Self-Supervised Machine Learning on Brain MRI Data

be.AI

S. Kuznetsov Supervised by Dr. Ivor J. Simpson

UNIVERSITY OF SUSSEX

MOTIVATION

- Lesion segmentation is crucial for various clinical tasks like medical diagnostics, prognosis and treatment.
- It is typically performed manually, rendering the process time-consuming and heavily reliant on comprehensive knowledge of human anatomy.
- The goal of our study is to methodically explore and compare a selection of self-supervised medical image analysis approaches for lesion detection and segmentation that circumvent the requirement of labeled data.

SELF-SUPERVISED ANOMALY DETECTION

A potential solution for anomaly detection entails utilizing self-supervised learning techniques to model healthy human brain distribution in latent space. During inference, anomalous data within input are identified as outliers, resembling radiologists' decision-making processes without manual annotation and solely relying on a *priori* knowledge of pathology manifestations.

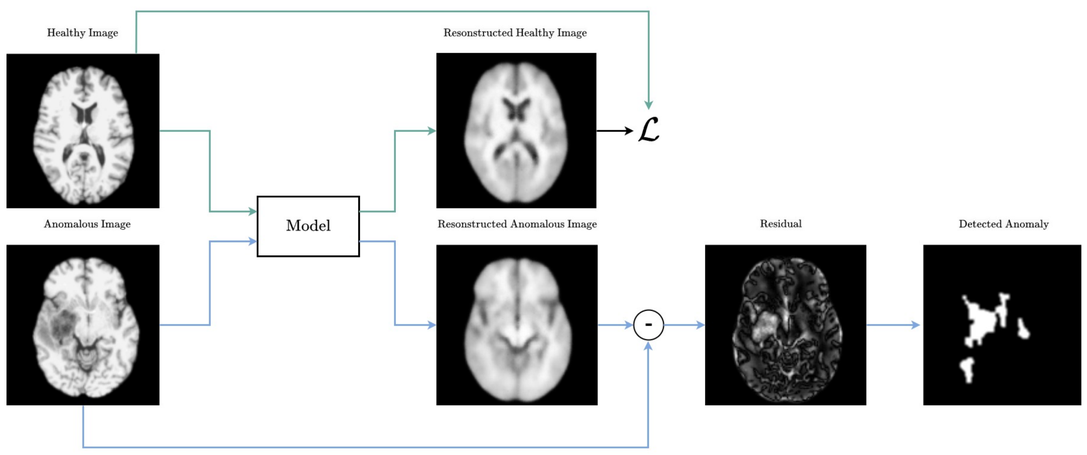


Figure 1: Self-Supervised anomaly detection.

METHODOLOGY

In our research, we employed probabilistic graphical models built upon the framework of Variational Inference to approximate the true posterior distribution of the data in the latent space $p(z|x)$ with another distribution $q_\phi(z|x)$. A prominent example of such a model is the Variational Autoencoder (VAE) [Kingma et. al. 2013], which learns the underlying data distribution by mapping it to a normative distribution in the lower-dimensional space. Since the model has not been trained on data containing anomalies, it might not be able to capture the anomalous features accurately. Instead, it will likely generate a reconstruction that resembles a healthy brain as closely as possible, given the input data. The anomalous regions can then be retrieved using image post-processing techniques.

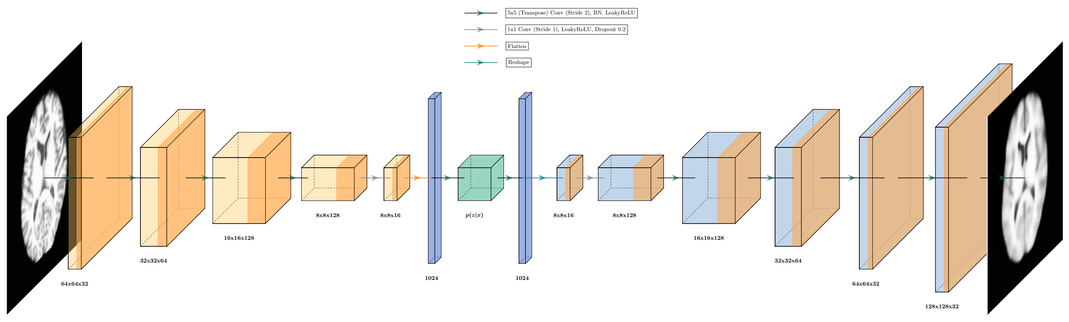


Figure 2: The architecture of Variational Autoencoder (VAE) model used in this study.

The primary objective of a VAE is to minimize the reverse Kullback–Leibler (KL) divergence between the ground truth posterior and the approximation of the posterior:

$$D_{KL} = \mathbb{E}_{z\sim q_\phi(z|x)}\left[\log \frac{q_\phi(z|x)}{p(z|x)}\right] \tag{1}$$

Which, after some simplifications, can be reduced to a computable loss function:

$$\mathcal{L}_{VAE}(x;\phi,\theta) = \underbrace{-\mathbb{E}_{z\sim q_\phi(z|x)}\Big[\log p_\theta(x|z)\Big]}_{\text{Reconstruction Loss}} + \underbrace{D_{KL}\Big(q_\phi(z|x)\|\, p(z)\Big)}_{\text{Regularization term}} \tag{2}$$

Where $\mathbb{E}_{z\sim q_\phi(z|x)}$ term in the reconstruction loss denotes the expected value, which is approximated by sampling from $q_\phi(z|x)$. The reconstruction loss can often be interpreted as Gaussian Negative Log-Likelihood (NLL) and the prior in the regularization term follows a standard Gaussian distribution.

LESION SEGMENTATION

Within the post-processing pipeline, we calculate the residual image by computing the difference between the original input image and its corresponding reconstruction. The expectation is that higher intensity values will be observed within the abnormal regions of the residual image. Subsequently, binary thresholding and morphological operations are applied to isolate the abnormal regions.

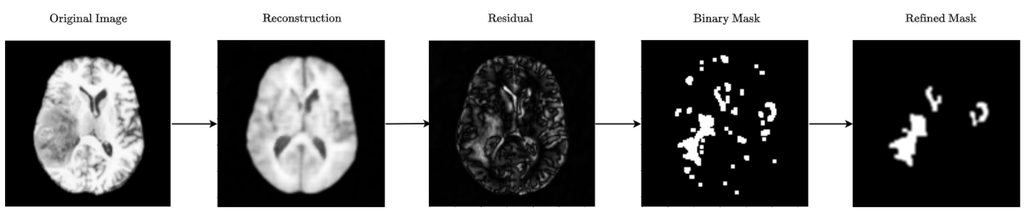
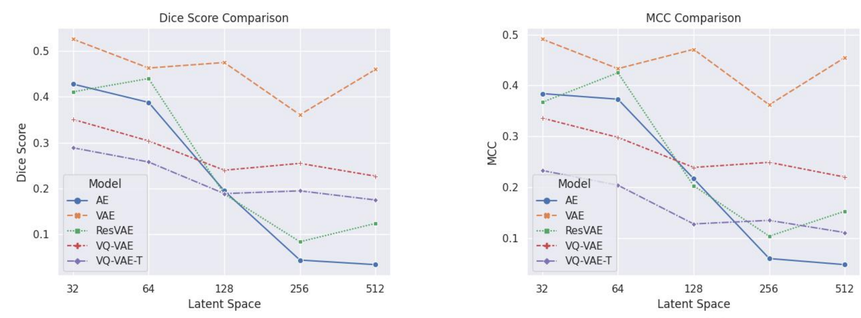


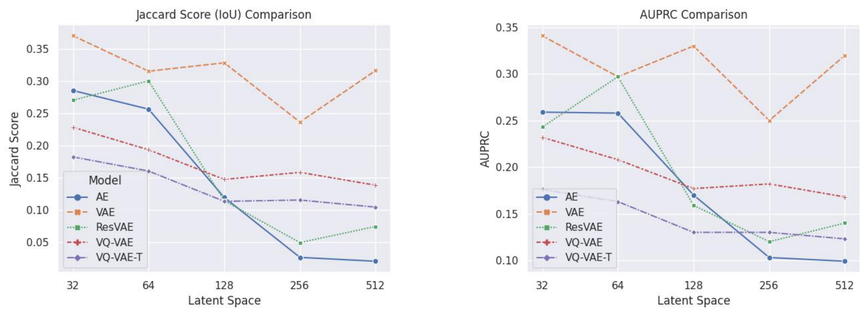
Figure 3: Postprocessing pipeline for a single brain image containing anomalies.

MODEL COMPARISON

In this study, we utilize an array of neural network-based models, including the standard Autoencoder (AE), Variational Autoencoder (VAE), Vector-Quantized Variational Autoencoder (VQ-VAE) [Van Den Oord et al. 2017], as well as a hybrid model that combines a VQ-VAE with a Transformer architecture [Pinaya et al. 2022]. We also conduct a comparative evaluation of the previously mentioned models for different latent space dimensions.



a Dice **b** MCC

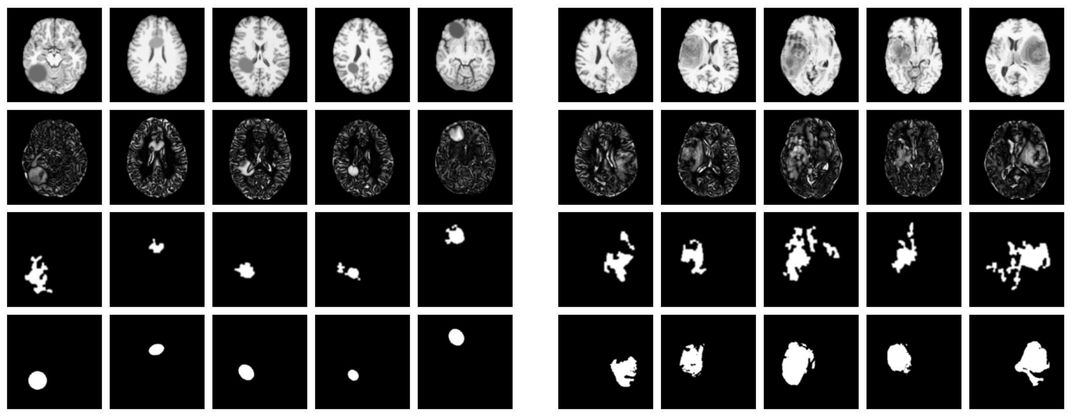


c Jaccard (IoU) **d** AUPRC

Figure 4: Evaluation of the models employed in this study on the BraTS'20 [Menze et al. 2014] dataset.

PREDICTIONS ON ANOMALOUS NEUROIMAGING DATA

The developed models were employed to analyze data coming from two distinct modalities: (1) data with synthesized anomalies and (2) real anomalous neuroimaging data.



a Synthesized anomalies. **b** Real anomalies.

Figure 5: Predictions of the best-performing model (VAE 32) on anomalous neuroimaging data. From top to bottom: Original input images, residuals between original and reconstructed images, predicted segmentation masks, ground truth segmentation masks.

DISTRIBUTION OF RESIDUALS FOR HEALTHY AND ANOMALOUS DATA SAMPLES

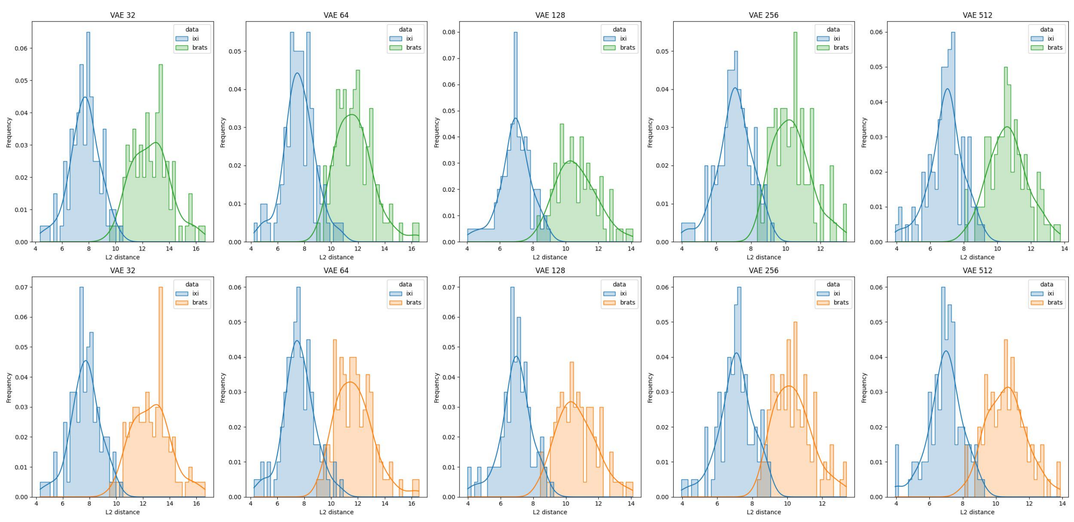


Figure 6: Histogram of $l2$ distances between original and reconstructed images from the used datasets for the VAE model. Top row: healthy images and synthesized anomalous images, bottom row: healthy images and real anomalous images.

CONCLUSIONS

- Developed and implemented pre-processing and post-processing pipelines for the analysis of brain MRI data
- Developed and trained multiple self-supervised machine learning models for the purpose of identifying and classifying anomalies within brain MRI images.
- Evaluated the performance of implemented models by comparing their respective metrics, determining the most efficient and accurate configuration.

- *"Auto-encoding variational bayes", Kingma, Diederik P and Welling, Max. 2013*
- *"The multimodal brain tumor image segmentation benchmark (BRATS)", Menze et al. 2014*
- *"Neural discrete representation learning", Van Den Oord et al. 2017*
- *"Unsupervised brain imaging 3d anomaly detection and segmentation with transformers", Pinaya et al. 2022*