Self-Supervised Anomaly Detection in Brain MRI Using Convolutional Autoencoders and Masked Autoencoders

Niamatellah Lahkim
IMT Atlantique, Brest, France

Summary This project studies self-supervised anomaly detection in brain MRI using reconstruction-based models trained exclusively on healthy slices. Two architectures were implemented under identical training conditions: a Convolutional Autoencoder (CNN AE) and a Vision Transformer-based Masked Autoencoder (MAE). The comparison focuses on reconstruction behaviour, error localisation, and anomaly detection performance. Our experiments show that, although the MAE produces blurrier reconstructions, it achieves competitive or superior slice-level ROC–AUC compared to the CNN AE. This suggests that convolutional and transformer-based representations may capture complementary aspects of brain anatomy and may both be valuable for medical anomaly detection.

Background

Annotating brain MRI is costly and requires expert knowledge. Reconstruction-based self-supervised learning provides a viable alternative by training only on healthy anatomy and detecting anomalies as regions the model cannot reconstruct. CNN autoencoders are widely used because they capture local spatial details [1]. Masked Autoencoders (MAEs), in contrast, rely on Vision Transformers that learn global context through random patch masking [2]. Their behaviour in medical anomaly detection remains under investigation [3]. This work offers an empirical comparison between both approaches under controlled and identical training conditions.

Aim

The aim of this research is to better understand how CNNs and MAEs differ in reconstructing healthy brain slices and how these differences influence anomaly detection. The objective is not to produce perfect segmentations but to analyse reconstruction quality, the behaviour of reconstruction errors inside tumour regions, global slice-level anomaly separation, and the strengths and limitations of each model.

Methods

Dataset and preprocessing. We use the BraTS 2021 dataset [4]. All T1-weighted MRI volumes are converted into 128×128 axial slices and normalised to $[0, 1]$. Only tumour-free slices are used for training, while testing includes both healthy and tumour slices. After subsampling, around 1,600 healthy slices were kept for training and validation.

CNN Autoencoder. A U-Net–style convolutional autoencoder with skip connections was trained using MSE reconstruction loss, allowing the model to reconstruct high-frequency details typical of healthy brain anatomy. Pixel-wise squared error during inference provides anomaly maps.

MAE Autoencoder. A Vision Transformer MAE (ViT-B/16), pretrained on ImageNet, was adapted for MRI reconstruction. Each slice is resized to 224×224, converted to three channels, and 75% of patches are randomly masked. The encoder processes the visible patches while a lightweight CNN decoder reconstructs the full image, later downsampled to 128×128.

Evaluation. Both models were trained using identical reconstruction loss, the same healthy slices, and the same batch size and number of epochs. Performance was evaluated using slice-level ROC–AUC, pixel-wise PR–AUC, and the comparison between mean reconstruction error inside and outside tumour regions.

Results and Discussion

To illustrate the behaviour of the models, Fig. 1 shows a tumour slice with its tumour contour overlaid, while Fig. 2 presents the reconstruction error inside the tumour region for both models.

CNN Autoencoder. The CNN AE produces sharp reconstructions and exhibits clear localised errors in tumour zones. Quantitatively, it reaches a slice-level ROC–AUC of 0.8676 and a PR–AUC of 0.9962. On representative slices, the reconstruction error inside the tumour is approximately 2.5 times higher than outside, confirming that CNNs detect anomalies through texture inconsistencies.

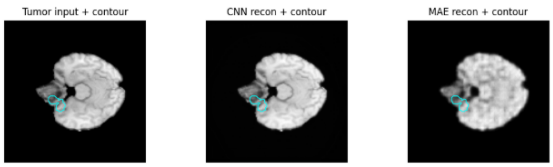

Figure 1: Tumour slice with contour overlay used for both CNN and MAE evaluation.

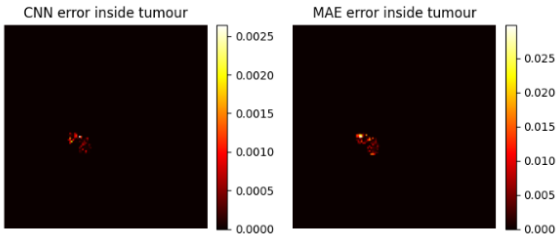

Figure 2: Reconstruction error inside tumour region for CNN AE and MAE AE.

MAE Autoencoder. The MAE generates blurrier reconstructions due to its low-frequency global representations, yet anomaly detection remains highly effective. It achieves a ROC–AUC of 0.9039 and PR–AUC of 0.9971. The model highlights tumour regions not by texture mismatch but through global structural deviations, leading to stable anomaly scores despite visual blur.

The models demonstrate complementary behaviours: the CNN excels at detailed local reconstruction, producing sharp reconstructions and precise error peaks, while the MAE achieves strong anomaly separation thanks to its global contextual understanding. These findings show that reconstruction sharpness alone does not determine anomaly detection performance; global structural coherence can be equally important.

Conclusion and Perspectives

This work provides a controlled and systematic comparison of CNN and MAE autoencoders for self-supervised anomaly detection in brain MRI. Both approaches can identify tumour regions without supervision but rely on fundamentally different mechanisms. Future work will explore hybrid CNN–MAE models, 3D reconstruction architectures, and advanced post-processing such as CRF or morphological refinement to improve anomaly localisation.

References

[1] C. Baur et al., *Modeling and Segmentation of Anomalies in Medical Images: A Survey*, Medical Image Analysis, 2021.

[2] K. He et al., *Masked Autoencoders Are Scalable Vision Learners*, CVPR, 2022.

[3] W. Pinaya et al., *Unsupervised brain anomaly detection and segmentation with transformers*, Medical Image Analysis, 2022.

[4] U. Baid et al., *RSNA-ASNR-MICCAI BraTS 2021 Benchmark*, arXiv:2107.02314.