Deep Learning-based Identification of Ischemic Brain Regions in Patients with Acute Ischemic Stroke

Undergraduate Thesis

Tianxi Hu
Supervisor: Professor Maged Goubran
Division of Engineering Science
Univerisity of Toronto

Contents

List of Figures

Acknowledgements

My deepest gratitude to my supervisor Dr. Maged Goubran, for this opportunity and his invaluable guidance, knowledge and insights. To Lyndon Boone, for his unwavering support and encouragement throughout this journey.

To Yuxin Lin for being on my side throughout all these years and Rui Ren for all the late-night conversations, although we are oceans apart. To Tia Fang for the endless hours spent together in the library. To Patrick Ly for his profound belief in me.

Lastly, to my family for their love, trust and motivation, as well as my cat, who offers moral support by sitting on my keyboard.

Abstract

Acute ischemic stroke (AIS) is marked by a reduced blood supply to brain tissues, leading to rapid cell death in minutes. Endovascular treatment (EVT) decisions for AIS patients rely on computed tomography perfusion (CTP) which identifies the two ischemic regions: core (irreversibly damaged tissue) and penumbra (potentially salvageable tissue). However, many challenges such as limited availability and delayed treatment time hinder the use of CTP in acute stroke treatment. This study aims to investigate the utility of Deep Learning (DL) in developing an automated method that produces core and penumbra estimation from widely available non-contrast CT (NCCT) and multiphase CT angiography (mCTA) sources. The dataset for this study is collected from the Sunnybrook stroke clinic and consists of 1002 subjects, with only 125 having non-empty RAPID-generated segmentation labels. U-Net baseline and UNETR models are trained on the labelled dataset for each of core and penumbra. The result shows good correlation with the labels (Dice = 0.321 for core and 0.459 for penumbra). In addition, self-supervised learning is used in modal pre-training to utilize the unlabelled data, which demonstrated great potential to increase the prediction accuracy.

Introduction

Stroke is the second main cause of death and disability worldwide, with ischemic stroke being the most common form accounting for more than 87% of all cases. Acute ischemic stroke (AIS) is marked by a reduced blood supply preventing brain tissues from getting oxygen and nutrients, leading to rapid cell death in minutes. Endovascular treatment (EVT) has been shown to significantly reduce disability and improve recovery compared to conventional therapy for AIS patients [1]. However, only 27% of patients who are eligible receive EVT. Additionally, each delayed 30 minutes in EVT decreases favorable outcomes by 11% [2]. Therefore, systems for automatic and rapid detection and treatment of AIS are critical to promote the chance of receiving appropriate treatment and reduce mortality for patients.

Current Computed Tomography (CT) imaging modalities used in the evaluation of AIS patients include non-contrast CT (NCCT), multiphase CT angiography (mCTA) and CT perfusion (CTP) [3]. In particular, patient selection and treatment decisions for EVT rely on CTP which identifies the two ischemic regions: core (irreversibly damaged tissue) and penumbra (potentially salvageable ischemic tissue). However, many factors challenge the use of CTP in acute stroke treatment: the insufficient implementation of perfusion imaging in primary stroke centers, incomplete standardization of image processing and the lack of expertise in image interpretation [4].

Under the supervision of Dr. Maged Goubran at the Sunnybrook Research Institute, we will be investigating the utility of Deep Learning (DL) in the prediction of brain tissue perfusion and infarction. The aim is to develop a DL-based automated method that produces core and penumbra segmentation from NCCT and mCTA sources alone, eliminating the need for CTP imaging. This is greatly valuable as it could improve AIS patient outcomes by providing critical perfusion information with substantially reduced time-to-treatment and need for resources. In addition, we are also interested in implementing methods in uncertainty estimation and model interpretability in order to study what information is used in our model to make predictions.

Background

1 Evaluation of Ischemic Core and Penumbra/Prognosis

The volumes in the ischemic core and penumbra are of great significance for the prognosis and outcomes of AIS patients. Both the core and the penumbra are irregular in shape due to a number of physiological factors. Their manual segmentation is time consuming and labor intensive, with inconsistency across raters [2]. Therefore, it is challenging for radiologists to manually annotate the lesion area at the pixel level.

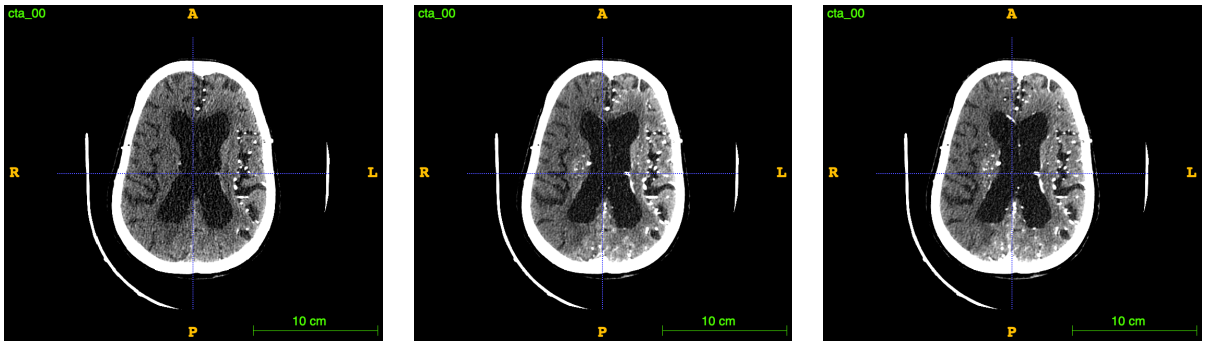
The parameter maps calculated from CTP images can be used to segment the ischemic core and deficit regions. Several commercial software, such as RAPID and F-stroke, have established automatic methods to use parameter threshold to predict core infarct area and ischemic penumbra. According to the definition by RAPID [5], a reduction in cerebral blood flow (CBF) to below 30% of normal brain tissue is associated with ischemic core [6] while the prolonged time-to-maximum (Tmax) of the tissue residue function to more than 6 seconds is associated with penumbra [7].

2 Alternative to CTP Imaging

State-of-the-art treatment selection centers around CTP, which identifies the two infarct regions: core and penumbra. However, CTP is not widely available in all stroke centers and could delay treatment due to its long scanning and postprocessing time. Due to those limitations, researchers have explored alternatives to CTP imaging in acute stroke treatment.

A study by Nguyen et al.[8] evaluated clinical outcomes in patients with proximal anterior circulation occlusion stroke selected for mechanical thrombectomy and concluded that there were no significant differences in patients selected with NCCT/CTA compared with those selected with CTP. These findings show that there is potential for the development of a simpler and more widespread stroke imaging paradigm without the use of CTP. In addition, studies have demonstrated that deep learning has great potential for automatically extracting the lesion features at the pixel level and segmenting core and penumbra volumes from NCCT/CTA sources alone. Wang et al.[9] worked on identifying acute ischemic volumes from NCCT and CTA and showed promising results with strong correlations observed with RAPID segmentations.

Our research has the same objective, but we wish to additionally leverage information



(a) CTA0 (b) CTA1 (c) CTA2

Figure 1: Three phases of mCTA. Note the high amount of contrast agent shown in white in the right hemisphere, indicating occlusions due to AIS in the left hemisphere.

on the differences between the left and right hemispheres observed in the CTA images of AIS patients. An example is provided in Figure 1, where the contrast agent injected shown in white is concentrated in the right hemisphere, indicating occlusions due to AIS in the left hemisphere.

A number of research have studied this asymmetry by either producing an differential map between the two hemispheres, or using them as separate inputs for a model to learn and use information on their differences. Work by Herzog et al. [10] on deep learning for diagnosis of early and progressive dementia used segmented images of brain asymmetry. A computer vision approach was used for the detection and segmentation of brain image asymmetries. After an image is skull-stripped and binarized, its centroid was calculated to centre the image and a segmentation of asymmetry was produced by reflecting and subtracting one side from the other. However, the algorithm was only tested on 2D slices of MRI images so it may not be as applicable to our work. Barman et al.[11] proposed a convolutional neural network design for automated detection of ischemic stroke from CTA images. The proposed model is sensitive to changes in symmetry of vascular and brain tissue texture which allows it to detect ischemic stroke and produces a classification for stroke/no stroke.

3 Self-Supervised Learning

Deep learning has boosted medical image analysis over the past years. Training a deep learning model with high performance requires a large amount of labeled data. However, medical imaging is a field where labeling data requires expert knowledge, and collecting large labeled datasets is challenging. As a result, it is often difficult to obtain

a sufficient number of labeled images for training, and unlabeled datasets are usually more common and accessible [12]. Therefore, boosting the performance of deep learning models by using unlabeled as well as labeled data becomes an important but challenging problem.

Self-supervised learning (SSL) presents one possible solution to this problem. It offers a way to lower the need for manually annotated data by pre-training models for a specific domain on unlabelled data. In this approach, labeled data are solely required to fine-tune models for downstream tasks. As a result, such algorithms have potential for substantial improvements in the field of medical image analysis. There are three common problems in medical imaging: classification, localization, and segmentation. For our project, we will apply existing SSL pre-training methods with the unlabelled data to increase the efficiency of the downstream image segmentation task.

A number of studies have proposed self-supervised learning strategies for medical images. Doersch et al. [13] studied the utility of predicting the relative positions of image patches, where a visual representation is learned by the task of predicting the position of an image patch relative to another. Chen et al. [14] proposed a method based on context restoration to better exploit unlabelled images, which improved the previous methods by learning features useful for different types of subsequent image analysis tasks. The effectiveness of this strategy has been verified in three common problems in medical imaging: classification, localization, and segmentation. In all three cases, self-supervised learning based on context restoration learns meaningful semantic features and leads to improved machine learning models for the above tasks.

4 Model Interpretability and Explainability

Systems based on deep learning have no inherent way of representing the uncertainty associated with a model's prediction and do not provide information on what input features influence a particular prediction. They are often referred to as "black boxes" as there is a lack of theoretical understanding on the underlying mechanics. [15].

However, interpretability and explainability are crucial for the safe and ethical use of AI in healthcare. Regulations such as the 2018 European General Data Protection Regulation have made it more difficult to use "black-box" models in healthcare since the decision-making process must be traceable [16]. As a result, deep learning-based models need to not only achieve a high precision in order to be helpful in a clinical setting. In

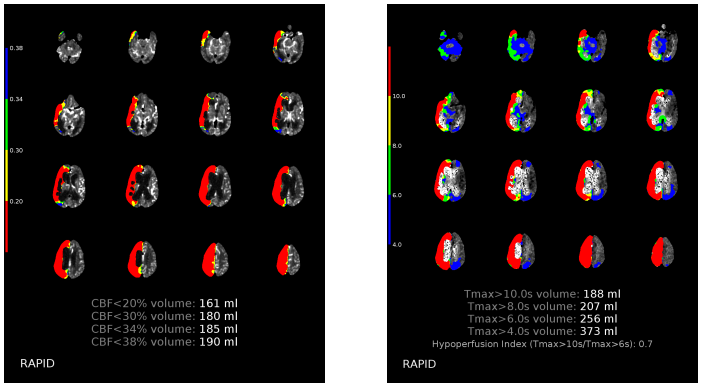
addition, interpretability and uncertainty in predictions must be well understood [17].

Methods

Data Preprocessing

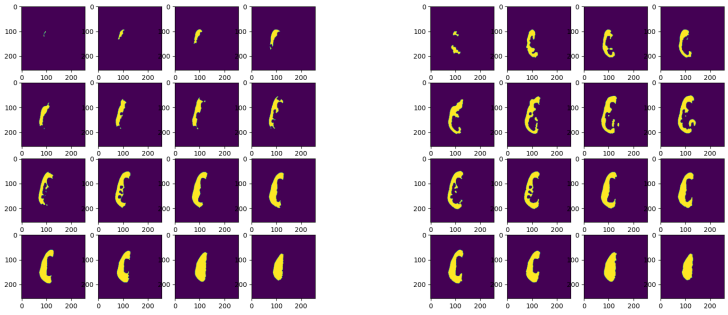
A dataset of ∼6000 ischemic stroke patient CT scans is collected from the Sunnybrook stroke clinic, among which 1002 subjects have been already processed to include NCCT, three phases of CTA, CTP and RAPID-generated perfusion paramter maps.

Ground truth labels for core and penumbra are extracted from perfusion parameter maps produced by RAPID with respective thresholds of CBF $\leq 30\%$ and $\geq$ greater than 6 seconds as shown in Figure 2 and Figure 3.



(a) CBF (b) Tmax

Figure 2: Rapid-based CTP segmentations for different CBF and Tmax thresholds. A patient with a high HIR=0.7 is shown with expected rapid ischemic core growth.



(a) Core with CBF $\leq 30\%$ (b) Penumbra with Tmax ≥ 6

Figure 3: Core and penumbra segmentation for subject shown in Figure 2.

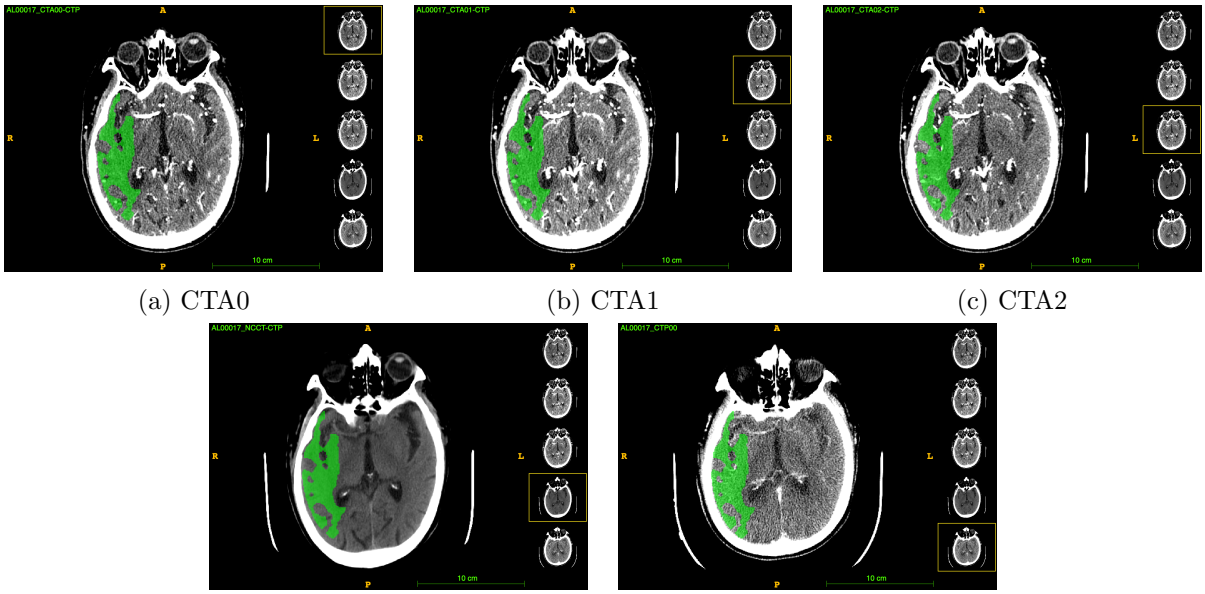
Registration

A preprocessing procedure is established in order to align CT images from different modalities for a deep learning model [18]. Specifically, image registration is needed which involves applying spatial transformation between structures within an image to match the data [19]. We choose to register both NCCT and CTA to CTP with the first perfusion image as the fixed reference.

Prior to registration, the reference CTP images are upsampled with linear interpolation to increase the z-resolution to the same as CTA. The upsampled images are then zero-padded in the z-axis to the same size as the first CTA image. NCCT is registered directly to CTP. A two-step method is used for registration of CTA to CTP. First, CTA is registered to NCCT through the same affine registration as NCCT to CTP. The transformation matrix is then combined with the NCCT to CTP transformation matrix and applied on CTA to map it to CTP's space. As the labels are in the same space as CTP, there is no need for registration and they are upsampled with nearest neighbor interpolation to increase the z-resolution to the same as CTA.

As images are registered with-in subjects, tissue deformation can be ignored. As a result, a registration method based on rigid transformation (rotation + translation) is a good choice for our task. The Similarity transformation implemented in ANTsPy [20] with mutual information metric is chosen as the registration method, which offers six degrees of freedom with a combination of scaling and rigid transformation.

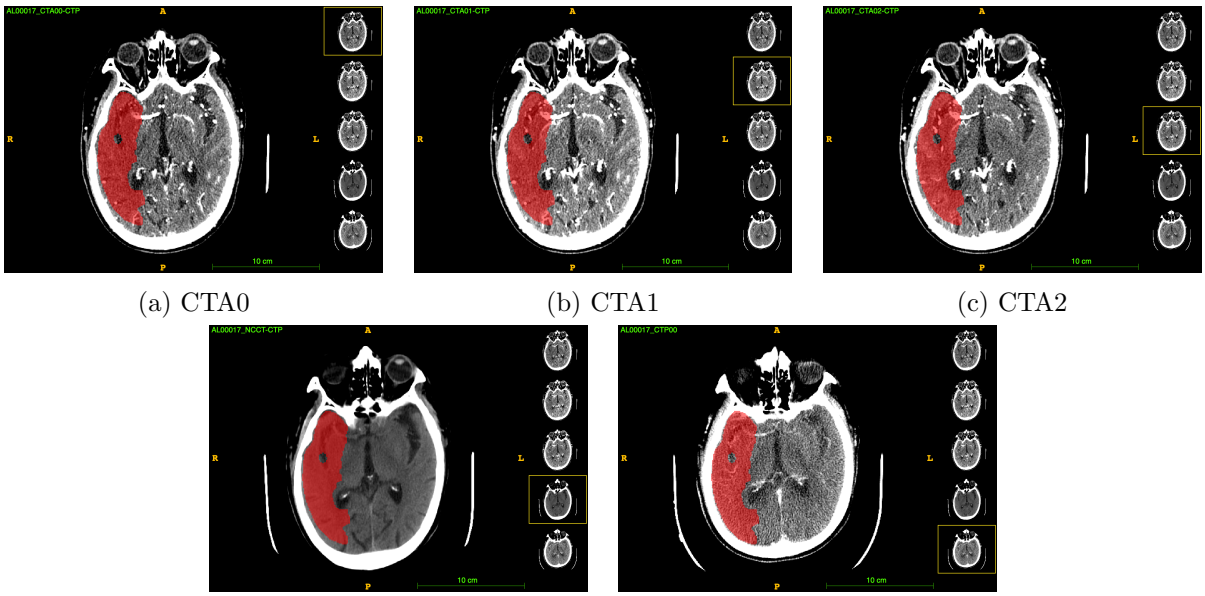
After registration, CTA and NCCT are cropped to the same dimension and field of view as the upsampled CTP to ensure their alignment with the segmentation labels. The final processed data for each subject consists of four 512×512×128 images to be used as model inputs, three for mCTA (CTA0, CTA1, CTA2) and one for NCCT, and two 512×512×128 labels for core and penumbra. Figure 4 and 5 visualize the input and label for one subject. The first image of the perfusion sequence, CTP0, is also included as reference image for registration.



(a) CTA0 (b) CTA1 (c) CTA2

(d) NCCT (e) CTP0

Figure 4: Axial slices at (256, 256, 64) of registered CT image sequence for subject AL00017 with core region ground truth label shown in green.



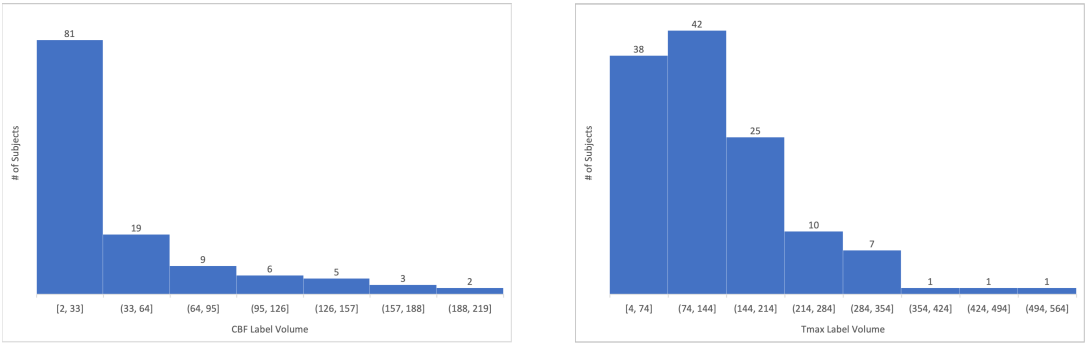
(a) CTA0 (b) CTA1 (c) CTA2

(d) NCCT (e) CTP0

Figure 5: Axial slice at (256, 256, 64) of registered CT sequence for subject AL00017 with penumbra region ground truth label shown in red.

Data Selection

To curate a labelled dataset, processed subjects are selected based on the condition that the subjects has a complete mCTA, CTP and NCCT sequence (3, 24, 2 images for each respectively) and both core and penumbra labels are non-empty. The resulting dataset $S_{labelled}$ has n=125 subjects in total, divided into train/validation sets with 80%/20% split. The distribution of label volumes for core and penumbra of the labelled input subjects is shown in Figure 6. Note that penumbra on average has larger volumes than core, and the distribution is right-skewed meaning that the majority of non-empty labels have small volumes.



(a) Core (b) Penumbra

Figure 6: Distribution of label volume for core (left) and penumbra (right) of the labelled input subjects.

The remaining subjects has at least one empty label and are gathered into an unlabelled dataset $S_{unlabelled}$ to be used for SSL pre-training. Due to the space constraint on the processing server, the unlabelled dataset only has n=95 subjects, divided into train/validation sets with 80%/20% split as well.

1 Deep Learning Pipeline

For the purpose of efficient and robust training, we constructed a modular deep learning pipeline with the following components:

- **Data Loader**: The labelled dataset (n=125) is divided into training/validation sets with a split of 80%/20%. Intensity normalization is applied channel-wise. Data augmentation through the application of various random transformations (i.e., flipping, zooming, adjusting contrast) is used to increase the size of the train-

ing dataset. Such techniques have been demonstrated to boost performance in deep learning image classification by up to 7% [21].

- **Model Training and Evaluation**: Our pipeline is modularized in a way that we can test a variety of deep learning architectures without changing the infrastructure. Models with selected data and network architecture will be trained and evaluated against spatial overlap-based metric dice coefficient and spatial distance-based metrics 95% Hausdorff Distance (HD) to determine their performance.

2 Model Architecture

U-Net

The U-Net architecture [22] which has consistently demonstrated strong results in medical image segmentation tasks is used as the baseline for our task.

Its architecture, as shown in Figure 7, is an encoder network followed by a decoder network. The encoder shown in the left half in is classification network where convolution blocks followed by a maxpool downsampling encode the input image into feature representations at multiple different levels.

The decoder is the right half of the architecture. The goal is to semantically project the discriminative features (lower resolution) learnt by the encoder onto the pixel space (higher resolution) to get a dense classification. The decoder consists of upsampling and concatenation with the feature map at the same level in encoder network followed by regular convolution operations.

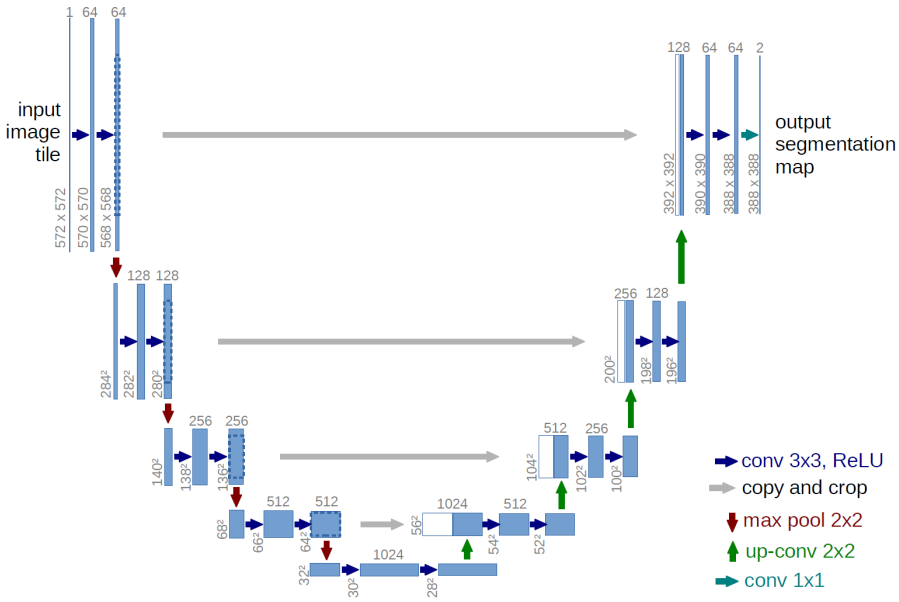


Figure 7: Architecture of U-Net [22].

UNETR

Introduced by Hatamizadeh et al. [23], UNETR, although having a similar name to the U-Net, is a Transformer-based architecture with 3D Vision Transformer (**ViT**) as the backbone. The architecture is shown in Figure 8.

Transformer [24] is a network based on attention mechanisms [25], which calculate attention weights for each pixel in the image based on its relationship with all other pixels. ViT implements multi-head attention, which extends this mechanism by allowing the model to attend to different parts of the input sequence simultaneously. UNETR uses a pure transformer as the encoder to learn sequence representations of the input volume, in this case a sequence of input patches with their positional information, effectively capturing the global multi-scale information. It shares some structural similarity with the U-Net, where features from multiple resolutions of the encoder are merged with the decoder via skip connections to compute the final segmentation output.

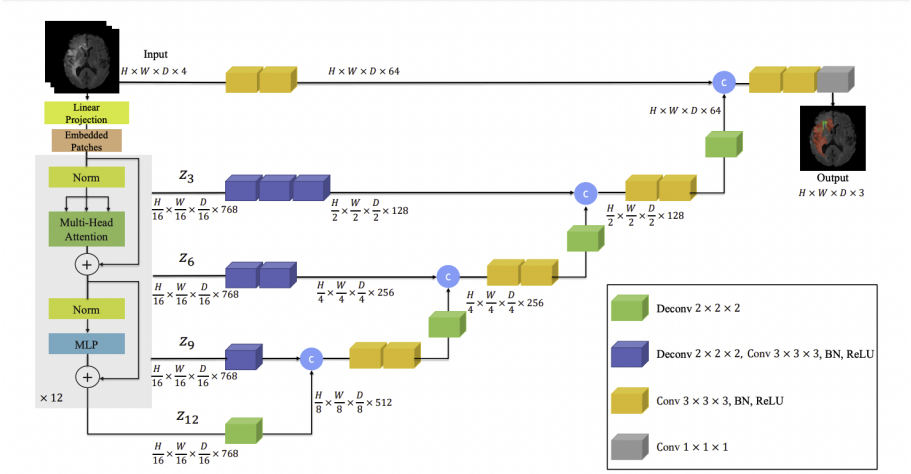


Figure 8: Architecture of UNETR [23].

SSL framework

The implementation of our SSL framework is based on MONAI [26]. The framework has two parts: we first pretrain the model on a self-supervised task, then we transfer the model embeddings and perform supervised learning on the downstream task, based on the idea that pretraining the representations on similar data can help the model recognize more comprehensive features. The overall workflow is summarized in Figure 9 and Algorithm 1.

Algorithm 1: SSL pre-training framework for segmentation tasks using transformer models.

Input: unlabelled SSL data $S_{unlabelled}$, labelled downstream data $S_{labelled}$

Result: transformer model ω^* that produces segmentation masks given input images

$\omega \leftarrow$initialize transformer model;

$S_{aug} \leftarrow$addAugmentation$(X_{unlabelled})$; `/* create augmented dataset */`

$\omega_{unlabelled} \leftarrow$self-supervised reconstruction$(\omega; S_{unlabelled}, S_{aug})$;

$\omega^* \leftarrow$transfer$(\omega_{unlabelled})$; `/* copy backbone weights */`

$\omega^* \leftarrow$train$(\omega^*, S_{labelled})$; `/* train on downstream segmentation task */`

return ω^*

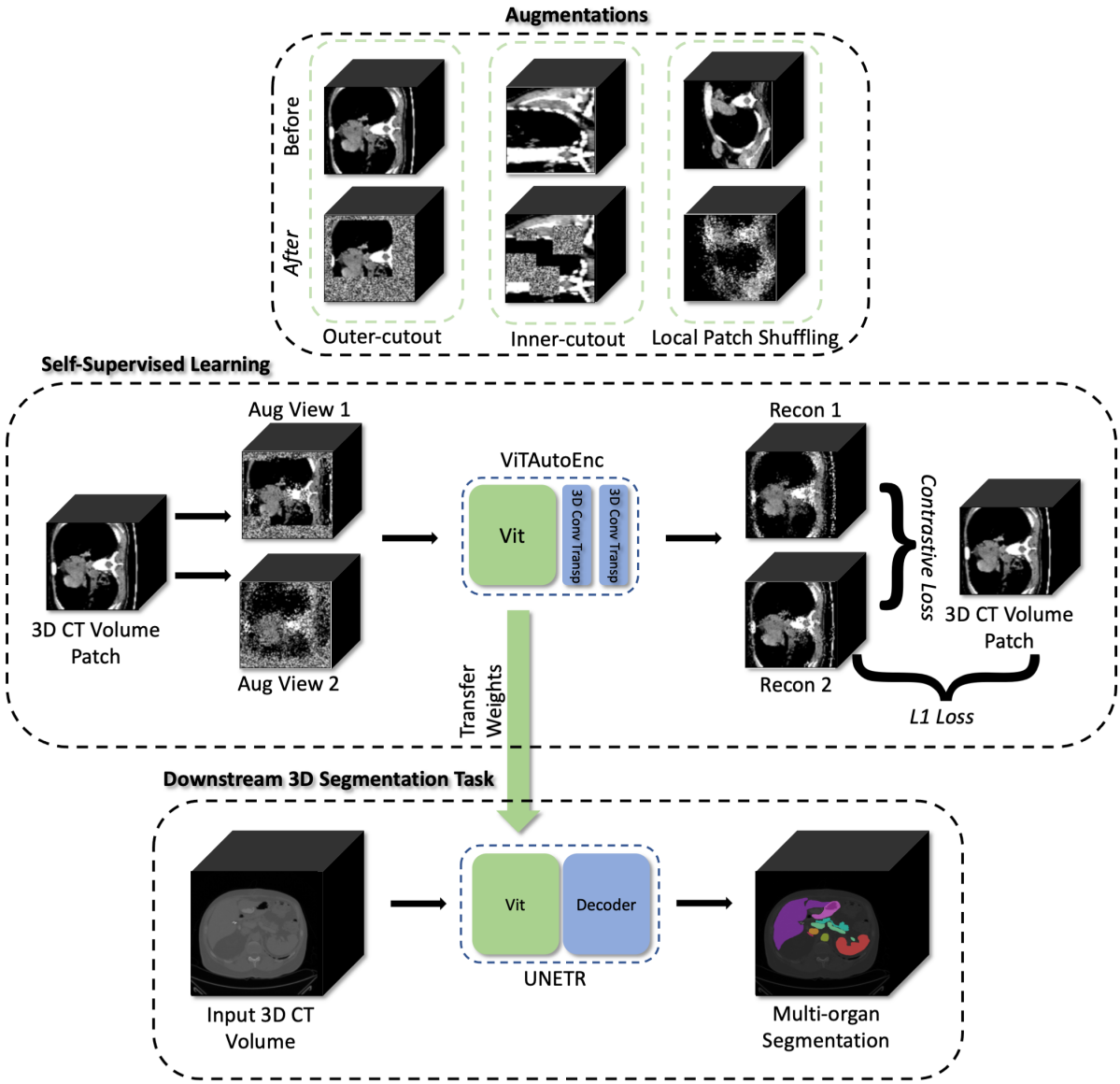


Figure 9: Workflow of the SSL framework with sample inputs from [26].

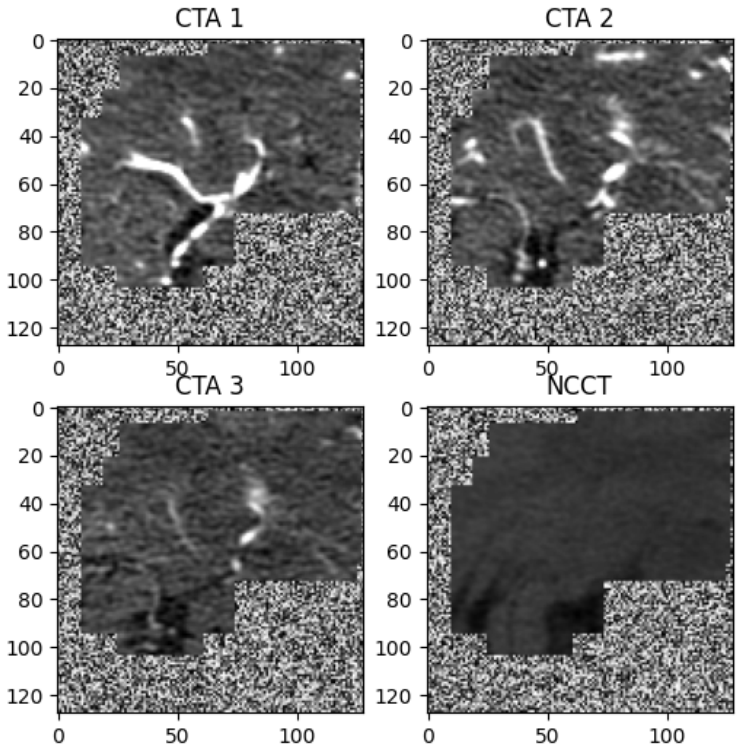


Figure 10: Sample input patches with noise augmentation to the ViT model.

Self-Supervised Learning The input images are augmented with operations such as in-painting [27], out-painting [27] and noise augmentation to the image by local pixel shuffling [14]. Sample input slices from our unlabelled dataset with noise addition are shown in Figure 10. The network then works to simultaneously reconstruct the two augmented views as similar to each other as possible via regularized contrastive loss [28] as its objective is to maximize the agreement.

Weight Transfer We transfer the weights from ViT to UNETR for further training on the downstream segmentation task. We only keep items of the ViT weights if they are part of the ViT backbone that is used in UNETR, and exclude the weights that are only used for formatting dimensions.

3 Training and Evaluation

Experimental Set-up

All models are trained on Compute Canada allocations with one v100l gpu and 6 cpu. One self-supervised ViT model is trained over 200 epochs with validation per 2 epochs and the following hyperparmeters:

- Learning Rate: 1e-4
- Batch Size: 4
- ROI Size: (128, 128, 48), total 8 3D volume (2 samples with ROI size were drawn per image)
- Loss Function: L1 Contrastive Loss
- Optimizer: Adam
- Temperature: 0.005
- Total Model Parameters: 98546961

The weights from the self-supervised model are then transferred to train the downstream UNETR models. A number of baseline U-Net and UNETR models are also trained with random weight initialization. The following hyperparmeters are used for the fully-supervised models with validation per 2 epochs:

- Learning Rate: 1e-4
- Batch Size: 2
- ROI Size: (128, 128, 48), total 8 3D volume (4 samples with ROI size were drawn per image)
- Loss Function: DiceCELoss
- Optimizer: AdamW
- Weight Decay: 1e-5
- U-Net Total Model Parameters: 1980906
- UNETR Total Model Parameters: 102203938

Loss Function

For the supervised models, an equally weighted sum of Cross Entropy Loss and Dice Loss, **DiceCEloss**, is used as the loss function. A lower DiceCELoss value indicates better performance and sharper decision boundary.

Validation Metric

The primary evaluation metric used is the Dice coefficient, which is also the most used metric in validating medical volume segmentation. It computes the ratio between the intersection and the union of prediction and ground truth segmentation as shown in Equation 1. A higher dice is better as it indicates a higher similarity.

$$Dice = 2 \times \frac{|y_{true} \cap y_{pred}|}{|y_{true}| + |y_{pred}|} \tag{1}$$

Hausdorff Distance is widely used as a dissimilarity measure in the evaluation of image segmentation, with a lower value indicating a higher similarity. It is sensitive to outliers so we use the distance at 95 percentile to mitigate that. Other measures such as a confusion matrix are also included for a more holistic evaluation [29].

Results

1 Fully-Supervised Models

Two fully-supervised U-Net and UNETR models are trained for each of core and penumbra prediction. The validation mean dice results from the four models as well as a UNETR model with supervised pre-training is summarized in Table 1 and Figure 11.

Table 1: Summary of best validation mean Dice for trained models over 50 epochs.

Model		
Name	Description	Best Dice
$UNet_{CBF}$	Random weight initialization for core prediction	**0.1394**
$UNet_{Tmax}$	Random weight initialization for penumbra prediction	**0.2671**
$UNETR_{CBF}$	Random weight initialization for core prediction	**0.2304**
$UNETR_{Tmax}$	Random weight initialization for penumbra prediction	**0.3946**
$UNETR_{Tmax}$	Initial weights transferred from ViT for penumbra prediction	**0.4050**

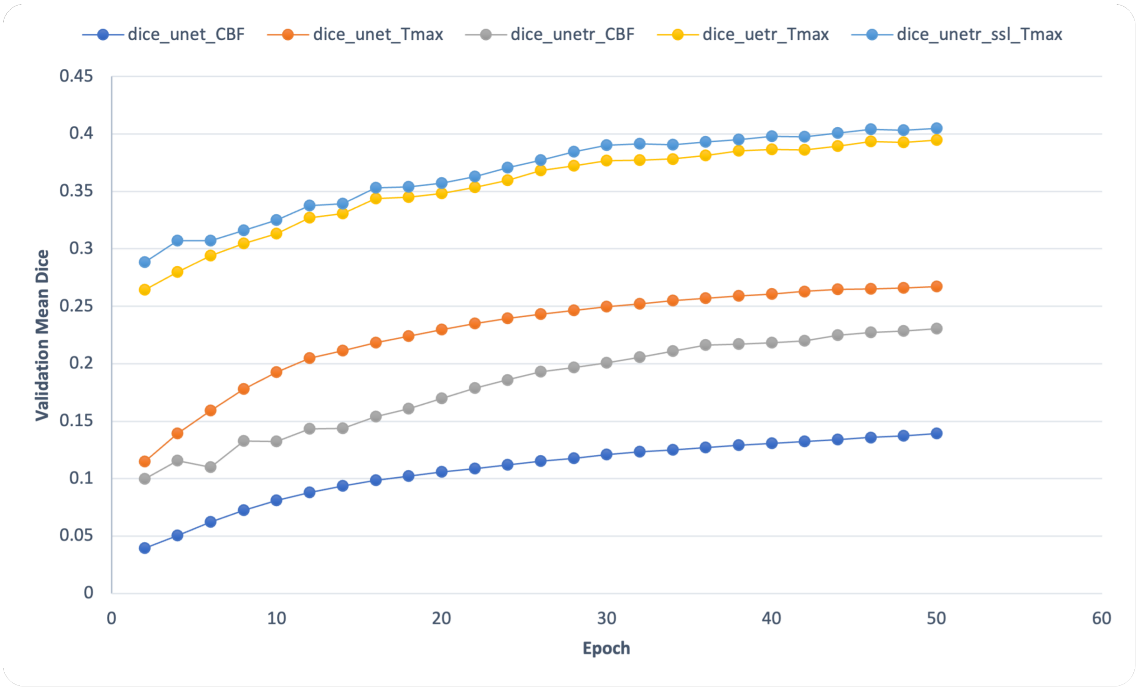


Figure 11: Summary of fully-supervised model training results over 50 epochs.

Note that the UNETR models outperform U-Net in both tasks. The advantage likely comes from UNETR's attention mechanism, which allows the model to focus on relevant features and ignore irrelevant ones. It can be particularly useful for this segmentation task, where the model needs to identify and segment region of interest from a complex background. In addition, the Transformer architecture is designed to better capture long-range dependencies in input sequences. By combining this with the U-Net architecture's ability to extract low-level features, UNETR can potentially extract more meaningful features from the input image.

Two best performing models, UNETRs for core and penumbra, are trained for longer

and achieved best dice = 0.3210 at epoch 320 and best dice = 0.4590 at epoch 176 respectively. However, when examining the validation subjects individually, it can be noted that the dice score fluctuates a lot by subjects.

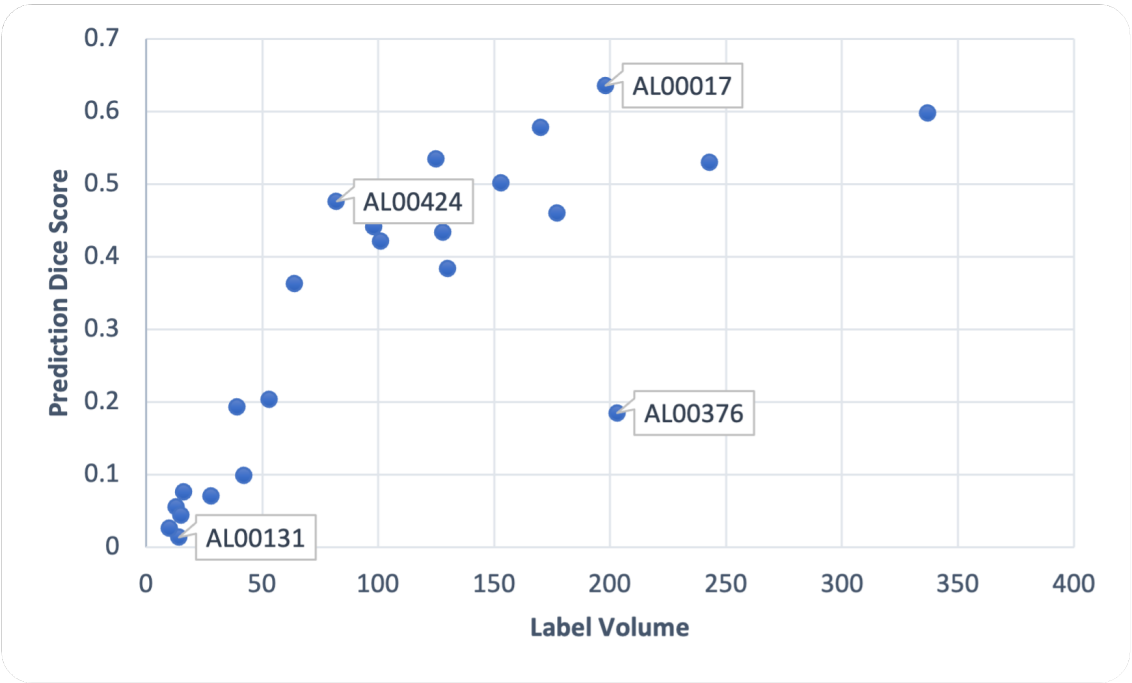
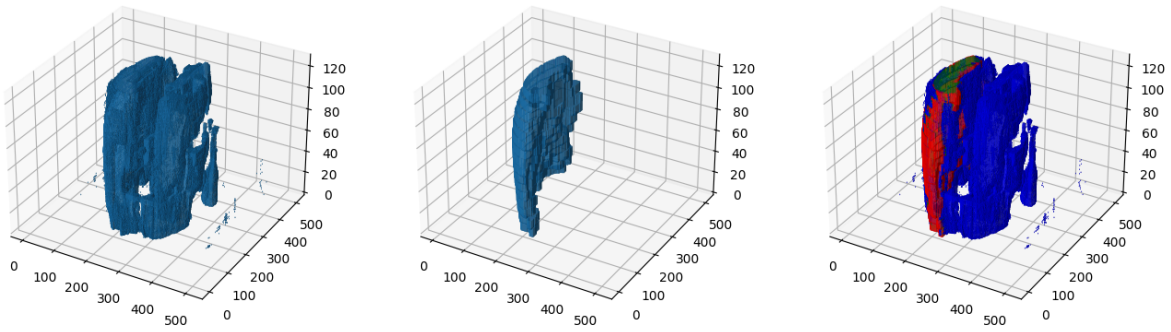


Figure 12: Dice vs penumbra volume by baseline U-Net model (avg dice = 0.3305).

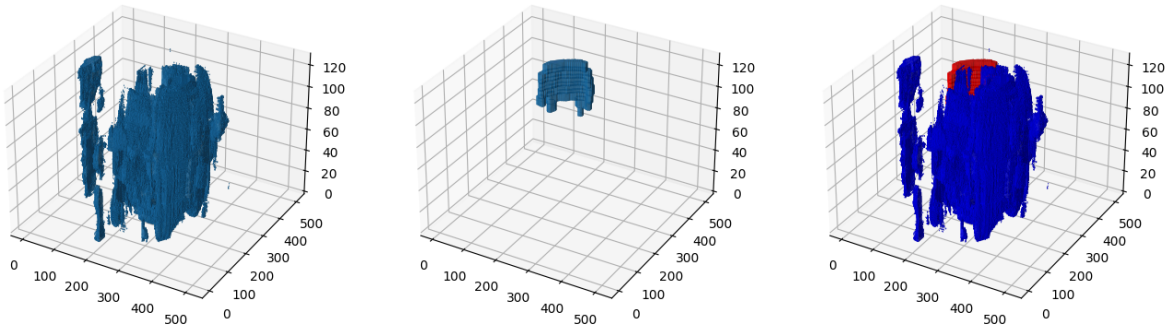
A plot of Dice vs. penumbra volume for the 25 subjects in validation set is shown in Figure 12, which shows that the higher the label volume, the higher the resulting dice score tends to be. In particular, the four subjects labelled on the plot are worth examining closer.

AL00017 achieved the highest Dice = 0.64 with a large label volume and **AL00131** achieved lowest Dice = 0.01 with a small label volume. The truth and prediction for these two subjects are visualized in 3D in Figure 13 and 14. In the comparison plot for these two figures, true positives (overlap between truth and prediction) are shown in green, false positives in blue and false negatives in red. Both predictions have a significant more number of false positives than false negatives, which lowers the dice score further for the subject with smaller label volume.



(a) Prediction (b) Truth (c) Comparison

Figure 13: Subject AL00017: large label volume with high dice = 0.64.



(a) Prediction (b) Truth (c) Comparison

Figure 14: Subject AL00131: small label volume with low dice = 0.01.

Although larger label volumes usually corresponds to a higher accuracy in prediction on the validation set, there are two outliers. **AL00424** has Dice = 0.19 with a relatively large label volume while **AL00376** has Dice = 0.48 with a smaller label volume.

2 Self-Supervised Models

Figure 15 shows one comparison of the validation loss and metrics over 100 epoch for UNETR penumbra models with random weight initialization and loaded pre-trained weights. Although the dataset used for pre-training only has a small size (n=95), models with pre-trained weights are shown to have consistently achieved higher validation Dice and lower validation loss. As a result, with an increased number of unlabelled subjects for pre-training, SSL has substantial potential for further improving our models.

(a) Dice

(b) DiceCELoss

Figure 15: Comparison of UNETR models for penumbra with and without pre-trained weights over 100 epochs.

Conclusion

In conclusion, we investigated the utility of deep learning in predicting core and penumbra volumes from mCTA and NCCT sources and achieved good results for subjects with large label volumes. However, our models are prone to predicting a high number of false positives for subjects with small label volumes. We also observed that self-supervised learning in modal pre-training using unlabelled data has shown potential in increasing the prediction accuracy.

Although the initial results are encouraging, there is still great room for improvement. Additional CT scans have been acquired from the Sunnybrook Stroke Clinic, which can be processed to increase the size of both our labelled and unlabelled dataset. Additional self-supervised tasks for pre-training can also be explored. One example is proposed by Taleb et al. [30] which extends the patch-based approaches in our current SSL framework to solve Jigsaw puzzles on multi-modal images. After a desired training accuracy is reached, we will also incorporate the use of uncertainty estimation and model interpretability techniques such as Monte Carlo Guided Backpropagation [17] to visualize the importance of different input features for prediction.

Bibliography

[1] M. Goyal, B. K. Menon, W. H. v. Zwam, *et al.*, "Endovascular thrombectomy after large-vessel ischaemic stroke: A meta-analysis of individual patient data from five randomised trials," *The Lancet*, vol. 387, no. 10029, pp. 1723–1731, 2016, ISSN: 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(16)00163-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S014067361600163X.

[2] L. Cui, Z. Fan, Y. Yang, *et al.*, "Deep Learning in Ischemic Stroke Imaging Analysis: A Comprehensive Review," *BioMed Research International*, vol. 2022, p. 2 456 550, Nov. 2022, ISSN: 2314-6133. DOI: 10.1155/2022/2456550. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9678444/ (visited on 01/23/2023).

[3] D. Birenbaum, L. W. Bancroft, and G. J. Felsberg, "Imaging in Acute Stroke," *Western Journal of Emergency Medicine*, vol. 12, no. 1, pp. 67–76, Feb. 2011, ISSN: 1936-900X. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3088377/ (visited on 10/14/2022).

[4] J. Demeestere, A. Wouters, S. Christensen, R. Lemmens, and M. G. Lansberg, "Review of Perfusion Imaging in Acute Ischemic Stroke," *Stroke*, vol. 51, no. 3, pp. 1017–1024, Mar. 2020, Publisher: American Heart Association. DOI: 10.1161/STROKEAHA.119.028337. [Online]. Available: https://www.ahajournals.org/doi/full/10.1161/STROKEAHA.119.028337 (visited on 10/14/2022).

[5] i. Inc, *Aneurysm, pulmonary embolism and stroke software platform powered by AI*, en. [Online]. Available: https://www.rapidai.com (visited on 01/31/2023).

[6] B. C. V. Campbell, S. Christensen, C. R. Levi, *et al.*, "Cerebral blood flow is the optimal CT perfusion parameter for assessing infarct core," eng, *Stroke*, vol. 42, no. 12, pp. 3435–3440, Dec. 2011, ISSN: 1524-4628. DOI: 10.1161/STROKEAHA.111.618355.

[7] J.-M. Olivot, M. Mlynash, V. N. Thijs, *et al.*, "Optimal Tmax threshold for predicting penumbral tissue in acute stroke," eng, *Stroke*, vol. 40, no. 2, pp. 469–475, Feb. 2009, ISSN: 1524-4628. DOI: 10.1161/STROKEAHA.108.526954.

[8] T. N. Nguyen, M. Abdalkader, S. Nagel, *et al.*, "Noncontrast Computed Tomography vs Computed Tomography Perfusion or Magnetic Resonance Imaging Selection in Late Presentation of Stroke With Large-Vessel Occlusion," *JAMA Neurology*, vol. 79, no. 1, pp. 22–31, Jan. 2022, ISSN: 2168-6149. DOI: 10.1001/jamaneurol.2021.4082. [Online]. Available: https://doi.org/10.1001/jamaneurol.2021.4082 (visited on 01/30/2023).

[9] C. Wang, Z. Shi, M. Yang, *et al.*, "Deep learning-based identification of acute ischemic core and deficit from non-contrast CT and CTA," *Journal of Cerebral Blood Flow & Metabolism*, vol. 41, no. 11, pp. 3028–3038, Nov. 2021, ISSN: 0271-678X. DOI: 10.1177/0271678X211023660. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8756471/ (visited on 01/26/2023).

[10] N. J. Herzog and G. D. Magoulas, "Deep Learning of Brain Asymmetry Images and Transfer Learning for Early Diagnosis of Dementia," en, in *Proceedings of the 22nd Engineering Applications of Neural Networks Conference*, L. Iliadis, J. Macintyre, C. Jayne, and E. Pimenidis, Eds., ser. Proceedings of the International Neural Networks Society, Cham: Springer International Publishing, 2021, pp. 57–70, ISBN: 978-3-030-80568-5. DOI: 10.1007/978-3-030-80568-5_5.

[11] A. Barman, M. E. Inam, S. Lee, S. Savitz, S. Sheth, and L. Giancardo, *Determining Ischemic Stroke From CT-Angiography Imaging Using Symmetry-Sensitive Convolutional Networks*. Apr. 2019, Pages: 1877. DOI: 10.1109/ISBI.2019.8759475.

[12] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "Self-Supervised Feature Learning for Medical Image Analysis," en,

[13] C. Doersch, A. Gupta, and A. A. Efros, "Unsupervised Visual Representation Learning by Context Prediction," in *2015 IEEE International Conference on Computer Vision (ICCV)*, ISSN: 2380-7504, Dec. 2015, pp. 1422–1430. DOI: 10.1109/ICCV.2015.167.

[14] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "Self-supervised learning for medical image analysis using image context restoration," eng, *Medical Image Analysis*, vol. 58, p. 101 539, Dec. 2019, ISSN: 1361-8423. DOI: 10.1016/j.media.2019.101539.

[15] A. Singh, S. Sengupta, and V. Lakshminarayanan, "Explainable Deep Learning Models in Medical Image Analysis," en, *Journal of Imaging*, vol. 6, no. 6, p. 52, Jun. 2020, Number: 6 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2313-433X. DOI: 10.3390/jimaging6060052. [Online]. Available: https://www.mdpi.com/2313-433X/6/6/52 (visited on 01/24/2023).

[16] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, *What do we need to build explainable AI systems for the medical domain?* arXiv:1712.09923 [cs, stat], Dec. 2017. DOI: 10.48550/arXiv.1712.09923. [Online]. Available: http://arxiv.org/abs/1712.09923 (visited on 01/26/2023).

[17] K. Wickstrøm, M. Kampffmeyer, and R. Jenssen, "Uncertainty and interpretability in convolutional neural networks for semantic segmentation of colorectal polyps," en, *Medical Image Analysis*, vol. 60, p. 101 619, Feb. 2020, ISSN: 1361-8415. DOI: 10.1016/j.media.2019.101619. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841519301574 (visited on 01/26/2023).

[18] S. Masoudi, S. A. A. Harmon, S. Mehralivand, *et al.*, "Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research," *Journal of Medical Imaging*, vol. 8, no. 1, p. 010 901, Jan. 2021, Publisher: SPIE, ISSN: 2329-4302, 2329-4310. DOI: 10.1117/1.JMI.8.1.010901. [Online]. Available: https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/volume-8/issue-1/010901/Quick-guide-on-radiology-image-pre-processing-for-deep-learning/10.1117/1.JMI.8.1.010901.full (visited on 10/14/2022).

[19] C. R. Maurer, "A Review of Medical Image Registration," en,

[20] *ANTs by stnava*. [Online]. Available: http://stnava.github.io/ANTs/ (visited on 01/30/2023).

[21] L. Perez and J. Wang, *The Effectiveness of Data Augmentation in Image Classification using Deep Learning*, arXiv:1712.04621 [cs], Dec. 2017. DOI: 10.48550/arXiv.1712.04621. [Online]. Available: http://arxiv.org/abs/1712.04621 (visited on 10/14/2022).

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," en, May 2015. DOI: 10.48550/arXiv.1505.04597. [Online]. Available: https://arxiv.org/abs/1505.04597v1 (visited on 01/31/2023).

[23] A. Hatamizadeh, Y. Tang, V. Nath, *et al.*, "Unetr: Transformers for 3d medical image segmentation," 2021. arXiv: 2103.10504 [eess.IV].

[24] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021. arXiv: 2010.11929 [cs.CV].

[25] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," 2017. arXiv: 1706.03762 [cs.CL].

[26] *Tutorials/ssl_finetune.ipynb at main · Project-MONAI/tutorials*, en. [Online]. Available: https://github.com/Project-MONAI/tutorials (visited on 04/26/2023).

[27] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ISSN: 1063-6919, Jun. 2016, pp. 2536–2544. DOI: 10.1109/CVPR.2016.278.

[28] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," en, in *Proceedings of the 37th International Conference on Machine Learning*, ISSN: 2640-3498, PMLR, Nov. 2020, pp. 1597–1607. [Online]. Available: https://proceedings.mlr.press/v119/chen20j.html (visited on 04/26/2023).

[29] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, no. 1, p. 29, Aug. 2015, ISSN: 1471-2342. DOI: 10.1186/s12880-015-0068-x. [Online]. Available: https://doi.org/10.1186/s12880-015-0068-x (visited on 01/31/2023).

[30] A. Taleb, C. Lippert, T. Klein, and M. Nabi, "Self-supervised Learning for Medical Images by Solving Multimodal Jigsaw Puzzles," en, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 2020.