSE + KL + Embedding loss	$2 \cdot 10^{-4} \rightarrow 10^{-6}$
VQ-VAE+Tr	4	500	Cross-Entropy	10^{-4}

Table 1. Training hyperparameters of the evaluated UAD methods.

To evaluate the ability of the trained models to generalize across various sources of data, we tested the methods using three different datasets of healthy brain structural T1 MRI scans: *i)* 160 scans provided by Hospital Sant Pau Memory Unit (Barcelona, Spain), *ii)* 260 scans from the OASIS-3 dataset [7], and *iii)* 580 scans from the IXI dataset [8].

In our study, we introduce synthetic anomalies into MRI images to test unsupervised anomaly detection methods. We add rectangular and elliptical anomalies by copying the original image and inserting regions of specified dimensions (width and height between 20 and 30 pixels) and intensity at random center points within the image. Half of the anomalies are rectangles, and the other half are ellipses. These synthetic anomalies simulate irregularities in MRI scans, allowing us to evaluate the performance of anomaly detection algorithms effectively.

Two aspects of each model were evaluated. First, their fidelity in reconstructing healthy MRI scans, which was measured in terms of the structural similarity index metric (SSIM). And second, their ability to detect synthetically added anomalies and produce anomaly maps by combining residual (L1 loss) and perceptual differences. To evaluate the accuracy of the anomaly segmentations, we reported the area under the precision-recall curve (AUPRC) and the area under the receiver operating characteristic (AUROC).

3. Results, Conclusions and Further Work

Table 2 presents the evaluation of the four UAD methods for both healthy MRI scan reconstruction (in terms of SSIM on the three test datasets) and synthetic anomaly detection accuracy. Moreover, Figure 2 illustrates qualitative examples of the results obtained by each method.

Figure 2 provides a qualitative example of our anomaly detection method, evaluating four models. Two different tests are shown: a healthy brain (first row) and a brain with a randomly introduced synthetic anomaly (second row). For each input image and model, the reconstruction and computed anomaly map are displayed. Our preliminary results indicate that the VQ-VAE + Transformers approach achieves the best balance, producing superior pseudo-healthy reconstructions and generalizing well across heterogeneous datasets. In contrast, while VQ-VAE delivers good results for healthy inputs, it fails to detect anomalies accurately. Swin-MAE can capture anomalies, as demonstrated, but it relies heavily on pixel intensities. Finally, the Reverse Autoencoder (RAE) is effective at detecting anomalies but alters the brain anatomy in its healthy reconstructions.

To enhance performance, research should investigate new loss functions such as perceptual or adversarial losses, and incorporate 3D architectures for volumetric data. Additionally, models should be trained on data from various datasets to enhance robustness and generalization ability. Additionally, these models should be tested using real anomalies, ranging from severe conditions like tumors to minor changes caused by disorders such as Down Syndrome or Alzheimer's disease to help identify potential biomarkers.

Table 2. Quantitative evaluation of Unsupervised Anomaly Detection architectures.

Method	OASIS	IXI	St. Pau	HCP	Synthetic Anomalies	
	SSIM ↑	SSIM ↑	SSIM ↑	SSIM ↑	AUROC ↑	AUPRC ↑
VQ-VAE [3]	0.96	0.96	0.96	0.96	0.80	0.28
Swin-MAE [4]	0.80	0.78	0.81	0.84	0.88	0.38
Reverse AE [5]	0.55	0.58	0.58	0.63	0.95	0.67
VQ-VAE + Transformers [6]	0.93	0.93	0.93	0.93	0.94	0.74

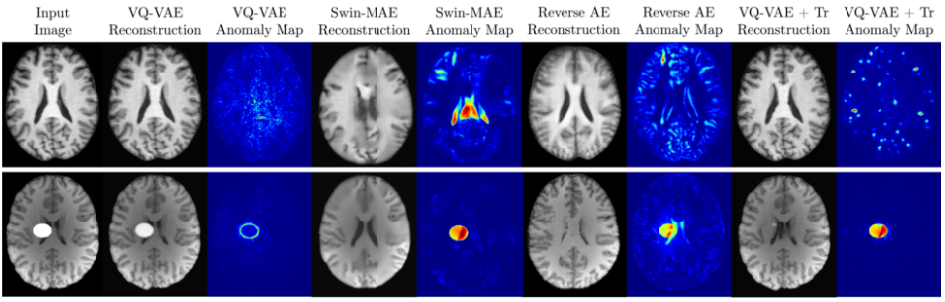

Figure 2. Qualitative results of the brain reconstruction methods (top row, healthy brains, bottom row: brains with synthetic anomaly): (a) Input image, (b) VQ-VAE (c) Swin-MAE, (d) RAE, (e) VQ-VAE + Tr.

Acknowledgements

This work was partly supported by Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) of the Generalitat de Catalunya (2021 SGR01396, 2021 SGR00706), the Spanish Ministry of Science, Innovation, and Universities under grant PID2020-113609RB-C21, the Fondation Jerome Lejeune under grant 2020b cycle-Project No.2001, and the Joan Oró grant (FI 2024) from the DRU of the Generalitat de Catalunya and the European Social Fund (2024 FI-2 00014).

References

[1] Bruno MA, Walker EA, Abujudeh HH. Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. RadioGraphics. 2015;35(6):1668-76.
[2] Baur C, Denner S, Wiestler B, Albarqouni S, Navab N. Autoencoders for Unsupervised Anomaly Segmentation in Brain MR Images: A Comparative Study; 2020.
[3] van den Oord A, Vinyals O, Kavukcuoglu K. Neural Discrete Representation Learning; 2018.
[4] Xu Z, et al. Swin MAE: Masked Autoencoders for Small Datasets; 2023.
[5] Bercea CI, Wiestler B, Rueckert D, Schnabel JA. Generalizing Unsupervised Anomaly Detection: Towards Unbiased Pathology Screening. In: Medical Imaging with Deep Learning; 2023. .
[6] Pinaya WHL, Tudosiu PD, Gray R, Rees G, Nachev P, Ourselin S, et al.. Unsupervised Brain Anomaly Detection and Segmentation with Transformers; 2021.
[7] LaMontagne PJ, et al. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease. medRxiv. 2019.
[8] Imperial College London UCL, of Edinburgh U. Information eXtraction from Images (IXI) Dataset;. Accessed: 2024-05-21. https://brain-development.org/ixi-dataset/.