Identifying patients with similar brain tumors using auto-encoding and approximate nearest neighbours

1st Tilen Limbäck-Stokin*
Undergraduate Programme
Dept. of Computer Science
University College London
London, United Kindgdom
tilen.limback-stokin.21@ucl.ac.uk

2nd Matt Williams
Computational Oncology Laboratory
Institute of Global Health Innovation
Imperial College London
London, United Kingdom
matthew.williams@imperial.ac.uk

***Abstract*—Background: Brain tumours are the leading cause of cancer death in the under-40's. Imaging of tumours is crucial to diagnosis, prognosis and assessing response to treatment. Existing computational approaches to brain tumour imaging are time and resource intensive. We introduce a simple, scalable technique to enable image-based query using a convolutional autoencoder and nearest-neighbour search to identify 'similar' brain tumour images to a query case.**

Methods: We used MRI scans from 102 patients from a public dataset. We trained a convolutional autoencoder to extract the latent features of the MRI and used that as a vector space to be queried via a nearest neighbour approach. We used this to identify the five most similar patients for each query patient. We assessed autoencoder performance using standard performance metrics, and the performance of the whole system by assessing how performance on a small set of previously unseen query patients.

Results: The autoencoder achieved reasonable performance, with an average PSNR of 42.15 dB and SSIM value of 0.46, and was relatively insensitive to noise injection. Images retrieved for the 7 query cases were able to successfully match our GBM input to other GBM patients 88% of the time and 60% for LGG. Location matching was reasonable, with 72% correct for GBM inputs and 70% for LGG inputs.

Conclusions: We have developed an image search approach based on CAE and nearest neighbours that allows us to find patients similar to a query patient. Performance is reasonable, given the small dataset, and importantly, is easily scalable.

***Index Terms*—auto-encoding, nearest neighbour, neural network, brain tumor, MRI**

I. Introduction

Brain tumours are the leading cause of cancer death in the under-40's. Imaging of tumours is crucial to diagnosis, prognosis and assessing response to treatment. Computational assessment of brain tumour imaging has been an active area of research, e.g. the MICCAI BRATS challenge [1]. Such work has focused on segmentation-based approaches. However, manual segmentation of brain tumour images is time consuming, which limits utility and widespread adoption of such approaches. [2]. More scalable approaches, such as weak labelling still need each slice to be manually labelled. There is value in being able to answer simple questions, such as 'which patients have tumours similar to this patient's tumour', which do not necessarily require segmentation, and may be achievable using simpler approaches such as example-based search technique, that allows us to find 'images like this one'. In this work, we show that a simple approach based on sequential use of an autoencoder and clustering allows us to effectively find patients with similar scans.

A. Technical Background

Autoencoders: autoencoders (AE) are an unsupervised machine learning technique that consists of training a neural network with layers of reducing size, followed by an expansion. AEs are used to achieve latent feature representation of the input, and training is conducted by comparing input-output similarity and hence is relatively simple and rapid. The encoder transforms input images from higher dimensions into lower dimensional representations using non-linear transformations until the final 'bottleneck' is achieved, when the decoder then attempts to reconstruct the input image, and the output of the bottleneck layer generates a latent space based on the input dataset. In this work we specifically use a *convolutional* autoencoder as the data we are working with are images [3] [4].

Autoencoder metrics & performance: Given that an autoencoder aims to recreate the input signal, the simplest and most commonly used measurement of error is the mean squared error loss function, which can be minimised as part of training. However, there is a well recognised problem with MSE and L2 based metrics not reflecting human perception of images. For this reason, most work also uses the Peak Signal-to-Noise ratio (PSNR) and Structural Similarity Metric (SSIM) which consider factors that are more appreciated by humans. PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index). Here PSNR in decibels (Db) is defined as:

$$\text{PSNR} = 20 \times \log_{10}\left(\frac{1}{\sqrt{MSE}}\right)$$

while SSIM is defined as follows [5]:

$$\text{SSIM}(x, y) = \left([l(x, y)^{\alpha}] \times [c(x, y)^{\beta}] \times [s(x, y)^{\delta}]\right)$$

Where l is the luminosity, c the contrast, and s the structure comparison between two signals x and y. Calculation of PSNR and SSIM are often also performed following addition of

gaussian noise to reduce overfitting. Typical PSNR values from the literature are above 40dB, and typical SSIM are 0.95 [6] [7], where a SSIM of 0.95 or greater means the reconstruction is similar enough to the original that humans cannot distinguish them. We use PSNR it as quantifies the quality of the reconstruction and SSIM as it evaluates the structure of the images.

Distances, Similarity and Search The latent feature space that is learned by an autoencoder can be used to identify similar images. The simplest is to consider the distance between two vectors in the latent space based on their cosine distance, defined as:

$$\text{cosine distance} = \sqrt{2 - 2\cos(u, v)}$$

This then gives us an intuitive way of measuring the distance between two images. We can then retrieve instances that are close to each other using well-described, efficient search techniques, such as K-Nearest Neighbours.

Slices, scans, patients and tumours We are interested in performing similarity-based query in medical imaging. Even if we only consider one scan per patient, that scan consists of multiple slices and adjacent slices are likely to be reasonably similar. Therefore a naive approach that retrieves the slices that are "most similar" to a slice of a scan for patient P, we are likely to return multiple sequential slices from another patient, P'. We therefore assume that for search purposes, each patient appears only once in the dataset, and when we identify similar slices from scans, we aggregate slice similarity to return scans that are similar (where each scan may contain multiple slices, each of which are similar to the query slice). Finally, different tumour types have different appearances on MRI. In our case, low-grade gliomas (LGG) and Glioblastoma (GBM) have very different appearances, with GBM (characteristically) having a necrotic centre and bright, contrast-enhancing ring outline, and LGG showing low signal on T1, with minimal contrast enhancement. In addition, standard MRI imaging will include multiple sequences that highlight different aspects of anatomy and function.

B. Related Work

Neighbourhood-based techniques have been widely used in medical image analysis, including the counting of kidney stones and leukemia cells in images [8] [9]. Previous work has taken a similar approach to us, but used wavelet transform and PCA for dimensionality reduction [10]. Similarly, there is extensive previous work on the use of autoencoders in medical imaging but much of that work has focused on denoising existing images, and generative approaches to expand datasets [11] [12]. There is also an extensive body of work on computational approaches to brain tumour imaging, but much of that is based around segmentation-based approaches, rather than querying for similar patients.

II. Methods

Data: We used MRI scans from two different public datasets(TCGA-GBM & TCGA-LGG), restricted to MRIs with 20 - 30 slices. This resulted in a total of 80 Glioblastoma Multiforme (GBM) patients and 22 Low Grade Glioma (LGG)

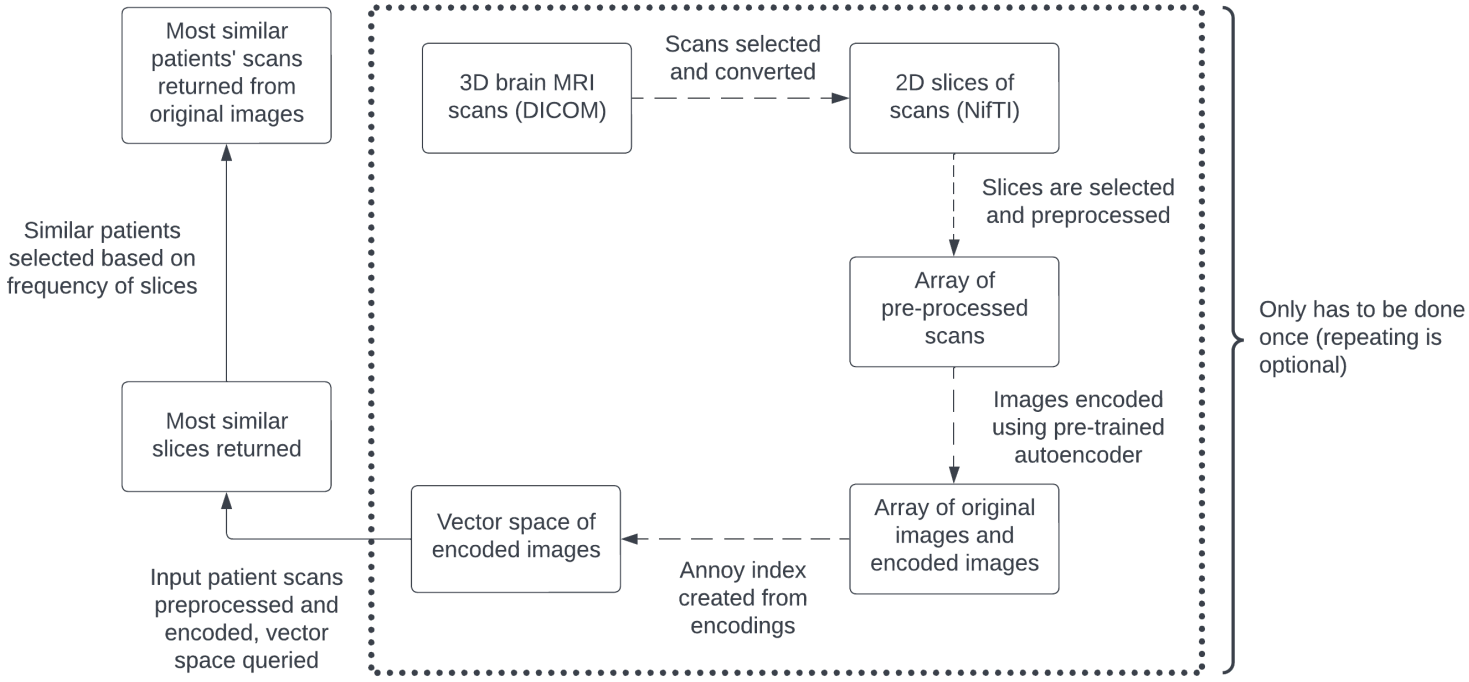

Fig. 1. Outline of pipeline for image search

patients [13] [14]. For this work we only considered axial T1-weighted post-contrast MRI sequences. Our code for the autoencoder, KNN, and evaluation is publicly available on gitlabs [15]

Image pre-processing: We converted MRI scans from DICOM to NIFTI and extracted the central 20 slices of each scan, resized them to 256x256, normalised using min-max normalisation, and added four cell padding vertically and horizontally. The final data therefore consists of 20 images, each 260x260, for each patient, for a total of 2040 images. We divided patients into 70/30 split train/test dataset, gving train/test datasets of 102 patients and 1428 train and 612 test images. The 7 patients in the test dataset were used as query patients to test our approach.

Extracting Latent Features: We trained a CAE on all slices in the the training dataset using architecture for MRI image reconstruction as previously described [16]. It consists of three convolution blocks, each consisting of a convolution layer and batch normalisation layer. After the first and second convolution blocks there is a max-pooling layer, which downsamples the input by two times. All the convolution blocks use filters of size 3x3, where the blocks, in order, have: 32, 64, and 128 filters. The decoding architecture uses only two convolution blocks, structured the same way as is in the encoder. Each convolution block is followed by a max-pooling layer, which upsamples the input by two times. The first block has 128 filters, the second 64 filters, and at the end of the decoder there is an additional layer with 1 filter to reconstruct the input. We trained with a batch size of 32, for 200 epochs, using the RMSProp optimiser. We then used the trained CAE on each slice in the dataset to extract the encoded latent feature representation from the autoencoder bottleneck. The bottleneck layer in our CAE has dimensions 65x65x128, giving a vector representation of length 540800.

Identifying similar tumors: We consider each query patient in turn. We iterate through each slice of the query patient and identify the 20 most similar images to the query slice. Then we look-up which patient those similar images belong to and aggregate the frequency of how often a patient occurs and select the five most similar patients for each query patient.

Implementation: We used publicly available MRI brain scans from TCIA, converted MRI from DICOM to NIFTI using the dicom2nifti library and defined the CAE in Keras. Latent space distances from the CAE were captured using the python *Annoy* library, which provides efficient approaches for calculating distances and searching. PSNR and SSIM were calculated using the functions available in scikit-learn.

Evaluation: To assess CAE performance, we calculated standard training metrics including loss and, once trained, we calculated PSNR and SSIM for a random subset of 10 images from the test set, and the same images with added Gaussian noise. To evaluate search we used the (unseen) set of test images, and classified them by tumour type (LGG vs. GBM), tumour laterality, and tumour location (anterior/posterior and medial/lateral) and asked experienced clinical staff for informal review of scan similarity.

III. Results

A. Training metrics

We measured standard deep-learning performance metrics such as model loss per epoch; as expected for CAE, this reduced very rapidly. The mean PSNR across the test images was 42.15 dB without any noise and 38.18 dB with Gaussian noise. SSIM scores were less good: 0.46 without any noise and 0.39 with Gaussian noise [17].

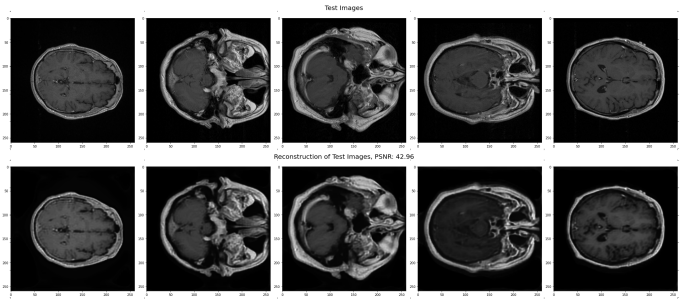

Fig. 2. Example PSNR without noise for five images.

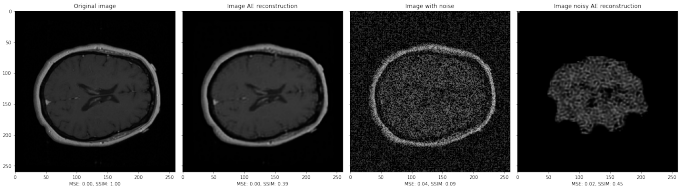

Fig. 3. Example SSIM results, with and without noise, for a single image.

Image search: We evaluated the performance of our trained system on the unseen dataset of 140 images (7 scans), split 2/5 between LGG and GBM. We return the 5 most similar patients for each query, and so assess our results across a total of 35 outputs. We assessed correctness by considering whether the tumour was of the same type, laterality and position anterior/posterior and lateral/medial. Results are shown below.

Tumor	N	Type	Laterality	AntPost	MedLat
GBM	25	88%	72%	65%	68%
LGG	10	60%	70%	60%	50%

TABLE I
TEST RESULTS FROM 7 QUERY PATIENTS

For the GBM patients, 88% of the 25 matched outputs were also GBM patients, but only 65% of the of the matched patients had a tumour in a similar location in the anterior/posterior classification. We show 5 query slices from a patient below, with the five most similar slices retrieved.

IV. Discussion

We have introduced, designed, implemented and tested a system that combines a convolutional autoencoder and near-neighbour clustering to provide instance-based querying of brain MRI scans in brain tumour patients. The autoencoder

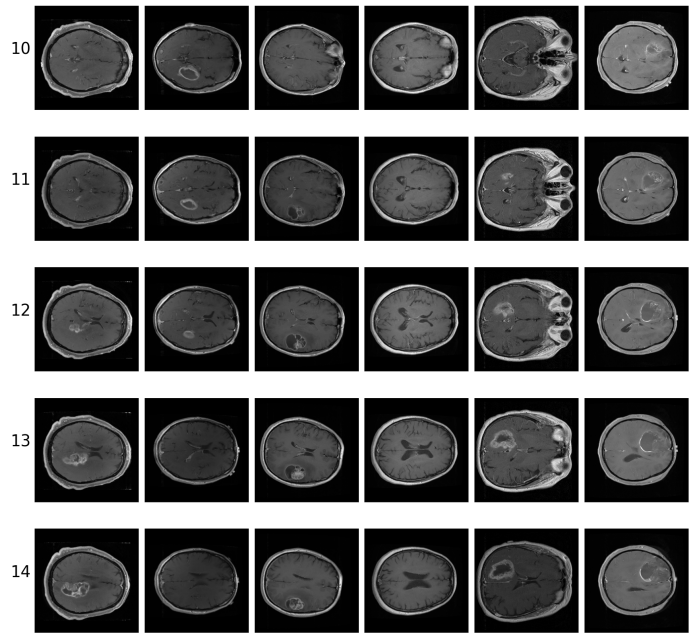

Fig. 4. The left most scans are of the input patient and then the scans from left to right are positioned based on how similar they are (the scans of each patient are in their own column). This example only shows part of the result with five scans (their indexes are visible on the left) of each the patients.

architecture performs reasonably well, as assessed by PSNR and SSIM and is reasonably resistant to noise injection. The combined system reliably returns scans that are similar to the instance used as a query, in terms of tumour type (LGG/GBM), location, laterality and size.

Standard measures of performance suggest that our CAE performs reasonably well: PSNR scores are in-line with results in the literature, although SSIM are lower than previous work and both PSNR and SSIM are reasonably resistant to noise injection. SSIM grades the actual structure of an image based on luminance, contrast, and structure rather than calculating the pixel by pixel squared difference (as for PSNR). Furthermore, although PSNR and SSIM measure CAE performance, we are interested in the latent-space induced by the bottleneck layer in the CAE, rather than performance.

Assessing performance of imaging-based technology outside of segmentation-based approaches is challenging: in our work, there is no widely-accepted metric that we can use. We have used a combination of 'computational' and 'clinical' metrics to try and capture how well our approach performs, but we accept that these are relatively informal approaches, and better metrics would enable more progress.

Failures to retrieve closely-matched images were informally evaluated, and were largely due to the lack of patients with similar brain tumours in our database. However, nearest neighbour searches, are approximate, and more detailed exploration of metric distances, the size of the retrieval set for each query, and optimisation of the underlying search structures are areas for further optimisation.

We would stress that this work is meant as a pilot, and we have deliberately chosen two types of tumour, with very different visual characteristics. As well as optimising metrics and data structures, we will expand this work to include a wider set of tumour types, and a larger, more varied dataset. However, unlike segmentation-based approaches, this work is relatively scalable, and expanding our work to use a larger dataset will be relatively easy.

REFERENCES

[1] Bondy, Melissa L et al. "Brain tumor epidemiology: consensus from the Brain Tumor Epidemiology Consortium." Cancer vol. 113,7 Suppl (2008): 1953-68. doi:10.1002/cncr.23741
[2] Sharma N, Aggarwal LM. Automated medical image segmentation techniques. J Med Phys. 2010;35(1):3-14.
[3] Ng, Andrew. "Sparse autoencoder." CS294A Lecture notes 72.2011 (2011): 1-19.
[4] Masci, J., Meier, U., Cireşan, D., Schmidhuber, J. (2011). Stacked Convolutional autoencoders for Hierarchical Feature Extraction. In: Honkela, T., Duch, W., Girolami, M., Kaski, S. (eds) Artificial Neural Networks and Machine Learning – ICANN 2011. ICANN 2011. Lecture Notes in Computer Science, vol 6791. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21735-7_7
[5] Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, April 2004, doi: 10.1109/TIP.2003.819861.
[6] Flynn, Jeremy & Ward, Steve & Abich IV, Julian & Poole, David. (2013). Image Quality Assessment Using the SSIM and the Just Noticeable Difference Paradigm. 8019. 10.1007/978-3-642-39360-0_3.
[7] Bondy, Melissa L et al. "Brain tumor epidemiology: consensus from the Brain Tumor Epidemiology Consortium." Cancer vol. 113,7 Suppl (2008): 1953-68. doi:10.1002/cncr.23741
[8] Verma, J., Nath, M., Tripathi, P. et al. Analysis and identification of kidney stone using Kth nearest neighbour (KNN) and support vector machine (SVM) classification techniques. Pattern Recognit. Image Anal. 27, 574–580 (2017). https://doi.org/10.1134/S1054661817030294
[9] Chatap, Niranjan J. and S. Shibu. "Analysis of blood samples for counting leukemia cells using Support vector machine and nearest neighbour." IOSR Journal of Computer Engineering 16 (2014): 79-87.
[10] N. H. Rajini and R. Bhavani, "Classification of MRI brain images using k-nearest neighbor and artificial neural network," 2011 International Conference on Recent Trends in Information Technology (ICRTIT), 2011, pp. 563-568, doi: 10.1109/ICRTIT.2011.5972341.
[11] L. Gondara, "Medical Image Denoising Using Convolutional Denoising Autoencoders," 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), 2016, pp. 241-246, doi: 10.1109/ICDMW.2016.0041.
[12] H. -C. Shin, M. R. Orton, D. J. Collins, S. J. Doran and M. O. Leach, "Stacked Autoencoders for Unsupervised Feature Learning and Multiple Organ Detection in a Pilot Study Using 4D Patient Data," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1930-1943, Aug. 2013, doi: 10.1109/TPAMI.2012.277.
[13] https://wiki.cancerimagingarchive.net/display/Public/TCGA-LGG
[14] https://wiki.cancerimagingarchive.net/display/Public/TCGA-GBM
[15] https://gitlab.com/computational.oncology/similarbrain
[16] https://github.com/AdrianArnaiz/Brain-MRI-Autoencoder
[17] A. Horé and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," 2010 20th International Conference on Pattern Recognition, 2010, pp. 2366-2369, doi: 10.1109/ICPR.2010.579.