UNIVERSITY OF AUCKLAND

Faculty of Engineering

Auckland Bioengineering Institute

Master's Degree Thesis

Stroke Lesion Segmentation and Functional MRI for Predicting Recovery

Supervisor:
Prof. Alan Wang
Prof. Justin Fernandez
Prof. Vickie Shim

Student:
Zion Lee

March 2024

Contents

List of Figures

List of Tables

1 Abstract

Stroke is a leading cause of death and long-term disability worldwide, which fosters the need for effective methods of diagnosis with lesion segmentation and prediction of recovery. Hence, this thesis presents a comprehensive approach that integrates deep learning techniques for stroke lesion segmentation from functional magnetic resonance imaging (fMRI) data, coupled with machine learning methods such as Random Forest (RF) and Support Vector Machine (SVM) for predicting recovery outcomes post-stroke.

The first phase of the thesis focuses on developing a reliable stroke lesion segmentation model comparable to the state-of-the-art deep learning architectures using a large public dataset. Specifically, families of Unets, derived from convolutional neural networks (CNNs) are employed to accurately and reliably delineate stroke lesions. These range from transfer learning of existing theoretical concepts such as Linear kernel attention (LKA), building model using a state-of-the-art method like transformers and dense blocks to attempting sophisticated methods such as diffusion. The adoption of enhanced ensemble method also contributed to generating a high performing segmented label with significantly high level of accuracy. Hence, our final segmentation model achieved an accuracy of 0.6513 as well as sufficient reliability in delineating stroke lesions and was comparable to the state-of-the-art method.

In the second phase, machine learning models, including random forest (RF) and Support vector machines (SVM), are employed to predict recovery outcomes following stroke based on extracted imaging features and clinical data from the Functional MRI data from two different sites. The delineated lesion using the segmentation model developed in the first phase as well as additional clinical variables to forecast functional recovery were used to predict the recovery of the patients, and highlighted the importance of lesion volume and its position relative to the somatosensory region of the brain for achieving high accuracy.

Overall, this thesis contributes to the advancement of stroke management by providing reliable tools for lesion segmentation and recovery prediction. The integrated approach combining deep learning with machine learning techniques offers a comprehensive framework for understanding stroke pathology and guiding clinical decision-making, ultimately improving patient care and outcomes in the post-stroke rehabilitation process.

2 Acknowledgements

I wish to offer gratitude to my fellow researchers in my group Triana and Mishaim for helping out with brain atlas, Grace and Jeremiah for helping out with their knowledge in deep learning segmentations in stroke, and Benjamin for sharing his knowledge in fMRI.

I also wish to acknowledge Parvez Ahmad for teaching me wholeheartedly with basics of machine learning and deep learning for medical images when I was lost at the start of the year and throughout.

Last but not least, I wish to express massive thanks to Dr.Alan Wang who was my supervisor. He was very helpful and approachable to ask any questions raised.

I was delighted to have done a project alongside them and wish to work with them again in the future should the opportunity come.

3 Introduction

3.1 What is stroke

A stroke occurs when the blood supply to the brain is disrupted, either by a blood clot, known as ischemic stroke, or by continuous bleeding in the brain, known as hemorrhagic stroke [3]. This disruption can cause brain cells to die which leads to a range of physical and cognitive symptoms. Stroke is well known for a leading cause of long-term disability in adults; and only a small number of patients achieve a full recovery sufficient for daily activities whereas more than 25% of patients need some assistance and about 50% experience long-term dependency [4].

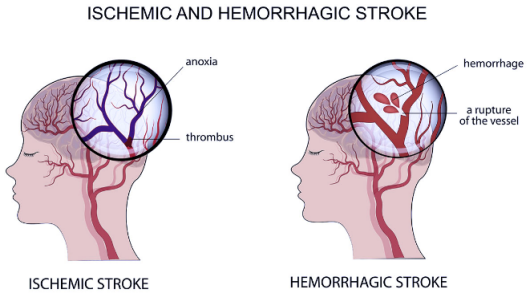


Figure 1: Detailed picture of ischemic and hemorrhagic stroke

Due to its severity and fatality, diagnosis of stroke at the earlier stage is important for faster and promoted recovery. Neuroimaging techniques such as CT or MRI are widely used in the modern world to accurately capture stroke lesions. Especially for MRI, an acronym for magnetic resonance imaging, it has been a state-of-the-art approach in delivering high-resolution images after its invention in 1974 by Paul Lauterbur and Peter Mansfield [5]. MRIs are found to be a much more valuable imaging technique in stroke lesion detection for their well-known sensitivity in detecting morphologically variant conditions such as stroke by providing better contrast for soft tissues and anatomical details than other modalities [6]. Next, it is up to the clinicians and neurosurgeons to decide whether any abnormalities that appear on the neuroimage obtained from MRI are related to the lesion of infarct after stroke. This process is time-consuming and expensive, as well as the possibility of human errors.

To assist in the decision-making process of these healthcare workers, machine learning techniques to automatically detect stroke lesions have recently become noticeable. Machine learning is a subset of artificial intelligence (AI) that focuses on creating models that enable computers to learn and make predictions based on the data given [7]. In the medical field, it is widely used for automatic detection and stroke is undoubtedly one of them. This process of creating models from an image of input is known as computer vision, with the use of deep learning, again a subset of machine learning that focuses on

using artificial neural networks to learn from large amounts of data. It involves training complex models, known as deep neural networks to perform tasks such as classification and segmentation. Deep learning with neural networks is a major approach in computer vision for stroke lesion segmentation and some commonly used yet powerful deep learning techniques/architectures in stroke lesion segmentation will be introduced below.

3.2 Deep learning architectures

3.2.1 U-net structures

In recent years, CNN (convolutional neural network) based on U-net architecture have been widely applied in biomedical image segmentation and have shown improved accuracy. U-net is a popular technique in semantic segmentation with U-shaped architecture which incorporates symmetric contracting (encoder) path and expanding (decoder) path with skip connections between the two paths to enhance detail retention. It has outperformed many visual recognition tasks since its introduction to the community and biomedical image segmentation was no exception [8] . Thus far, many extensions of U-nets including U-net++ and 3D-U-net have been widely developed and applied to stroke image segmentation, and undoubtedly many state-of-the-art architectures employed U-net architecture as a backbone to their proposed architectures, settling in as a clear trend in stroke lesion segmentation [9] [10].

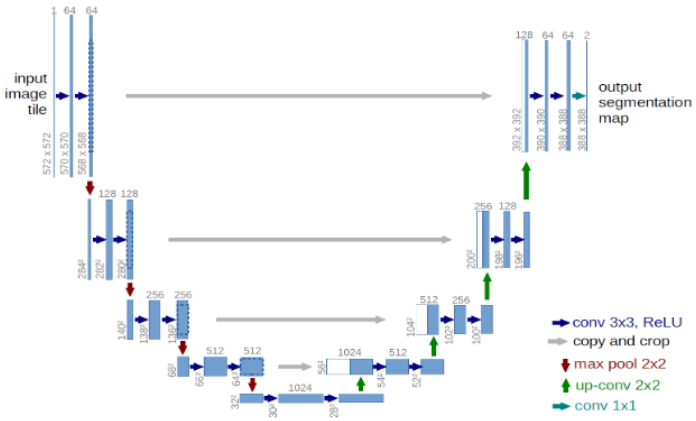


Figure 2: Structure of a U net showing encoder (left side) and the decoder (right side)

3.2.2 Transformers

In addition to Unet, transformer based techiniques have been utilized in the field of medical segmentation and titled as state-of-the-art method. Transformers have recently gained popularity in computer vision tasks with its success in natural language processing (NLP) sector [11]. Unlike traditional convolutional neural networks (CNNs), vision transformers (ViTs) rely on self-attention mechanisms to capture long-range dependencies between im-

age patches allowing them to handle large-scale images with high accuracy and efficiency [12]. The basic architecture of a ViT consists of an input patch embedding layer, a transformer encoder, and a fully connected output layer [13]. The transformer encoder consists of multiple layers of self-attention and feedforward neural networks, which process the patch embeddings to extract features and learn contextual relationships between patches. One advantage of ViTs is that they can be pre-trained on large-scale image datasets using self-supervised learning, which involves training the model to predict missing parts of an image. This pre-training process allows the model to learn generic image features that can be fine-tuned for specific downstream tasks. In terms of stroke lesion segmentation, transformers have been added to the encoder side of the backbone U-net structure to improve the dice coefficient score of the segmentation results. Some of the recent transformer infused U-net structures introduced to stroke lesion segmentation will be outlined below.

3.2.3 GAN

To endeavor further into various deep learning methods such as unsupervised learning, vast amount of studies have been done to developing models such as generative Adversarial Networks (GANs). GANs have shown great potential in medical image analysis, including stroke lesion segmentation. GAN-based techniques work by training a generator network to produce realistic images, while simultaneously training a discriminator network to differentiate between real and generated images. In stroke lesion segmentation, the generator network can be trained to produce segmentations of stroke lesions based on input medical images, while the discriminator network can assess the quality of the generated segmentations. This process can be iteratively improved until the generated segmentations are indistinguishable from real segmentations. GAN-based techniques have shown promising results in stroke lesion segmentation, achieving high accuracy and reducing the need for manual annotation.

3.2.4 Clustering

If we get into more traditional method to segment the regions of interest, clustering method can be highlighted. Clustering is a technique in machine learning and data analysis that involves grouping similar objects or data points into clusters based on their attributes or characteristics [14]. Clustering methods aim to find patterns or structures in the data that can be used for tasks such as segmentation, classification, or anomaly detection. There are various types of clustering algorithms applied to stroke lesion segmentation, including partition-based clustering (such as K-Means), Fuzzy C-means clustering and hierarchical clustering [15] [16].

3.3 Functional MRI

Due to the need of studies regarding the activation linking to the functionality of the brain, functional MRI techinique have emerged in the medical field. Functional MRI is a profound and widely used method in predicting severity and recovery of various brain related symptoms such as Alzheimer's disease, Parkinson's disease, Schizoprenia and stroke [17][18] . It is well known for brain activity modelling and functional connectivity analysis to provide a milestone for a prolonged prediction of outlined brain symptoms [19][20]. Functional MRI utilises a technique known as blood oxygen level dependent (BOLD) contrast to construct a feature brain map that capture neural activity related to blood flow, driven by the oxygen consumption of individual brain cells [21]. As MRI is a radiation free technique, functional MRI is well-known for minimally invasive functional image acquisition method. Its stability and capability in delivering high resolution images both spatially and temporally, have leveraged rapid growth of Functional MRI utilization in the field of cognitive neuroscience [22].

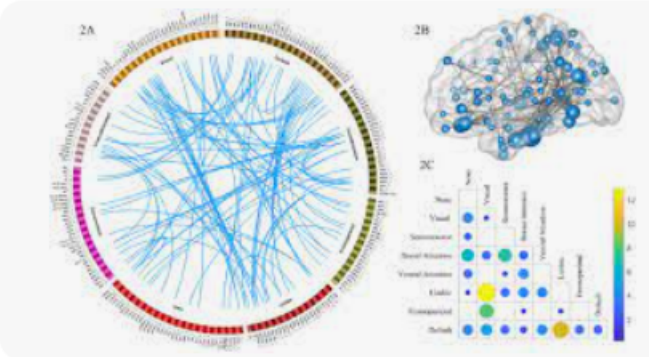


Figure 3: Detailed picture of ischemic and hemorrhagic stroke

Functional connectivity is an interaction between neural activities under time series and unlike anatomical connectivity, it is possible to take place between pairs of anatomically unconnected regions [23]. The functional connectivity data can be pulled out for further analysis such as seed-to-voxel-based or region-to-region (ROI).

3.4 Main contributions

This paper is structured as follows:

- Provide a comprehensive literature review on the recent trend in deep learning models in medical diagnosis field
- Engage on extensive data analayis and deep learning model building for stroke lesion segmentation

- Focus on functional MRI data analysis and feature extraction to create a machine learning based prediction of stroke recovery
- Conclude by suggesting future improvements and succeeding research questions to be attempted

4 Literature Review

4.1 Stroke lesion segmentation

Understanding the fast paced progress in the field of deep learning and machine learning, only papers published from 2021 to 2023 have gone through a literature review to capture the most recent trends in stroke lesion segmentation. A total of 40 papers have been found and analyzed under the keywords; stroke lesion segmentation, stroke prediction, medical image segmentation, and biomedical machine learning

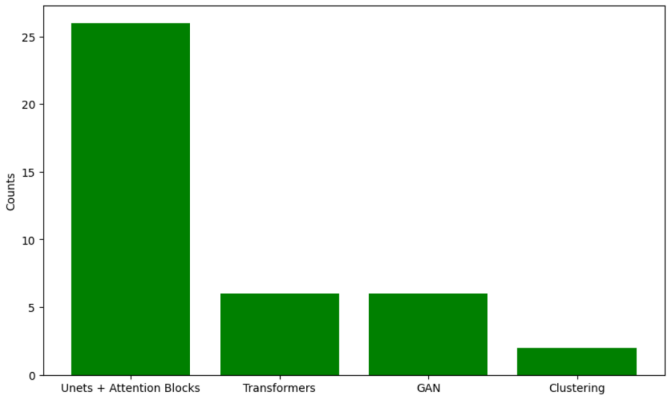


Figure 4: Number of segmentation architecture introduced in this literature review

4.1.1 U-net based architectures

4.1.1.1 Non-3D Unets

Reducing inter-site discrepancy within the dataset is important for a reliable segmentation. Yu et al. created a self-adaptive normalizing network by adopting a 2D U-net architecture with a site-classifier in the bottleneck [24]. The author additionally utililzed the fact that stoke occurs unilaterally in most cases and inspired by pseudosymmetry scheme of the human brain, flipped the lesions to one side of the hemisphere to produce a smaller image for training. Ou et al. came up with a 2.5D network architecture specified for segmenting highly discontinuous imaging data such as DWIs, extending U-Net by replacing traditional convolutional layers with proposed Lambda+ layers to capture both inter-slice and intra-slice contextual information [25].

4.1.1.2 3D U-nets

Soltanpour et al. made use of adopting MultiresUnet as a backbone that CNN layers that consists of three different filter sizes with 3×3, 5×5, and 7×7 convolution operations instead of adopting only 3×3 convolution throughout to enhance the capability of

capturing lesions with different sizes and shpaes at varying resolutions [26] [27]. Tursynova et al. initially demonstrated the use of a classical shaped 3D U-net with 4 symmetric encoder and decoder blocks to test on the new set of computation tomography dataset, ISLES 2018 [28]. Zhao et al. suggested a multifeatured map fusion network (MFMF-Network) where numerous weakly labelled subjects are used to train the classification branch, and several fully labelled subjects while freezing the parameters of backbone VGG 16 truncated model were used to train a modified segmentation branch and generate class activation mapping (CAM) for post processing [29]. Omarov et al. created a modified U-Net architecture that some convolution layers were replaced with dilated filter convolution layers which expands the field of the view of each filter, allowing the model to include more contextual information of the background [30]. Leaky ReLu activations for all feature maps has also been utilized for calculating convolutions across the network [10]. Lin et al. created an R2U-RNet architecture adopting R2U-Net as a backbone with a proposed residual refinement unit (RRU) where each input image containing two image channels are generated for inference at multiple scale [31]. Random label flipping technique was applied as part of a noisy-label training scheme to reduce uncertainties from manual label annotations. Verma et al. suggested the first-time usage of ATLAS v2.0 dataset by adopting a deep 3D U-net model with 5 encoder blocks and 5 decoder blocks with average symmetric surface distance (ASSD) to measure contour distances between the two input segmentations [32]. Wong et al. proposed a model that outputs twice increased resolution from the decoder against the encoder input image by using Subpixel embedding that captures local structure through feature and spatial expansion [33]. Kumar et al. introduced an asymmetric residual network that aim to solve a prevalent vanishing gradient problem that degrades the learning rate of the network and a class imbalance problem that the stroke lesion part takes up only a small part of the entire brain region [34]. Wong et al. suggested a rotation reflection equivariant U-net inspired by a fact that the segmentation technique must be robust to rotation or reflection regardless of the input image [35]. Moon et al. adopted both 2D and 3D multi-modal U-nets employed to FLAIR and DWI pre-processed using ANTs (Advanced normalization tools) software and found that 2D multi-modal technique performed the better than other 3D or combined techniques [36]. Du et al. suggested a novel attention-guided multimodal recovery framework that reduces intraclass inconsistency where there exists a large variability between areas of the stroke lesion and interclass indistinction where stroke lesion resembles healthy brain tissue in appearance [37]. Bousselham et al. created a temperature profile for acute ischemic strokes to enhance segmentation accuracy incorporated with classical 3D U-net [38]. Lee et al. used a classifier to classify affected tissue against the normal tissue in a detailed fashion using patch-based attention module as ground truth segmentation only shows an inclusive block of the two [39]. Alquhayz et al. proposed a network separated into three parts; an initial classifier model trained to extract the slices of stroke lesions from a whole brain image, a secondary classifier that finds difference between patches that consist of stroke lesions, and finally a 2D modified U-net that that segments out the lesion patches extracted from the patch classifier [40]. Li et al. suggested a multiscale U-net model with dilation convolution that produces richer feature maps by combining low-level feature mapping with high-level feature mapping [41]. This network performed better in segmented tasks under identical network parameters. Li et al. proposed a novel target-aware supervision residual learning

framework (TSRL-Net) considering imbalance between positive and negative sample in which a creatively target-aware loss function is designed to dilate strong attention regions while paying higher attention to the positive sample losses and compensate for negative sample losses around the target region [42].

4.1.1.3 Attention blocks

To reduce class imbalance between the lesion and the healthy tissue, Abramova et al. incorporated a novel restrictive balanced sampling technique to reduce common class imbalance problem when using patch-based method as well as a squeeze-and-excitation blocks in the encoder part of the network to learn channel-wise feature dependencies by adaptively weighting the feature maps [43]. Similarly, Karthik et al. suggested a consecutive method using classifiers before the segmentor to classify lesion pixels against the healthy pixels in advance to reduce class imbalance. Additionally, the author proposed a multi residual attention block and a masked dropout layer to suppress unnecessary spatial activations from transmitting to the encoder blocks and to selectively regulate the receptive field awareness [44]. Shi et al. proposed a cross modal cross attention (C2MA) module in the bottleneck of the 3D U-net and non-local operations to reduce long-range dependencies analysed between different modalities by taking two of the encoded feature maps as an input in a parallelized manner [45]. To improve the blurriness in small lesion contours, Hui et al. proposed a dual-path attention compensation U-Net (DPAC-UNet) that consists of a primary path network, a core network that segments and outputs final segmentation annotations, and auxiliary path network which correct possible errors when auxiliary attention compensation coefficients are delivered to the primary path [46]. Both path networks are modified attention U-Net models and are structurally identical. Unlike most studies with attention blocks attached to the encoding path, Huang et al. proposed a network with attention blocks inserted to both sides. An Encoding Mutual gain adaptive similarity block (MGAS block) was designed to enhance the ability of the network to capture long-range spatial by extracting global contexts through feature similarity whereas a decoding global context awareness block (GCA block) was designed to supply global context for low-semantic features to promote restoration of localized pixels to distinguish weak boundaries [47]. Shin et al. proposed an efficient U-net structure which employ novel e-blocks inspired by depth-wise separable convolution and a global-feature attention block (GA-block) that improves segmentation performance by capturing receptive features of each pixel between the encoder and the decoder [48]. As medical image segmentation does not fully take structural symmetry information into account, Bao et al presents a novel mirror difference aware network (MDAN) that uses a differential feature augmentation (DFA) module in the encoding path to specify semantic pathological asymmetries and aim at reducing noise in these asymmetrical non-lesion areas [49]. Sheng et al. came up with a cross-attention and deep supervision U-net (CADs-Unet) that interactively shares local spatial information using cross-spatial attention module instead of common self-attention module to reduce the uncertainty in positional information of stroke lesion images [50].

4.1.1.4 Transformers

Feng et al. connected ViTs after convolutional blocks to create a novel convolution transformer block (CT block) which consists of two convolutional layers and a transformer module where convolutional layers extract local information to the transformer to utilize its global receptive field [51]. Gu et al. proposed using an existing Swin transformer module, an inherent transformer module that accepts shifted windows approach and hierarchical transformer characteristics, instead of ViTs to tackle a problem in which the memory use and computational power increases with respect to the image size in the ViTs module [52] [53]. Luo et al. connected ViTs to the encoder part of the U-net to improve the global context feature recognition as well as adding Multi head cross attention modules in the decoder part to attain highly precise spatial information recovery [54]. Unlike conventional architectures that fuse transformers into the encoder part of the network, Liu et al. proposed a long-range hybrid features network (LLRHNet) which adopts transformers in the bottleneck part of the network to learn the long-range pixel dependencies in an image from tokenized patches [55]. Wu et al. introduced a CNN-based method as the auxiliary generator and transformer-based method as the primary segmenter to build the backbone network where segmentor is divided into initial segment generator and final segmentation predictor by applying a transUnet architecture [56], [57]. Wu et al. added a subsampling interactive transformer (SiTR) with dimension attention mechanism to not only capture both global and local semantic features but also extract channel features between multiple dimensions [58].

4.1.1.5 GAN

Platscher et al. trained their model using various image to image translation models (ITM) such as pix2pix, SPADE and cycleGAN to synthesize magnetic resonance images to with and without stroke lesions from semantic feature maps and furthermore generate synthetic lesion masks [59]. By combining the feature maps and the generated lesion masks, authors created a large synthetic dataset of stroke images that improved the segmentation performance. Van Voorst et al. proposed a novel follow-up-to-baseline GAN architecture to convert follow-up scan with stroke lesion to a generated baseline scan without a lesion through creating a difference map subtracted from the follow-up scan, where the map was later used to automatically obtain lesion segmentation [60]. Wang et al. proposed a consistent perception GAN (CPGAN) that generates segmentation based only on small number of labelled samples assisted by using similarity connection module that capture multi-scale feature information of the images [61]. Ghnemat et al. suggested a mutation technique integrated with Euclidian distance map to serve as an input to GAN architecture using patches generated from these mutated images to generate a synthetic dataset enhancing the dice coefficient of the proposed GAN model by 2.54% [62]. Wang et al. aimed to accomplish CT samples generation and cross-modality translation differentiation by integrating a Siamesed auto-encoder architecture into GAN where a Gaussian mixture translation module (GM-Trm) is further proposed to deliver incorporated translation loss to learn an intrinsic mapping between the latent space and

the multi-modal translation function [63]. Chen et al. proposed a novel semi-supervised segmentation framework that combines improved mean teacher and adversarial network where a student model and a teacher model was used to segment the target and generate the signed distance maps of image surfaces, and a discriminator network was subsequently employed for extracting hierarchical features and distinguishing the signed distance maps [64].

4.1.1.6 Clustering

Isa et al proposed a pseudo-colour k means clustering algorithm which performs colourization on greyscale images to intensify the visual perception and deliver more information [65]. The algorithm then segments the region of interest (ROI) using K-means clustering. To tackle the problem of clustering steps entangled in the local optima and does not progress toward the global optima, Thiyagarajan et al. proposed an arithmetic optimisation based K-means (AOK-means) method where the number of clusters in MRI image segmentation is determined by identifying the break in point of the aggregated arithmetic mean of sub-images of the whole brain MRI slice, and the number of sub-images equals the number of clusters [66].

4.2 Discussion

4.2.1 Computational efficiency

In this review, few studies have outlined the importance of computational efficiency in stroke lesion segmentation. As we observe more and more studies to attempt cloud-based method for storing outputs for overall automatic medical image segmentation, reducing computational load is essential for decreasing the memory uptake of the segmentation system. One popular method involved using patch-based method where small and overlapping patch samples were produced during image translation process in medical imaging as processing convolutions on the entire image require large amount of memory. Multiple studies have significantly improved their efficiency with patch sampling. Others include reducing number of convolution layers within each block to create a lighter convolutional block or generating asymmetrical decoder with fewer number of filters than the encoder as varying the number of filters in the decoding path does not alter the network performance because it only decodes information per requirement. A growing interest in improving the computational efficiency only highlights its importance and due to recent introduction of memory-heavy techniques such as transformers or unsupervised learning with numerous synthetic datasets, reducing the memory usage while improving accuracy will be a key to better performance.

4.2.2 Supervised, unsupervised and semi-supervised method

Selection of learning method is of great importance in machine learning as it directly affects the overall performance of the model. In the review, all three types of learning methods have been identified with the supervised learning being selected the most as many papers involved the use of U-net architecture originated from CNN, the leading method in supervision learning in medical image segmentation. The introduction of transformer techniques to supervised learning further boosted the accuracy drastically. In recent years, unsupervised learning has emerged as a strong competitor to supervised method that long dominated the field of stroke lesion segmentation for its unnecessity to provide manual annotation. Obtaining manual annotation is time-consuming, costly and susceptible to lesion inconsistency from subjective differences among radiologists, and applying unsupervised learning vastly reduce these drawbacks. In the review, unsupervised learning, typically GAN, achieved notable dice score coefficient comparable to supervised learning on the public dataset [63]. Overall, we could not identify which learning method is the best as multiple evaluation criteria such as accuracy and efficiency must be considered. Specifically, in terms of the dice score coefficient metrics, supervised learning outperformed the others. However, since unsupervised and semi-supervised have only been introduced to this field recently and as these methods possess their own clear advantages, further studies would guarantee their success in stroke lesion segmentation.

4.2.3 Loss function

Loss functions play a critical role in machine learning, as the choice of loss function can significantly impact the model's ability to learn and generalize and serve as a primary measure of how well a model is performing on a given task. Majority of the studies in the review adopted either dice loss or binary cross entropy loss or addition of both, especially for CNN based architectures. In addition, we identified growing number of papers adopting multiple loss functions as their total loss function. For instance, binary cross entropy loss function performs well on balanced classes where the number of lesion pixels are roughly equal to the number of non-lesion pixels whereas dice loss function becomes particularly useful in imbalanced classes with smaller number of lesion pixels. By combining these two loss functions, one can target wider range of lesions with different shapes and sizes, particularly competent in stroke lesions well-known for its great variability in morphology. Furthermore, appropriate selection and application of loss functions to simultaneously achieve multiple objectives such as segmentation and classification significantly increased the performance, especially in architectures with multiple consecutive paths or in GAN. For instance, using voxelwise absolute difference between the generated and real images (L1-loss), and binary cross entropy loss in the discriminator for classifying these images simultaneously have notably improved the dice coefficient score against only using L1-loss [60]. Since applying multiple loss functions often lead to overfitting and reduction of computational efficiency, the choice of loss functions and its number must be based on the type of segmentation performed, and carefully evaluated through various validation protocols to ensure optimal performance and generalization.

4.3 Objective and purpose of this project

This research aims to assist neurosurgeons in improving their way of diagnosing stroke by creating a helpful tool for delivering accurate yet reliable results by utilizing machine learning and deep learning techniques. Our objective leads to a clear purpose:

- To create a high-performing stroke lesion segmentation model comparable to the state-of-the-art method using a deep learning technique
- Use automatically segmented lesions obtained from our segmentation model alongside the Functional MRI results to predict the recovery of the patients using machine learning

5 Stroke lesion segmentation

5.1 ATLAS 2.0 Datset

The selection of proper data is an important milestone in achieving desired performance and output in deep learning. In the field of stroke segmentation, there is an innate difficulty in obtaining a large number of datasets due to the high cost of running MRI trials and hiring neuroradiologists to delineate stroke lesions from the brain, and the need for a large and robust dataset is highlighted to improve the reliability in segmentation performance.

ATLAS (Anatomical Tracings of Lesions After Stroke) v2.0, an updated version from the previous v1.2, is a set of publicly available 655 T1w MRI images obtained from 44 different research cohorts worldwide and harmonized by the ENIGMA Stroke Recovery working group. The anonymizing procedure of removing the facial images and intensity standardization step has been performed as part of normalization scheme. For our study, we selected ATLAS v2.0 following:

- The quality and size of the dataset, as currently, no dataset exceeds 250 stroke images,
- The robustness and reliability endorsed by the accredited institution,
- And finally, the possibility of interaction with other users for feedback and comparing the performance of the model as well as room for transfer learning

5.2 Methodology

For creating a deep learning model, we utilized Tensorflow and Pytorch interchangeably and project MONAI, a website containing built-in deep learning layer for users, was used for Pytorch models when building complex blocks such as transformers.

For training and testing of the models, we used high performing computer (HPC) provided by the University of Auckland and Tmux software for training the models in the background without the need to keep the computer on.

5.3 Preprocessing

Preprocessing of images before running a model is a critical and necessary step in computer vision tasks. Its primary purpose is to increase the quality of the images and reduce any artifacts present. In segmentation tasks for medical imaging including stroke lesion segmentation, common methods include co-registration, intensity normalization, skull

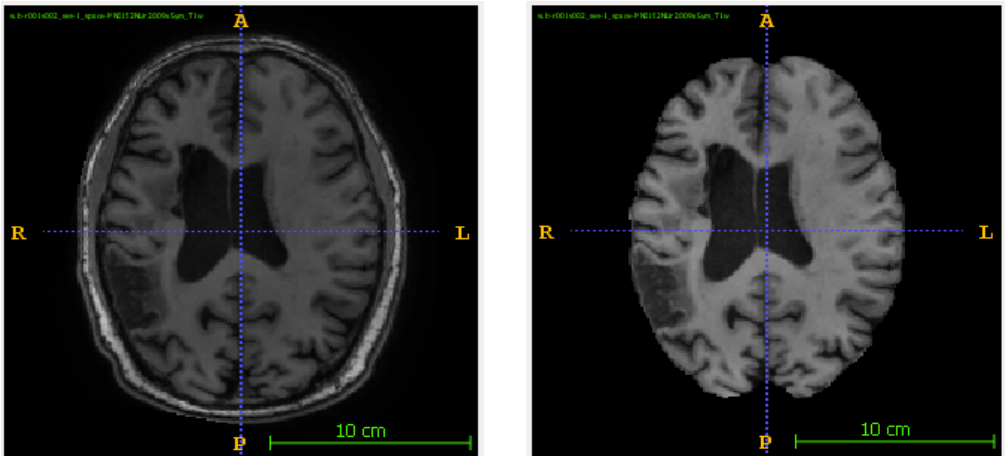
stripping, bias field correction and manual correction by specialists. For our dataset, few of the preprocessing techniques have been applied to all our images including both training and testing images to enhance the segmentation results. The protocol to preprocessing our dataset is as follows:

- **Intensity Normalization**: Intensity normalization is a common technique in the context of medical image processing. It is used to standardize the pixel or voxel intensities within an image to ensure the intensity values across different images or modalities remain consistent [67]. For our dataset, z-score normalization technique is applied, a widely used method in preprocessing medical images that subtracts the mean from each intensity value and divide them by the standard deviation in which the intensities are rescaled to have a mean of 0 and a standard deviation of 1 [68]. The equation is as follows:

$$Z = \frac{X - \mu}{\sigma}$$

  Where Z is the normalized intensity, X is the original intensity value, μ and σ are the mean intensity value of the image and its standard deviation respectively.

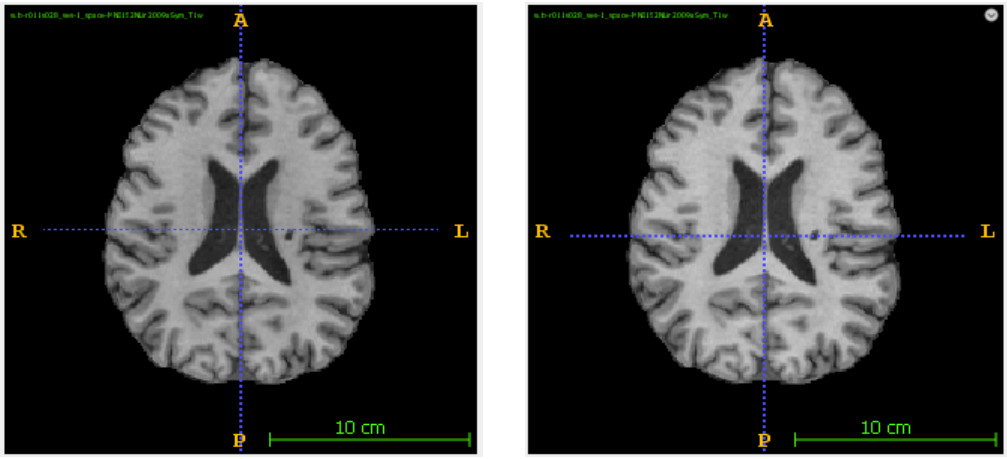
- **Skull-stripping**: It is also known as brain extraction or skull-stripped brain segmentation which involves the removal of non-brain structures such as the skull, scalp, and other extracranial tissues isolating only the brain and relevant intracranial structures of interest. This process is crucial for reducing any artifacts and overfitting during training. Our images have been processed using BET extraction provided by FSL software, kept at 0.5 intensity scale to keep the extraction quality consistent across the entire dataset. Through skull stripping, we were able to achieve dice score in the range of 0.021 to 0.044 higher than the non-skull-stripped images depending on the deep learning model used.



(a) Original (b) Skull-stripped

Figure 5: Before and After Skull-stripping has been performed

- **Bias correction**: Bias correction is a process to remove intensity inhomogeneities or so-called bias present in medical imaging data, particularly magnetic resonance imaging (MRI) scans. These intensity variations caused by imperfections can lead to inaccuracies in segmentation particularly when delineating stroke lesions. Hence, quality enhancement is essential, especially in stroke lesion segmentation where intensity contrast is an important factor in identifying the lesion as most stroke scans do not appear notably different between the lesion and healthy tissue as much as other brain scans like brain tumour. N4 bias correction scheme was used for our dataset based on the concept of iteratively estimating the bias field and then dividing the image by this field to correct for inhomogeneities [69]. N4ITK software for ANTs (Advanced Normalization Tools) preprocessing based on Python was used.



(a) Original (b) N4 Bias corrected

Figure 6: Before and After N4 bias correction has been performed

As observed from the figure above, the contrast of the brain area is notably improved. Given all the images in the training set, the intensity values of the brain area increased from a mean of 40.75 to 74.10.

- **Image Augmentation**: Various augmentation techniques such as cropping and flipping have been applied to the images. NnUnet data augmentation toolkit was used for our images. Since nnUnet was prominently used in 2D medical image segmentation, additional stacking was implemented to build a 3D image with a mask located at the last z-dimension following the convention. As a result, images were cropped to 128 x 128 x 189 which reduced the computational expense of our model and since most of our brain size fit between 128 x 128 on the 2D plane, no data was lost. During the training, the images were expected to rotate 90°, 180° and 270° with a single flip.

5.4 Cross Validation Method

In machine learning tasks including computer vision, cross-validation is an important technique to ensure that the performance of the model is kept consistent and generalized [70]. Although a wide range of cross validation schemes are acknowledged, five-fold cross-validation is recognized as the most common cross-validation method. In general, five-fold cross-validation splits the entire dataset into 20 percent test set and 80 percent training set and ran five times, each containing different training sets and test sets. Each trained versions are then ensembled to give a much more generalized predictive representation of the entire dataset.

For our dataset, thorough cross-validation was needed as we were dealing with a largest public dataset ever used in stroke lesion segmentation. However, in most papers that involve stroke lesion segmentation, the authors simply suggested the usage of five-fold cross-validation without specifics on how they managed the dataset and each corresponding fold. For instance, our tests showed that training with images under certain lesion size criteria gave a high variability in the dice score. There was a large difference in training with only large lesions and tested on the small lesions, and vice versa. The total number of images (534 for training and 131 for testing) was kept the same throughout.

Criteria	Large lesion	Small lesion
Dice Score	0.4112	0.7748

Table 1: Differences in test result under biased lesion training

Hence, an idealized cross-validation method suitable for stroke lesion segmentation was required.

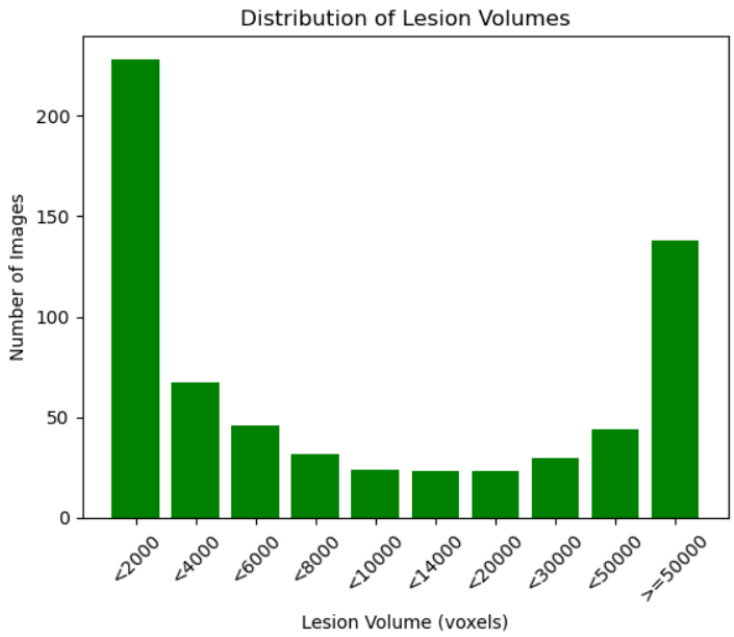


Figure 7: Histogram of ditribution of lesion volume in ATLAS 2.0 dataset

As notable from the above histogram, lesion volumes are not evenly distributed but more biased to each ends; lesion volume less than 2000 voxels and larger than 50000 voxels. Among our 655 images, these take up more than half of our dataset and as small labels are a well-known cause of overfitting and class-imbalance, an optimal method to evenly allocate small labels is essential.

A Greedy algorithm is arguably the most popular algorithm in combinatorial optimization which characterizes local optimum (selections) at each stage in search of a global optimum [71]. This algorithm has been used to create 5 set of folds each containing 131 images and almost equal total lesion volume. A brief process of the algorithm is given below

Algorithm 1 Greedy Algorithm for Lesion Volume Distribution

```
Sort files in descending order based on lesion volumes.
Initialize empty sets and total lesion volumes array.
for each file do
    Identify the set with the minimum total lesion volume.
    Add the file to the identified set.
    Update the total lesion volume for the chosen set.
end for
Create a new DataFrame to represent the distribution.
Save the DataFrame.
```

Following the algorithm, five set of dataset with almost equal lesion volume has been created.

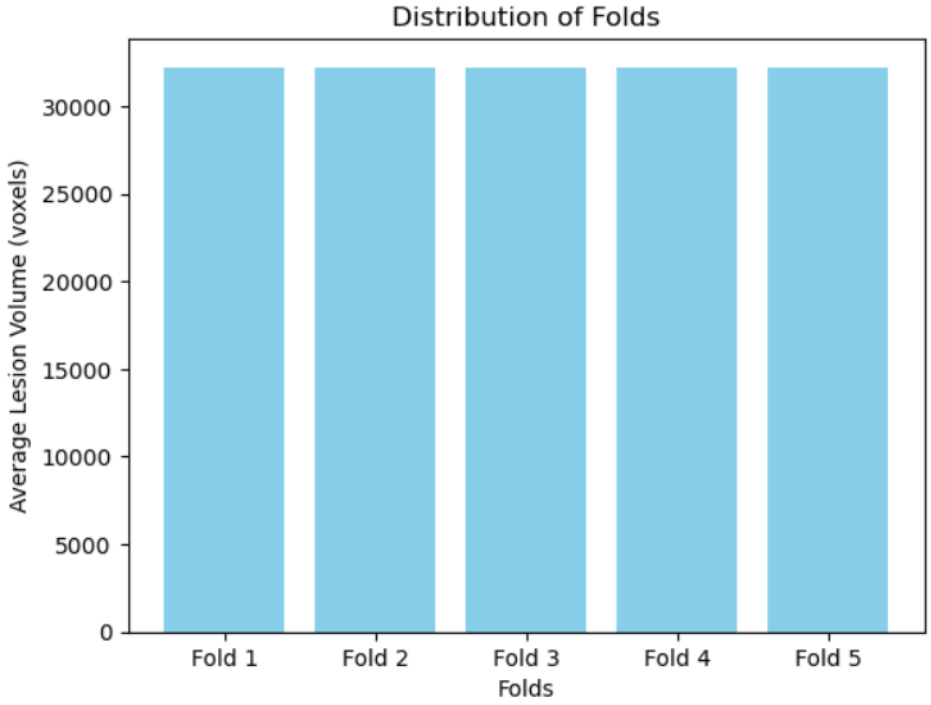


Figure 8: A balanced distribution of lesion volumes to five folds

Among the folds, maximum and minimum average lesion volume was 33222.19 and

33222.08, corresponding to less than 0.005% difference. Next, five fold cross-validation has been performed internally within each fold. each results have than been ensembled using a stacking method that stack multiple labels simultaneously to create the best label [72]. The result of each fold after internal five fold cross-validation is given below.

Fold	**1**	**2**	**3**	**4**	**5**
Dice Score	0.5925	0.6089	0.5978	0.6366	0.6247

Table 2: Dice Score for each Fold used as test set

It is shown that fold 4 scored the highest. Hence, all the models created have used fold 4 for testing and the rest for training to ensure the best possible dice score while keeping the entire dataset evenly distributed in terms of the lesion volume. This is because for ATLAS 2.0 dataset, no seperate testing set can be accessed or evaluated after the competition ended which is why all the papers use only the training set (655 images) to evaluate the model performance and validity. In real life cases where a set of stroke images have been acquired for testing, each fold can be trained and ensembled meaning there will be 25 training in total. This method has been implemented for the dataset in chapter 2.

5.5 Loss function

Loss function, also known as a cost function, is a key component in machine learning algorithms that quantifies how well a model's predictions match the actual ground truth during training. The goal of the training process is to minimize this function or differences in order to enable more accurate predictions [73]. In computer vision tasks such as object detection, image analysis, or segmentation, it is inherently better to use loss functions that have significant classification power for its necessity to delineate and classify different regions or objects within an image. For our stroke lesion segmentation models, the loss functions used will be introduced below. Most loss functions have been referenced by the built-in units from project MONAI or Tensorflow.

- **Binary Cross Entropy Loss**: Binary cross entropy (BCE) loss is a branch of cross-entropy loss that quantifies the difference between two probability distributions; the predicted probabilities and the actual ground truth labels. Specifically, binary cross entropy outputs the probability between two classes only which in our case the brain and the lesion, denoted as 0 and 1 respectively. It is most commonly used in simple Unets and CNN models for stroke lesion segmentation [27]. The formula is given below:

$$H(p, q) = -\left(y \cdot \log(p) + (1 - y) \cdot \log(1 - p)\right)$$

  where y represents the binary label either 0 or 1, and p represents the predicted probability.

- **Dice loss** Dice loss, also known as dice similarity coefficient. measures the similarity of overlap between the predicted labels and the ground truth labels. The dice loss is then denoted as 1 - DSC (Dice score coefficient) to penalize the model harsher when the similarity between them is low. The equation is given as:

$$\text{Dice\_loss} = 1 - \frac{1}{c}\sum_{i=0}^{c}\frac{\sum_j^N 2g_i^j * p_i^j + \epsilon}{\sum_j^N g_i^j + \sum_j^N p_i^j + \epsilon}$$

where,

$$g = \{0, 1\}, p = \{0, 1\}$$

or Simply:

$$\text{Dice Loss} = 1 - \frac{2 \times (A \cap B)}{A \cup B + \epsilon}$$

Since it uses spatial overlap to calculate the similarity, dice loss is well-known for countering the similar imbalance problem called class imbalance where the background, the brain in this case, takes up much more weight against the lesion especially when the lesion size is very small. Class imbalance is prevalent in binary cross-entropy loss models hence why dice loss is preferred in medical imaging models [41] [58].

- **Focal loss:** Similar to Dice loss, Focal loss focuses on reducing the class imbalance between the background and the label [42]. However, Unlike dice loss, focal loss focuses on emphasizing misleading and hard-to-delineate labels and suppressing well-identified labels by changing the model weights accordingly, thereby improving its classification power. Its equation is as follows:

$$\text{focal\_loss} = -\frac{1}{L_i}\alpha(1-g)^\gamma * \log p$$

where,

$$g = \{0, 1\}, p = \{0, 1\}, \alpha = [0, 1]$$

- **Enhanced Weight loss:** To address and improve the power of the loss function to take account of the lesion volume while training, we propose an imbalanced weight (IW) unit. This is done by dividing the mean lesion volume in the training images by the maximum lesion volume, shown as:

$$\text{IW} = \frac{\sum_i^N L_i}{n * (\max_i^N L_i)}$$

The enhanced weight loss is therefore shown as:

$$\textbf{Enhanced\_Weight\_loss} = IW * \text{BCE\_loss} + (1 - IW)(\text{focal\_loss} + \text{dice\_loss})$$

As smaller IW represents the larger class imbalance within the training subject, enhanced weight loss is designed to penalize BCE loss when the class imbalance is high and simultaneously enhance both focal loss and dice loss. Since class imbalance was a critical issue for ATLAS 2.0 dataset, improvement in dice score has been observed.

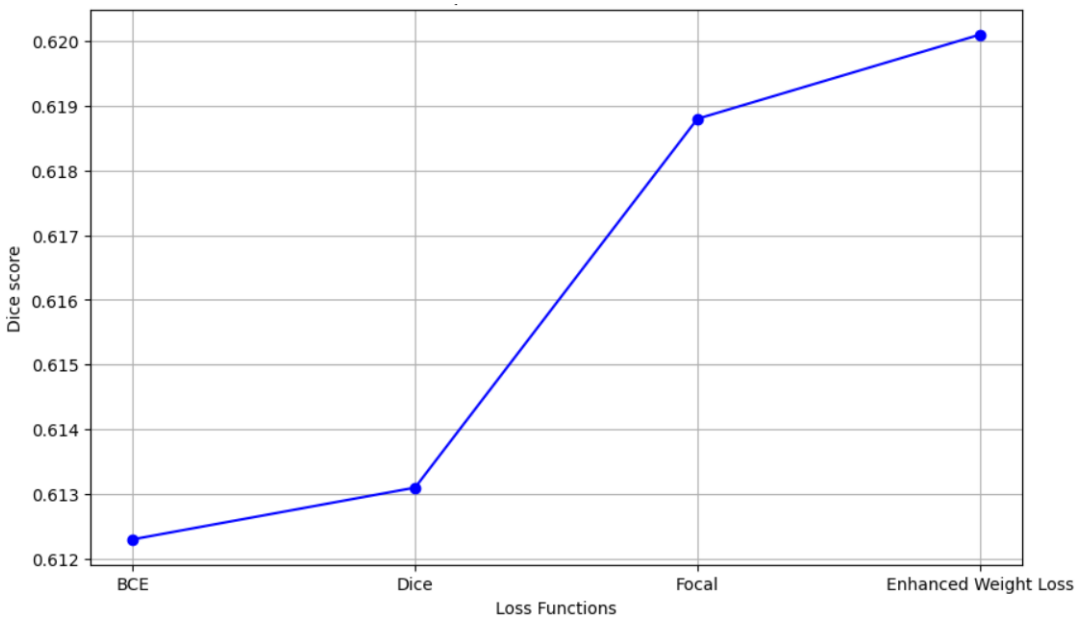


Figure 9: A figure showing dice score of different loss functions performed using Dense-net

5.6 Metrics and Scales

Metrics are important in the field of machine learning as they provide quantitative measures to evaluate the performance, accuracy, and effectiveness of models. Metrics used to represent the performance of my deep learning models are outlined below.

- **Dice Score:** As introduced in the loss function section, dice score is used to quantify the similarity or overlap between two sets, often applied in the context of image segmentation or binary classification tasks. It is commonly used to evaluate the performance of segmentation algorithms where it measures the overlap between the prediction and the given inputs; in our case, the lesion and the brain. Dice score has been selected as my primary metric to directly compare with the published literatures that use the same dataset, ATLAS 2.0.

- **Hausdorff distance:** It is a measure of the distance between two sets of points in a metric space. Specifically, it is used in the context of comparing two geometric shapes or point sets such as computational geometry or 3D images. For stroke lesion segmentation, the lesion label and its prediction have been compared to evaluate its matching. The unit used to measure the distance is in mm, converted from the magnetic resonance image voxels.

5.7 Deep learning models

5.7.1 LKA-net

Attention mechanism is seen as an adaptive selection process that selects specific features and automatically reduces noise to the corresponding input features [1]. LKA (Large Kernel Attention) mechanism follows this general attention mechanism criterion in which it creates large kernel convolutions as shown in the figure below.

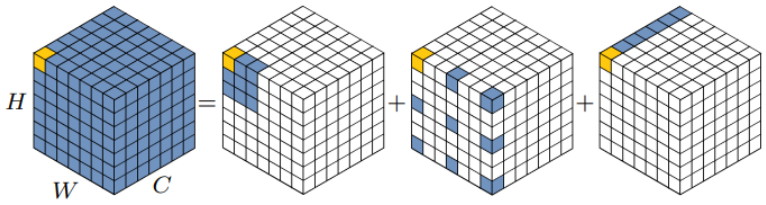
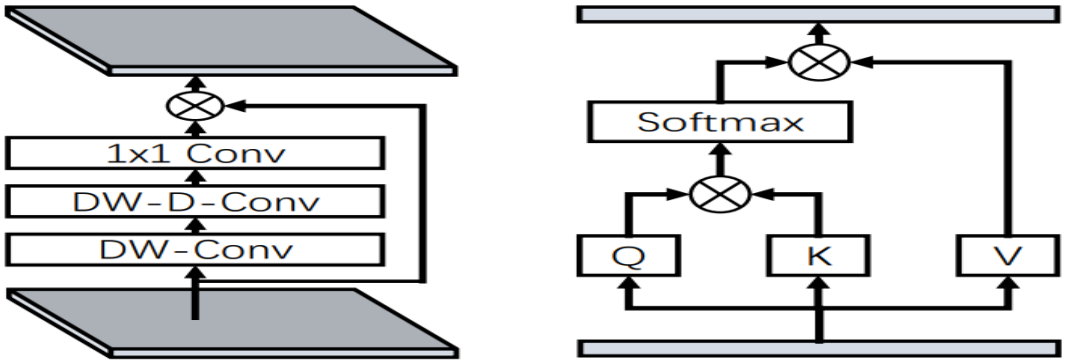


Figure 10: A figure showing Large Kernel Convolution block decomposed into different convolution components; from the left: DW-Conv, DW-D-Conv and a pointwise convolution [1]

As notable from the figure, large kernel convolution can be decomposed into three main components. A depthwise convolution (Dw-Conv) that captures local contextual information, a depthwise dilation convolution (DW-D-Conv) that captures long-range dependency, and finally a 1 x 1 convolution block that provides adaptability to channel dimensions when used with attention mechanism. When creating a K x K size kernel, it is made of a $\frac{k}{d}$ by $\frac{k}{d}$ DW-D-Conv with dilation rate d, (2d - 1) by (2d - 1) DW-Conv and a 1 x 1 pointwise convolution. Hence, For a 21 x 21 size large kernel, DW-D-Conv is 7 x 7, Dw-Conv is 5 x 5. The pointwise convolution was always kept to 1 x 1 in size. These large kernels are then fed through the attention mechanism to create an LKA block for each channel. The attention mechanism is shown below.

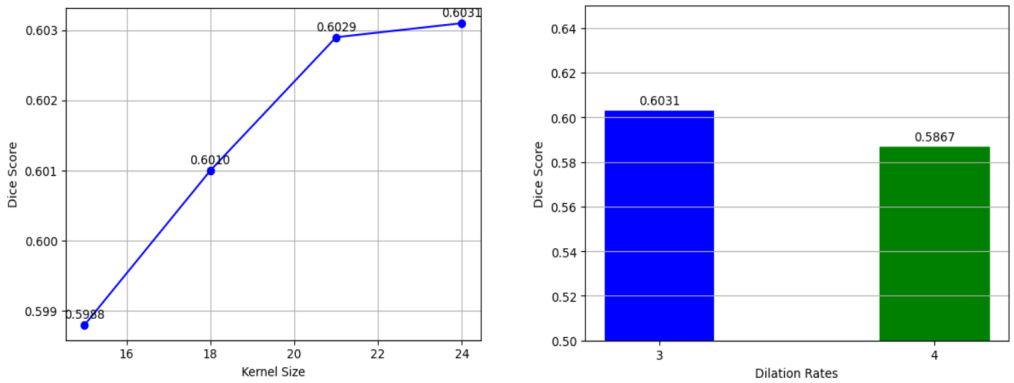


(a) Large Kernel Attention module

(b) General Attention mechanism

Figure 11: A diagram showing Large Kernel Attention Module and the referenced attention mechanism

Since our stroke MR images are 3 dimensional in shape, the kernels were all given an extra dimension with equal size. Hence, the kernels used for our studies were, for example, 21 x 21 x 21 size large kernel, 7 x 7 x 7 Dw-D-Conv, 5 x 5 x 5 DW-Conv and 1 x 1 x 1 pointwise convolution. Kernel size for the decoder of the U-net was kept at 3 x 3 x 3 throughout, and both the encoder and the decoder had 5 blocks of channels following the U-net convention.



(a) large Kernel convolution difference against dice score at dilation rate of 3

(b) Dilation rate difference against dice score

Figure 12: A figure showing comparison between each components in LKA

As notable from the figure above, a larger kernel size generally gave a higher dice score suggesting that the segmentation power was directly proportional to the size of DW-D-Conv (Dw-Conv was kept constant by keeping the dilation rate at 3). Moreover, figure b) suggests that reducing the dilation rate, therefore increasing the DW-Conv size, has a negative effect on the performance of the model.

5.7.2 Transformers

Use of transformers in computer vision, especially medical imaging has been brought to the light recently and as of 2023 it is seen as a state-of-the-art method in achieving a high-end performance in both object detection and segmentation. Numerous papers have suggested the use of 4 to 5 blocks of channels both in the encoder and the decoder, possibly due to a computationally expensive procedure. Here, the performance of the transformer method has been tested on 6 blocks of channels to make use of our high performing computers.

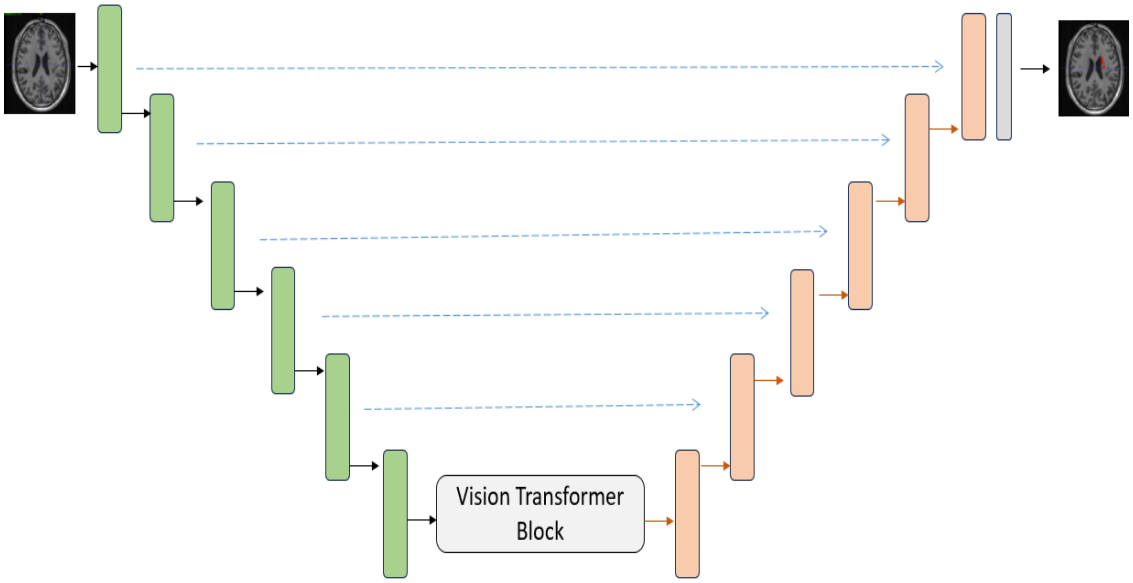


Figure 13: A figure of a 6 blocks of channel Vision transformer model

From the diagram, green blocks in the encoder represent a nominal convolutional block that performs convolutional layer (Conv3D), RELU activaton and batch normalization in a sequence, followed by maximum pooling (MaxPool3D) shown as a black arrow. In the decoder side, orange blocks represent a residual convolutional block that operates in a sequence of upsampling (transposed convolution) shown as an orange arrow, concatenation of skip connections shown as blue dotted arrows, convolutional layer, RELU activation and batch normalization. The vision transformer block is located at the bottleneck position to process the entire image as a sequence of patches to effectively capture the global relationship between each patch.

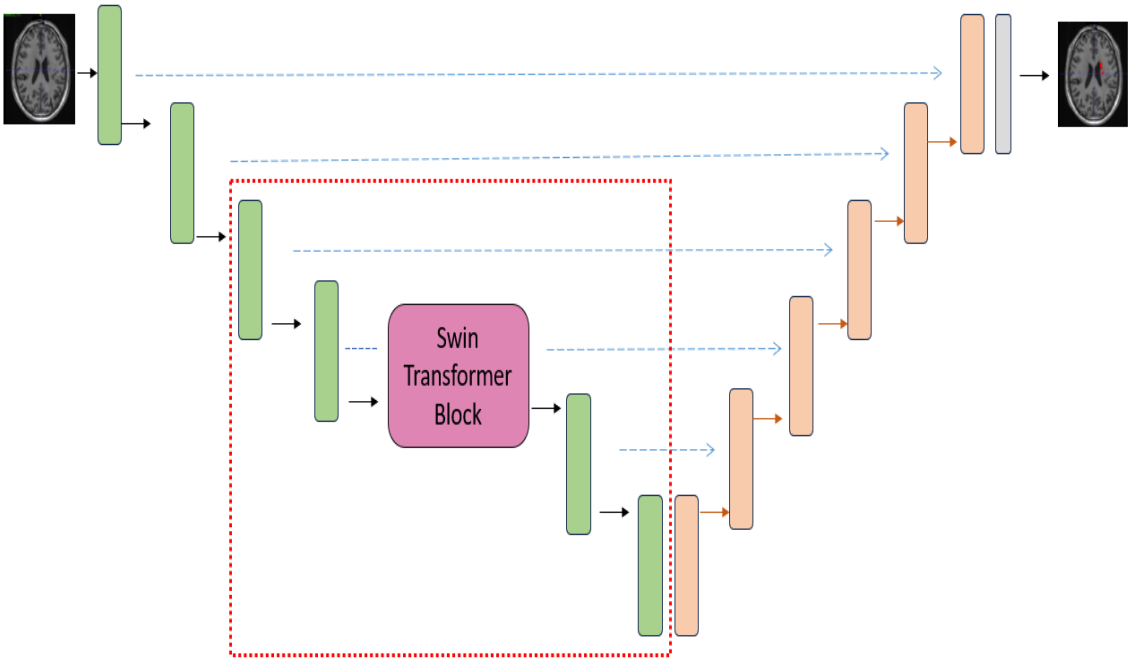


Figure 14: A figure of 6 blocks of channel Swin transformer model; red dotted square represents the boundary of Swin transformer installation

In the case of swin transformer, the position of the transformer block is not located at the bottleneck but between convolutional blocks in the encoder. This is due to the hierarchical processing capability of swin transformers as opposed to traditional ViTs. The performance of the swin transformer model with respect to the location has also been tested for an ideal location.

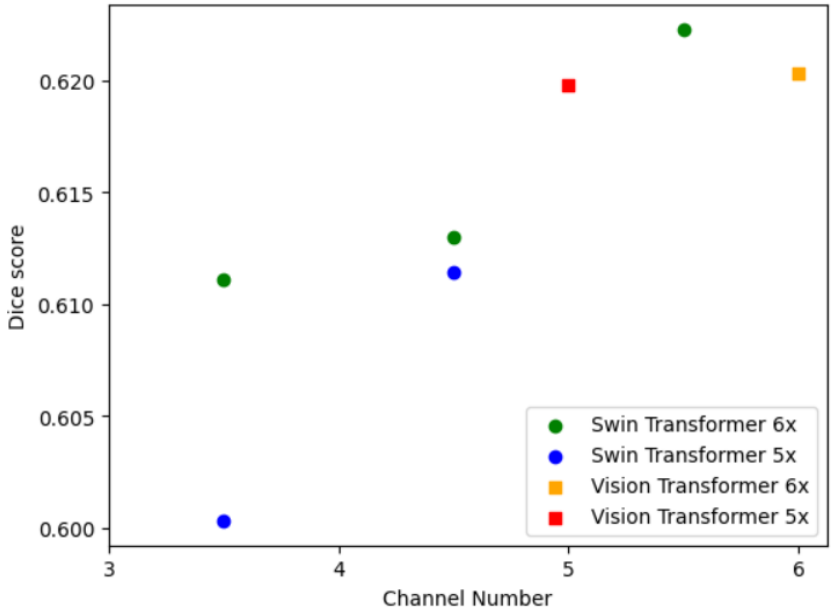
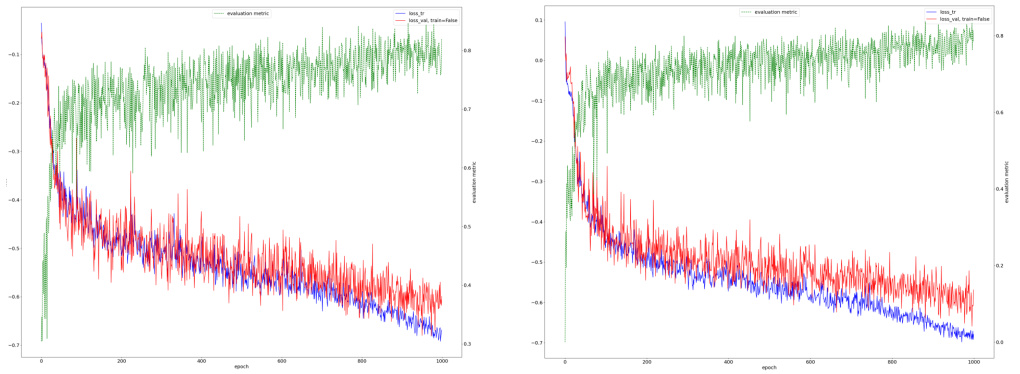


Figure 15: A graph showing the performance of transformers in 5 blocks and 6 blocks of channels

The channel number represents the channel block location where the transformers are

installed. For instance, for a 6-block channel model, swin transformers were installed between blocks 3 and 4, 4 and 5, and finally 5 and 6. In general, an increment in dice score was observed as the transformer block was located closer to the bottleneck. In terms of the ViTs, only a slight change in the dice score was observed by creating a deeper model.



(a) Vision Transformer (b) Swin Transformer

Figure 16: Changes in dice score and loss throughout training of the two transformers

It is also notable that the vision transformer experienced greater fluctuation in dice score during training as opposed to swin transformer. Despite the fluctuation, vision transformer managed to achieve a similar final dice score of above 0.8 to that of swin transformer.

5.7.3 Dense-net

This model has been inspired by the semantic segmentation model designed by the University of Auckland PhD candidate Parvez Ahmad and served as part of a collaborative project in applying semantic segmentation and Dense blocks to the ATLAS dataset and validate its segmentation performance [74]. Semantic segmentation partitions an image into multiple segments to assign a class label to each pixel in the image effectively creating a detailed pixel-level segmentation [75]. Hence, it is effective in segmenting a multi-class subset of images such as flair, T2 and T1 all obtained from the same patient.

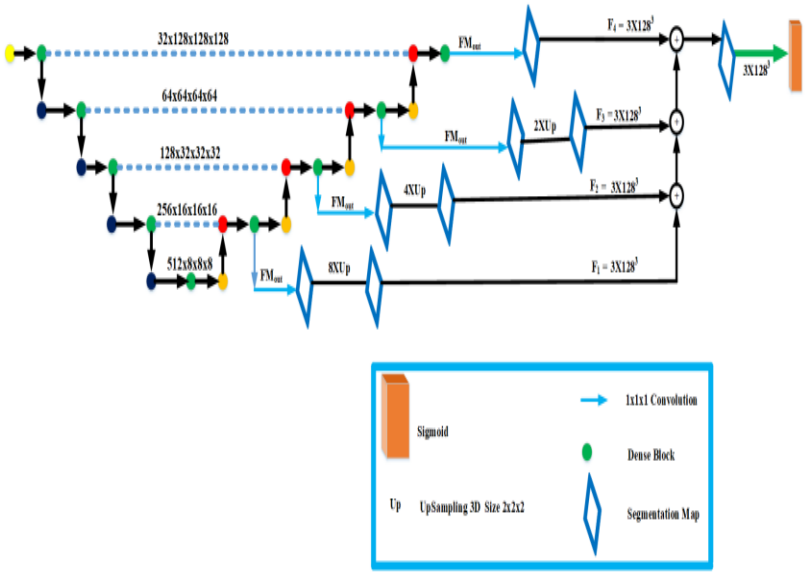


Figure 17: An overview of MH dense Unet suggested by Parvez Ahmad

This network is known for its delicate hierarchical structure where Dense connections allow deep supervision that fosters smoothing of the gradients flow, and subsequently decreasing training parameters. To cater to the necessities of the ATLAS dataset, few changes to this network has been made. Since ATLAS only has one modality which is T1, the multi-class ensembling block has been removed, improving computational efficiency further.

After 5-fold cross-validation and ensembling, we achieved a dice score of 0.6201, utilizing the Enhanced weight loss function

5.8 Diffusion model

As opposed to the common and popular methods such as CNN or Transformers, the diffusion probabilistic model (DPM) recently gained attention among researchers for its generative power in deep learning where ground truth is not present [76]. However, recent studies suggest that diffusion models can also perform meaningful segmentation tasks in

the computer vision field [77]. This led to research in the medical field, resulting in a paper by Dr. Junde Ju, the first to perform diffusion-based segmentation on a large-size medical image dataset such as brain tumour (BRATS) or An open access thyroid ultrasound image database (DDTI) [2][78] with published papers. Hence, there were no papers to present date that diffusion models have been used on segmenting stroke lesions and it is worthwhile to give attention to this newly built state-of-the-art method for the first time. Moreover, due to his pioneering work, my codes and methods are built upon the basis of this paper.

Diffusion models are composed of two stages, a forward diffusion stage and a reverse diffusion stage. In the forward process, the segmentation label x0 Gaussian noise is gradually added through a series of steps T that creates a fully noised image. In the reverse process, a neural network is trained to reverse the noising process (denoising) to recover the original image data. The equation of the diffusion model is as follows:

$$p_\theta(x_{0:T}) = p_\theta(x_0) \prod_{t=0}^{T-1} p_\theta(x_{t+1}|x_t) \tag{1}$$

$$p_\theta(x_{0:T-1}|x_T) = \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t) \tag{2}$$

where equation (1) and (2) represent forward and reverse processes respectively, $p_\theta(x_0)$ is the initial probability of distribution, $p_\theta(x_T)$ is the latent variable distribution and Gaussian noise at that state is denoted as $\mathcal{N}(x_T; 0, I_{n\times n})$ with I as the raw image. To keep the procedural symmetry between the forward process and the reverse process, the noise image is recovered step by step along the time series to obtain the final clear segmentation.

Following the diffusion probabilistic model, the segmentation model is embedded afterward to perform a segmentation task that creates a mask onto the tested image. The step estimation function used for this is given below:

$$\epsilon\theta(x_t, I, t) = D\left((E_t^I + E_t^x, t\right), t)$$

where E_t^I is the conditional feature embedding that uses the raw image, and E_t^x is the segmentation map feature embedding. This information is then delivered to a decoder denoted as D for the reconstruction of both the segmented mask and the original raw image. The author of the paper used Residual Unet with an arbitrary number of RESNET layers beginning from 4 where a residual block consisted of two convolutional blocks each with group-norm and SiLU activation layer for the segmentation encoder, and a simple Unet for the segmentation decoder [79]. The summarized model is depicted in the figure below:

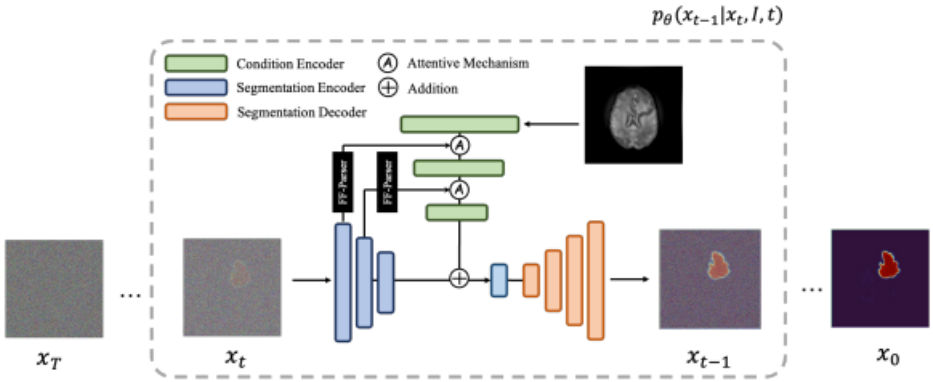


Figure 18: A figure of the diffusion segmentation model suggested by the paper [2]

In addition, we tried to enhance the segmentation power by adding a transformer into the encoder since the original model consisted of ResNet34 only. The enhanced version of the segmentation model is given in the figure below:

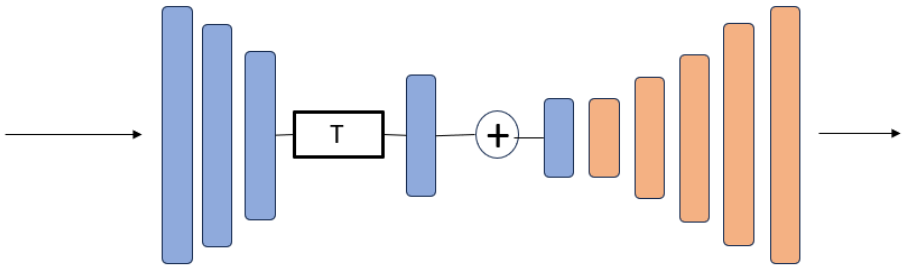


Figure 19: A figure of the enhanced diffusion segmentation model with transformer encoder and residual unet decoder with varying number of layers

The transformer encoder is very similar to the model given above, with a sliding window transformer in the middle layer denoted as T. The decoder has been replaced with the Residual layers created with project MONAI. The model has been trained with an image of 192x192x128 cropped to the suitable size and the batch size was kept at 4 to reduce the computational memory as much as possible. Diffusion steps; a number of time increments through which the diffusion process evolves, were kept at 1000 following the convention.

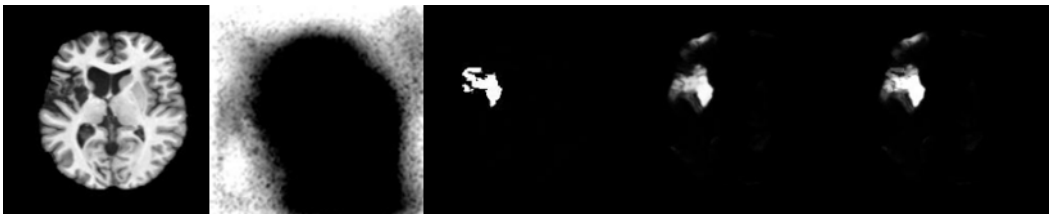


Figure 20: From left; Brain sample, Noised brain during sampling, segmented label, binary label at 0.3 binary point, and binary label at 0.7 binary point

Performance of diffusion models			
Model combination	**Dice score**	**Hausdorff distance**	**Computational memory usage (MB)**
Original+4 residual layers	0.3735	108.30	44,003
Original+8 residual layers	0.3993	82.99	48,667
Original+12 residual layers	0.4014	89.38	52,221
Swin Transformer+4 residual layers	0.4038	85.55	68,965
Swin Transformer+8 residual layers	0.4171	84.40	77,498

Table 3: Metric and computational efficiency comparison of various diffusion model combinations

In general, diffusion models suffered a lack of segmentation power against other methods such as transformers or attention mechanisms. Although an increase in the number of layers and adding transformers did improve the overall dice score, they were not high enough to compete with the other methods. A stable inverse correlation between the dice score and the Hausdorff distance was observed.

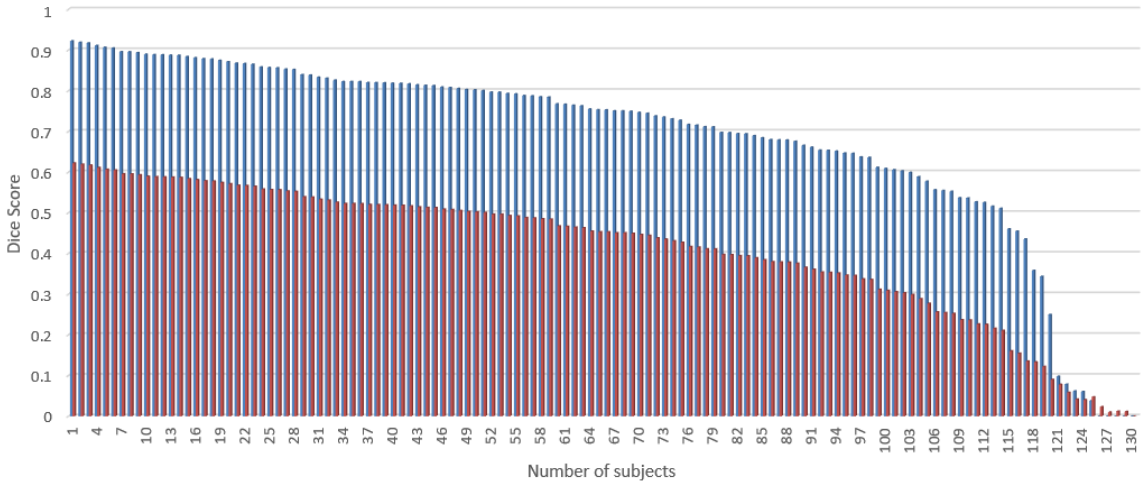
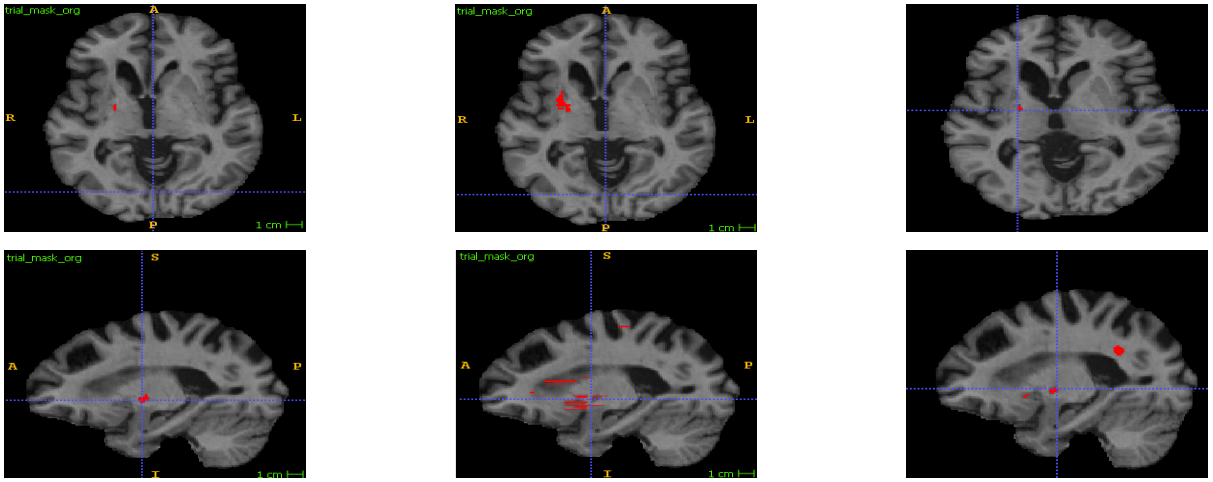


Figure 21: Performance of diffusion model (red) against swin transformer model (blue)

Although diffusion model performed poorly on most of the images in the dataset, a comparable dice score was observed near the dice scores close to zero, meaning very small or no overlap was found. These lesions turned out to be small lesions (lesion volume

< 500) with high contrast (visible brain holes). From the figure below, we can observe that the transformer model introduced a few false positives along the z-axis dropping the overall dice score. Although some noise was spotted, the diffusion model is arguably performing better in segmenting small lesions with reduced caveats in overfitting.



(a) Ground Truth (b) Diffusion (c) Transformer

Figure 22: Performance of diffusion model against transformer model in small lesion. Top; Axial-Coronal-plane and bottom; Saggital-axis

5.9 Enhanced Ensemble Method

Ensembling is a technique of combining multiple individual models to improve overall predictive performance rather than relying on the output of a single model, providing diversity in segmentation powers leading to higher accuracy [80].

To further improve the performance of our model, an enhanced ensemble method has been proposed. The enhanced ensemble method brings the prediction of each model into one place through ensembling; specifically, by leveraging a stacking generalization method that overlays predictions to create a large and comprehensive metadata model. Although ensembling results from the same model are frequently performed on the five-fold cross-validation, ensembling between different models has been under-studied in the medical imaging field. Hence, all our proposed models from above have been ensembled to achieve notable enhancement in the overall performance.

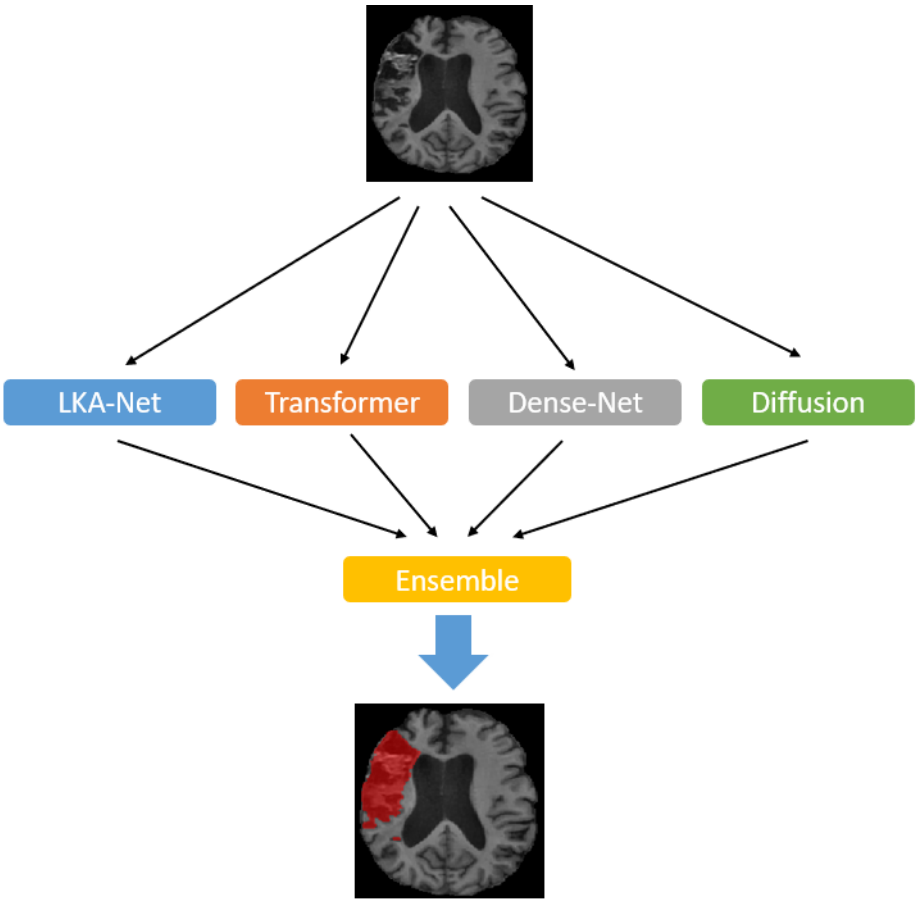


Figure 23: A diagram of Enhanced Ensemble method

The algorithm below ensures that the ensemble model only keeps the dice score higher than the previous one, and keeps the old value constantly updated every time a new model is introduced. This is because when overlapping another segmentation on top of another and by simply increasing the size it is not guaranteed to achieve higher dice score and can be prone to increasing the Hausdorff distance. Hence, selection of segmentation

Algorithm 2 Algorithm for Enhanced Ensemble Method

```
Prepare a list of predictions from an initial model.
for each model do
    Stack the current model prediction onto the target model prediction
    if Dice score of stacked prediction < previous prediction then
        Keep the previous prediction
    else
        Update to the stacked prediction
    end if
end for
Binarize the prediction to remove any artifacts.
Save the list of predictions.
```

to ensemble was chosen with care as it can introduce false negatives fatal in a medical field. For instance, diffusion model only performance on high standard for small lesions, therefore only segemented labels with lesser than 1000 voxels have been considered for ensemble.

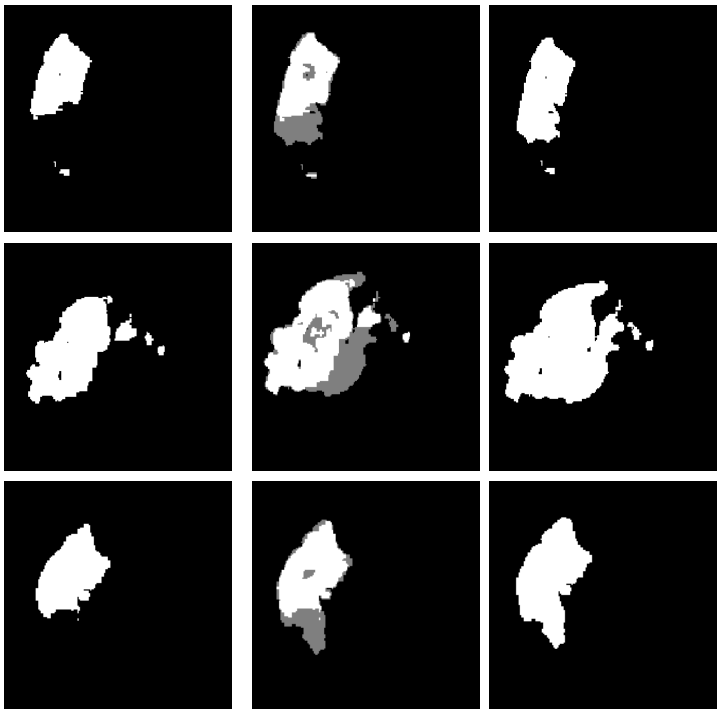


Figure 24: From left; Segmentation of LKA-net, ensembled between LKA-net, dense-net, and Swin Transformer, and the final ensembled label. Each row represents x, y, and z axis respectively

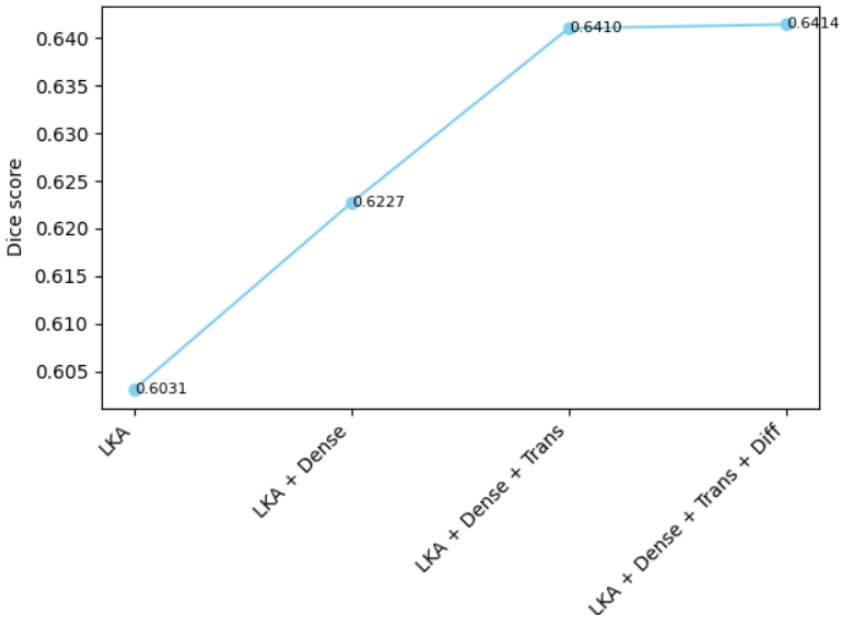


Figure 25: Improvement in performance of Enhanced Ensemble method

As depicted in the figure above, the enhanced ensemble method achieved dice score of 0.6414, an increase from our best-performing method of 0.6366. When ensembling the diffusion model, only small lesion (lesion volume < 500 voxels) labels were ensembled as significant noise was introduced in larger lesions.

5.10 Overall performance of models

The performance of all models presented in this thesis has been compiled as below.

Model	Dice score	Hausdorff distance
Diffusion	0.4171	84.40
LKA-net	0.6031	47.11
Dense-net	0.6213	45.49
Transformer (swin)	0.6366	44.38
Enhanced Ensemble	0.6513	41.64

Table 4: Dice score and Hausdorff distance of suggested models

Overall, Hausdorff distance dropping was proportional to the dice score going up which is ideal for achieving a compatible high performance model. Enhanced ensemble method achieved both highest dice score and lowest Hausdorff distance as expected, and presented as the best performing method regarding this thesis.

5.11 Post-processing

Post-processing of images, a refinement technique that additionally enhances visualization after running a model, offers significant advantages like noise reduction, feature enhancement, and standardization [81]. In medical image processing, we are faced with a greater challenge such as poor signal-to-noise ratio, low spatial resolution, and indistinct contrast between the anatomical features [82]. Moreover, changes made must be reliable and controlled so that it does not compromise the clinical decision-making process, and provide credibility that artifacts are removed (e.g. motion bias in MRI).

Post-processing techniques used on our images are outlined in chronological order below.

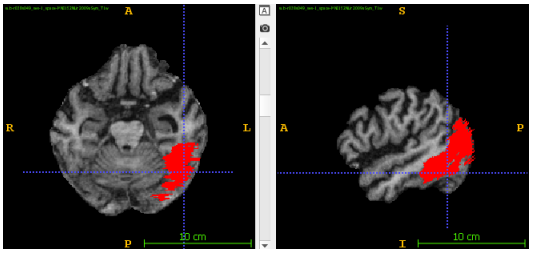
- **Thresholding:** generally a first technique to perform in image processing that converts a grayscale image into a binary image. The process sets a threshold value and subsequently classifies each pixel value in the image based on whether its intensity value is above or below the threshold [83]. Since our images are primarily in a binary form (0 for the brain and 1 for the lesion), binary thresholding with a thresholding value of 0.5 has been applied to the images.
- **Morphological Operations:** Morphology is a large extent of operations that modify the images based on shapes. Morphological image processing aims to eliminate defects such as inconsistent texture and uneven boundaries, and further maintain the structure of the image [84]. Morphological operations are known to be reliable only on the correlated ordering of pixel values, rather than their numerical values. Hence, they are focused more on binary images [85]. Erosion and dilation

are two common methods for this technique. Erosion shrinks (erodes) the boundaries of label objects in a binary image where it scans the image with a small kernel and sets the pixel value to 0 (background) whereas dilation acts in the opposite way of expanding the boundaries and setting the values to 1. By simultaneously applying the two techniques, boundary smoothing and broken structure connection can be achieved.

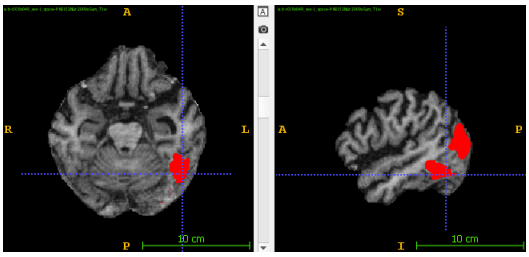
- **Gaussian Smoothing:** Gaussian smoothing, also known as Gaussian blur, is a well-known and traditional technique to suppress and reduce noise [86]. It is performed by the convolution of the image with a Gaussian kernel of standard deviation σ where the size of the standard deviation determines the level of smoothing. If σ is high, it will lead to more aggressive smoothing.

- **Connected Component Analysis (CCA):** Connected component analysis aims to identify and label connected regions in a binary image. Each region is designated with a unique label according to its spatial connectedness against nearby pixels; 4 or 8 connectivity depending on the orientation of edges. These labels are then used to create more distinct boundaries.

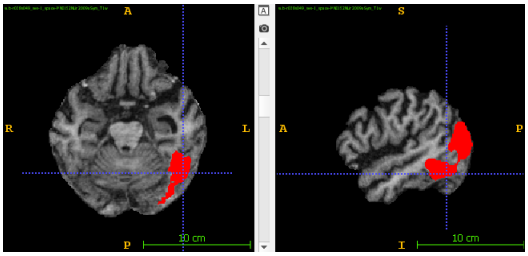
- **Zero Filling:** If the reference segmentation is empty in the prediction, any false positive predictions are counted as zero, resulting in a Dice score of 1. This prevents the Dice coefficient from being undefined due to division by zero. Consequently, the algorithm is placed in the shared rank for this particular test case, region, and metric. This characteristic can be used to enhance the average rank of the proposed model and improve its overall accuracy [87]. Hence, all zero components are converted to 1 removing any false positive predictions.



(a) Ground Truth



(b) Before Post-processing



(c) After Post-processing

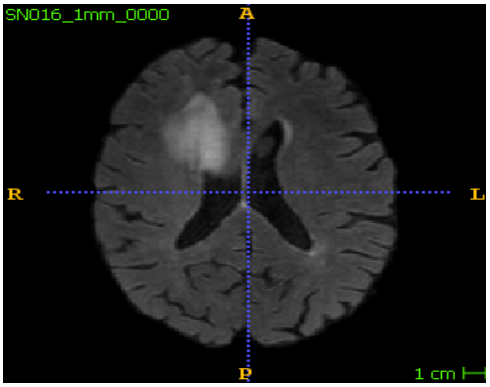
Figure 26: Effect of Post-processing on the prediction

	Dice score	**Hausdorff distance**
Before Post-processing	0.6414	41.64
Binary Thresholding ... (1)	0.6414	41.64
(1) + Morphological Operation ... (2)	0.6511	40.70
(2) + Gaussian smoothing ... (3)	0.6513	40.58
(3) + CCA ... (4)	0.6581	36.69
(4) + Zero Filling	0.7221	36.69

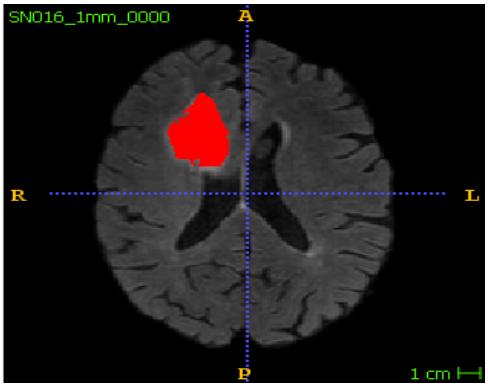
Table 5: Dice score and Hausdorff distance after post-processing

5.12 Results on our own dataset

We have then ran a segmentation on the fMRI T1 dataset that will be introduced in chapter 2. Preprocessing, Enhanced ensemble method and post-processing as outlined in this chapter have all been utilized. The segmentation results are as shown below.

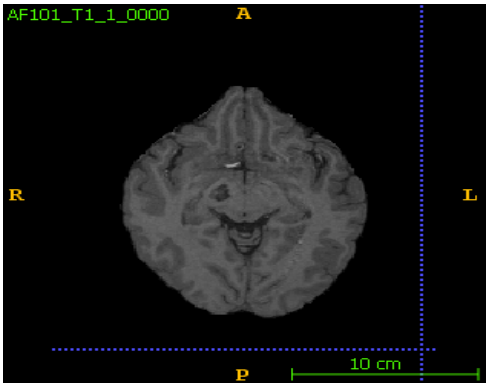


(a) preprocessed original image

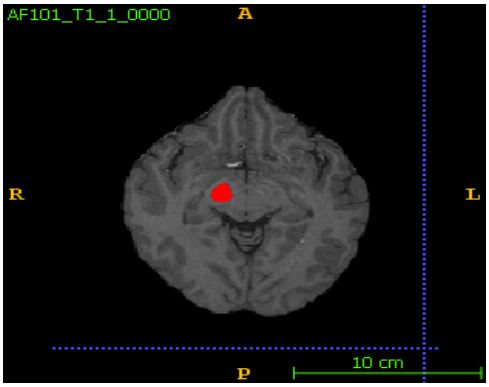


(b) predicted lesion overlayed

Figure 27: Resting state T1 image segmented with enhance ensemble method



(a) preprocessed original image



(b) predicted lesion overlayed

Figure 28: Task based T1 image segmented with enhance ensemble method

Although we cannot specify the dice score of these segmentations as we do not have ground truth labels provided for these images, it is visible that the segmentation model have performed sufficiently well to create a reliable prediction of infarcted lesions. These predictions have been used as input to the stroke recovery prediction performed in chapter 2.

5.13 Discussion

From our models, we found that increasing the number of layers both in the encoder and decoder to create a deeper supervision model increased the dice score, which is an expected result as deeper layers are able to capture more abstract and high-level features from the input data that is interpreted as a clear advantage for tasks such stroke lesion segmentation, as it is notorious for its complex patterns and morphologies. Especially in terms of the transformer models, the deeper layer scheme holds true and is found to be more effective when swin transformers rather than vision transformers are installed as

dice score of swin transformer installed between encoder layers 5 and 6 scored the highest as observable from figure 15. There is a clear correlation between installing the swin transformer toward the bottleneck and an increase in dice score. Although learning from hierarchical features at the earlier stage may help capture the diverse characteristics of stroke lesions such as morphology and shape, having a transformer at the deeper layer will increase its capacity to capture long-range dependencies and global context information that becomes particularly useful when the lesion span across the multiple regions of the brain [88]. Since stroke lesions are found anywhere in the brain with large variance in size and morphology, it could explain why deeper layer transformers perform better in stroke lesion segmentation. In terms of the vision transformers, there is no change between the 5 layer and the 6 layer models implying either the model has reached sufficient capability to effectively capture abnormalities in the image or overfitting may have occurred as deeper layers are prone to overfitting due to increased number of parameters to learn complexity in the image [89].

In terms of the diffusion model, several factors that led to the lack of segmentation performance were identified. Firstly, stroke lesions showed a much greater variability not only in size and morphology but also in contrast intensity against the healthy brain. The dataset used to achieve state-of-the-art performance such as BRATS clearly showed higher contrast between the lesion and the healthy brain under the same modality which was crucial for the segmentation model to detect abnormalities after removing Gaussian noises. Since its performance was poor even after our contrast enhancement preprocessing method such as bias correction was implemented, methods to enhance the contrast like manually increasing the contrast of the lesions specifically rather than the entire image, or advanced contrast techniques such as contrast-enhanced imaging that uses contrast agents., or perfusion imaging that assesses blood flow dynamics could be considered [90]. Another factor was in capturing spatial information. Inherently, the diffusion segmentation can only be applied to 2D images due to the noise generation which forced the MR images to be sliced along sagittal axis and perform segmentation on each of the axial brain images. Hence, the spatial information that assists in capturing complex structures and patterns within the volume has been depreciated leading to a significant decrease in its ability to effectively delineate structures that extend across multiple slices or exhibit complex spatial configurations. Moreover, this explains why the transformer on the segmentation model did not increase the segmentation power significantly unlike observed in 3D images. Although transformers are well-known for decreasing long-range dependency by capturing global contextual information utilizing its large receptive field, it appears particularly effective when convoluted in a 3-dimensional image meaning that the segmentation power is not as strong in 2-dimensional image segmentation, specific to our case.

Our proposed loss function; Enhanced weight loss, shows improvement in segmentation for ATLAS dataset. This is not only because of incorporating widely used loss functions such as dice, BCE and Focal but also because of the strength in our imbalanced weight unit. The need for this unit was raised initially as stroke lesions possess inherent variability in morphology, shape and size unlike other medical images such as brain tumours. Hence, a loss function suitable for training wider ranges of lesion size was required. The enhanced

loss function clearly acted as a channel to reduce class imbalance as focal loss and dice loss were given higher weight when the class imbalance was high in the training set, calculated from the imbalance weight unit. Vice versa, when the class imbalance is small, BCE loss was given higher weighting to promote faster convergence and avoid overfitting. Since class imbalance is large for small lesions that are less than 1000 voxels, Enhanced weight loss proposes a viable method for training small lesions with higher accuracy.

Our cross-validation method was a clear suggestion to train models with equal variance in lesion sizes, which helps to increase the segmentation accuracy when encountered with a new image to test, which will be a common case when introduced in real clinical situations. It also has advantages in validating and comparing the results of published literature as we do not know how the training set was created and the chance of bias is present, but using a training set from our cross-validation method assures the result from the trained model is bias-free.

Cost, more specifically computational expenditure was a major issue for our models. This was because our main image modality, MRI, is a highly resolution-heavy method with a 3-dimensional structure, consuming a large amount of computational memory. There were several reasons behind the increased consumption of computational memory such as the large image size (233 x 197 x 189) or the ambiguity of stroke lesion labels, the deeper layer model was found to be the main contributor. This issue was prevalent when constructing our deep-learning models that share the application of U-nets where the images are convoluted and pooled multiple times going deeper towards the bottleneck, and generally, a higher number of convolutional layers led to a higher number of parameters to learn that increased the computational cost significantly. This could be observed in our transformer models as creating a deeper layer caused the computational memory usage to rise significantly, possibly due to the quadratic nature of attention mechanisms in which each token interacts with all other tokens in the sequence, resulting in a substantial memory and computation overhead [91].

Use of the transformer itself also had a great impact on heavy memory usage. In our diffusion model, installation of Swin transformer in the segmentation encoder roughly increased the computational usage by 55 percent whereas the deeper layer only contributed to 10 percent increase. One of the reasons behind this would be that the diffusion models were sampled at 2D slices individually unlike our transformer models that perform segmentation from the 3D image directly, so the impact of adding a self-attention mechanism innnate to the Swin transformer to the model will be more critical than having additional convolutional layers, as convoluting with 2D images will introduce lesser parameters than in 3D. Therefore, to reduce memory usage, we imported the residual layers on the decoder side to both models to compensate for the heavy memory usage on the encoder side.

Time complexity which is a measure of time consumed for executing algorithms as a function of inputs rather than measured against the memory consumption was another outstanding limitation for our project. For instance, N4ITK bias correction took more than a month to process all our images for training due to its availability in CPU only. The diffusion model also suffered greatly as training up to 1000 epochs took 23 days and more than a week to sample the test images to produce the segmentation as sampling

had to be performed on each of the 2D slices. Although not as extreme as the diffusion model, all our models took at least 3 days to train up to 1000 epochs. Due to this time complexity, there was a huge time gap between modifying the code and viewing the results which slowed down the improvement and iteration process. Utilizing parallelized training can be a solution to this, where the training data can be split and slotted into multiple memories. Unfortunately, the university policy limits the usage of only 2 GPUs at a time but should we gain access to parallel training, the time taken to train our model will be significantly decreased. Reducing the size of our input data could also be a good approach. For instance, rather than using 128 x 128 x 128 image as of now, we can reduce the size of the window to 64 x 64 x 64 for images with a lesion size of less than that. This will definitely reduce both time consumption and computational expense but its segmentation performance must be comparable to the segmentation result of the 128 x 128 x 128 image. Although Only a few studies included time complexity metrics in their results and were often overlooked when implemented on segmentation tasks with complex domains with intricate regions, we believe that including these metrics as part of the evaluation parameters would increase the quality of performance assessment. Reducing time consumption for software system is crucial for its success, and medical lesion segmentation techniques must also aim to decrease the time spent for execution for its success in the commercial field.

Despite its necessity to remove false positive predictions, some concerns have been raised about the credibility in zero filling post-processing technique. Firstly, only the results using NNunet have utilized this method and therefore it is not a universal technique adopted by many methods [87]. There was no paper involving this post-processing method regarding stroke lesion segmentation introduced in the literature review. Practically, converting zero values to 1 will clearly reduce the reliability of the overall dice score as it is the only factor that changed; no physical changes to the segmented lesion have been made which corresponds to no change in Hausdorff distance after zero filling. Our zero filled value of 0.7221 in fact shows higher dice score than the current state-of-the-art method for ATLAS 2.0 of 0.7098 whereas the dice score with the actual post-processed lesion is only 0.6581. It is hence very difficult to propose with credibility that this is the dice score we get, and should consider using zero filling after further studies have been performed.

6 Functional MRI

6.0.1 Methodology

Functional MRI analysis is performed using multiple softwares such as SPM12 (Statistical parametric mapping) fMRI toolkit, FSL and CONN toolbox, all in most recent versions. Specifically, FSL and SPM have been vastly used for preprocessing of fMRI images whereas CONN has been used for first (individual) and second (group) level analysis of functional MRI. While using CONN toolbox, the approach to fMRI analysis varies between two conditions, resting state or task-based, suitable for unique conditions required for each of them.

6.1 Resting state fMRI

Resting state fMRI is a method of functional imaging to investigate neural connectivity between brain regions of interest. As the name 'resting state' represent, a patient is put under an undisturbed state without any explicit form of testing. The absence of task makes resting state fMRI a powerful tool for evaluating brain connectivity in patients who cannot complete motor-related or cognitive tasks [92].

We used a dataset of 25 patients from our collaborators from University of Leuven, Belgium. These patients developed upper limb sensorimotor impairments afer stroke, recruited from the University Hospital Leuven and the University Hospital St-Luc Brussels. Both T1w images and fMRI were obtained between 5 to 9 days after the stroke onset.

Since in-house fMRI data from our collaborators is in a raw form, sufficient preprocessing to both Anatomical and functional images had to be applied for a reliable analysis. Hence, preprocessing step was required to reduce the effect of various artifacts that can potentially compromise the interpretation during functional MRI fluctuations such as physiological or motional instances, and to achieve a higher signal-to-noise ratio. Moreover, preprocessing provides better correspondence between the brains of different subjects so that the analysis reliability is increased for the entire dataset.

6.1.1 Preprocessing

The preprocessing protocol for resting-state functional MRI is given below:

- **Slice timing correction**: An initial and essential step of pre-processing of resting-state images, as it is based on correlations between time series of different voxels, and existing noise or artifacts can disrupt the correlation between each voxel resulting in a faulty analysis [93]. A signal phase is adjusted to address temporal sampling discrepancies within dynamic causal modeling. The prediction is then calculated by

incorporating model parameters and the corresponding sampling time series [94]. Our images have been corrected with 4800 ms repetition time (TR) as suggested by the acquisition protocol using SPM12 slice-timing correction (STC) procedure.

- **Spatial alignment**: The spatial alignment also known as motion correction aims to remove movement artifacts by registering functional MRI images to a reference image (first scan of the first session). This is done by estimating the derivatives of the deformation field related to head motion and resampling the functional data with respect to the deformation field of the reference image. Our images have been processed using SPM12 realign&unwarp procedure under Susceptibility Distortion Correction (SDC) scheme where the absolute deformation state of the reference image is corrected along the phase-encoded direction [95]. The figure below depicts a head motion in terms of translation and rotation of a resliced and aligned image. By convention, it is recommended to remove any runs that moved more than a size of 1 voxel which in our case is 1 mm or if there is a sudden spike of more than half a voxel size which is 0.5 mm. Only subjects that fall under these criteria have proceeded to the next stage.

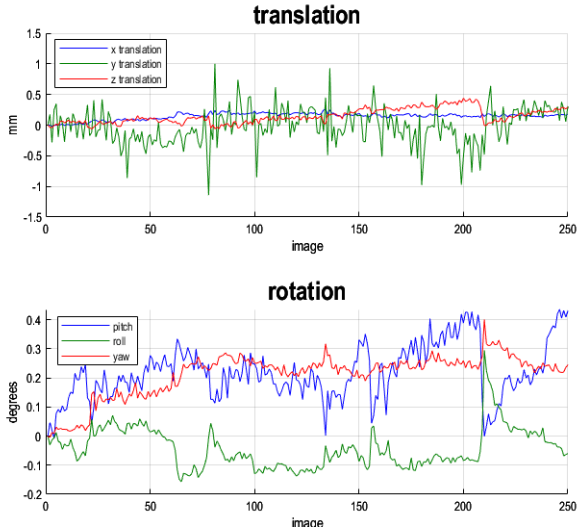


Figure 29: Translation and rotation of the patient during spatial alignment

- **Coregistration**: Both functional and anatomical images for the subject are overlaid to maximize the mutual information, since the heads may be moved in different manner in the acquisition of anatomical and functional images. The mean functional image is then coregistered to high resolution anatomical image, and corresponding functional images are resegmented to align with the reference image from spatial alignment. SPM12 Coregistration procedure is used for our dataset.

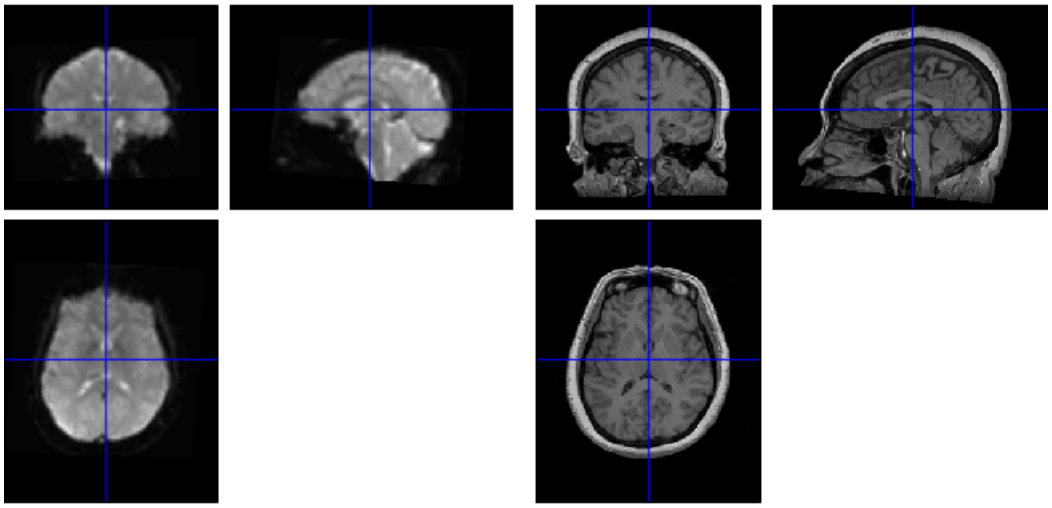


Figure 30: Functional (left); and Anatomical (right) images after coregistration

- **Normalization**: Here, both functional and anatomical images are aligned and spatially registered to a standard space to decrease variability between subjects. The most commonly used reference templates are Talairach brain and Montreal Neurological Institute (MNI) [96]. For our dataset, we used MNI 152 1mm template for anatomical images and MNI 152 2mm template for functional images. These have been re-normalized to MNI 152 1mm overall afterward through estimate&reslice procedure to [180, 216, 180] template size with 1mm voxel spacing; [-90 -126 -72;90 90 108] coordinate-wise. SPM12 Normalization procedure is used.

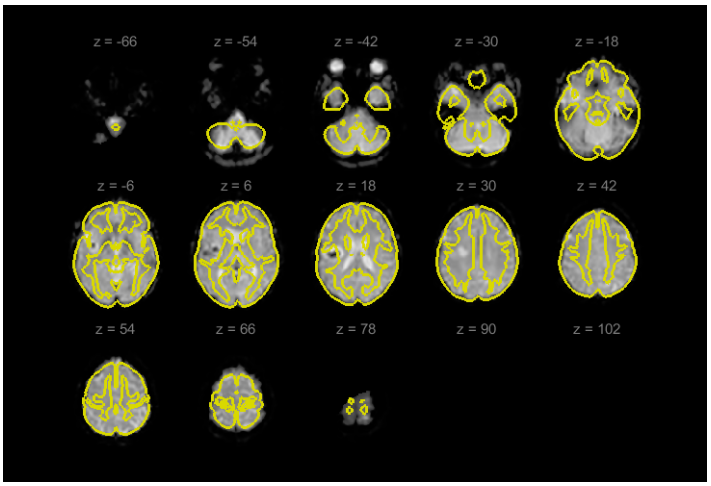
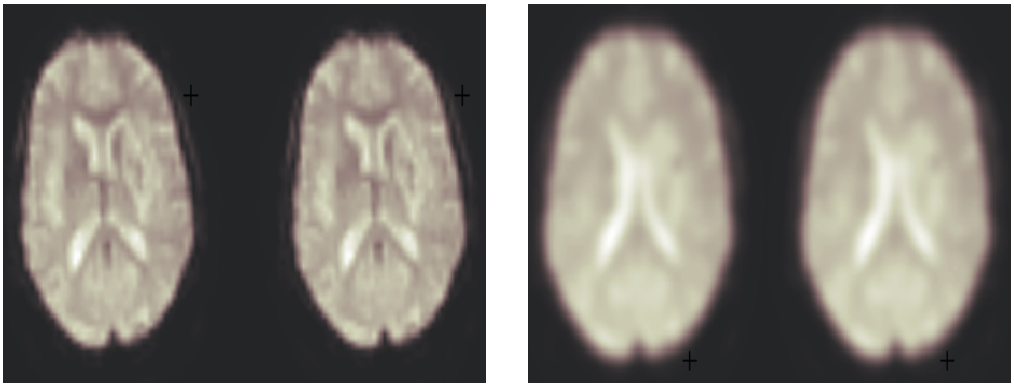


Figure 31: MNI template (yellow) overlayed onto the functional image

- **Smoothing**: Smoothing is a process of removing high-frequency information to increase the BOLD signal-to-noise ratio. Spatial convolution is utilized with a Gaussian kernel of 2mm full-width half maximum (FWHM) to the original image that replaces the intensity of each voxel by a weighted average of the neighbor voxels; reducing residual variability [97]. For our analysis, FWHM of 2mm has been used following the convention that the kernel must be equal to or less than twice the size of the voxel. SPM12 Smoothing procedure has been used.



(a) Before smoothing (b) After Smoothing

Figure 32: Before (fig 4a) and After (fig 4b) smoothing has been performed

- **Denoising**: Denoising is a process used to remove or reduce various sources of noise and artifacts allowing for a cleaner and more accurate representation of brain activity. CONN toolbox combines linear regression and temporal band pass filtering to tackle this issue. For linear regression, factors are identified as potential confounding effects to the estimated BOLD signal are estimated and removed separately for each voxel and for each subject and functional run/session. Next, temporal frequencies below 0.008 Hz and above 0.09 Hz are filtered out from the BOLD signal, emphasizing slow-frequency fluctuations while simultaneously reducing the impact of noise sources such as physiological signals or head motion. Discrete cosine transform windowing operation is used to minimize edge-related artifacts [98].

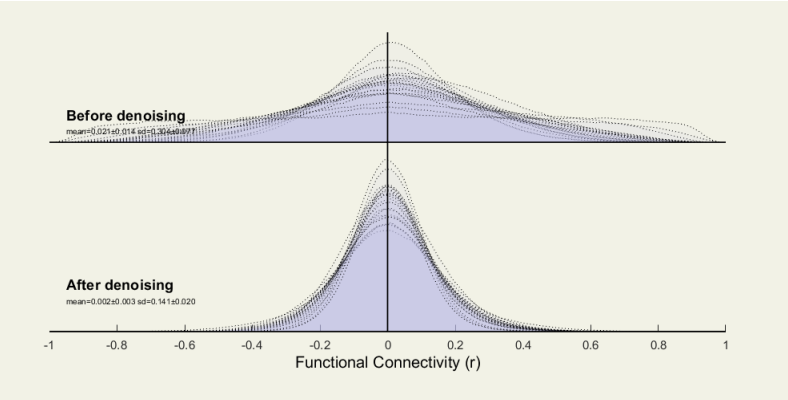


Figure 33: Effect of denoising in functional connectivity variability

6.1.2 Seed-Based Connectivity maps

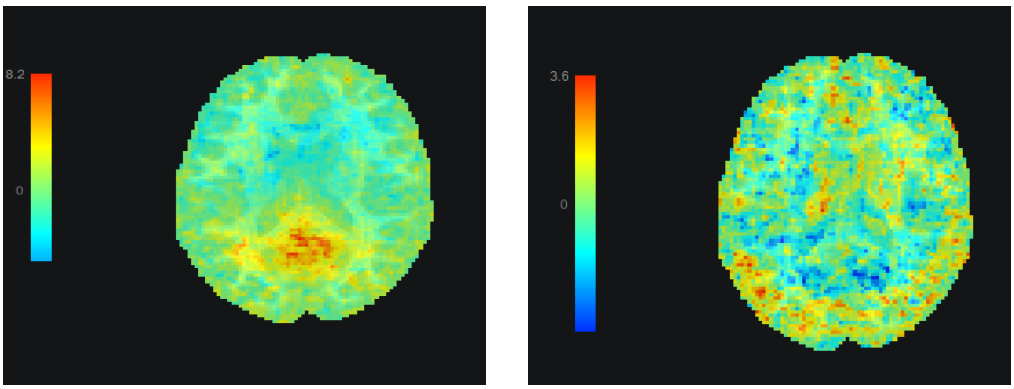
The selected subjects have then gone through BOLD signal analysis for functional connectivity analysis. CONN toolkit for resting state fMRI analysis exploits SBC (Seed-based connectivity) measures to construct a functional connectivity map to characterize the connectivity patterns with a pre-defined seed or ROI (Region of Interest). Such type of analysis is used when we are interested in a few or individual subjects.

The SBC maps represent the intensity of functional connectivity between the seeded ROI and every other voxel in the brain. It is computed as the Fisher-transformed bivariate correlation coefficients between an ROI and each other voxel both measured in terms of BOLD time series. The equation is as follows:

$$r(x) = \frac{\int S(x,t)\, R(t)\, dt}{\left(\int R^2(t)\, dt \int S^2(x,t)\, dt\right)^{1/2}}$$

$$Z(x) = \tanh^{-1}(r(x))$$

where S is the BOLD time series at each voxel processed to zero mean, R is the measure of average BOLD time series of specific an ROI, r is the spatial map of Pearson correlation coefficients (PCC), and Z is the SBC map of Fisher-transformed correlation coefficients for the chosen ROI. The connectivity pattern with an individual seed falls under the same area, indirectly signifying the level of homogeneity within this region. It also directly measures inter-regional connectivity strength.



(a) PCC (b) LPC

Figure 34: BOLD distribution of SBC maps in posterior cingulate cortex (left), and lateral parietal cortex (right)

Utilizing our knowledge of SBC, Seed-to-voxel based analysis was performed. We selected regions related to the primary motor cortex and the sensorimotor cortex; precentral gyrus, postcentral gyrus and lateral sensorimotor network as our seeded ROI under the assumption that the level of activation of these motor cortex would affect the recovery outcome of the stroke patients. The ROIs were defined from Harvard-Oxford atlas, a probabilistic atlas divided into 96 cortical and 21 subcortical structural areas, and independent component analysed brain networks derived from human connectome project (HCP). The voxel of interest was represented as the greatest connectivity value inside the largest cluster of voxels. This cluster uses inference based on Gaussian Random Field theory to obtain a statistical parametric map of T-values estimated using a General Linear Model [99].

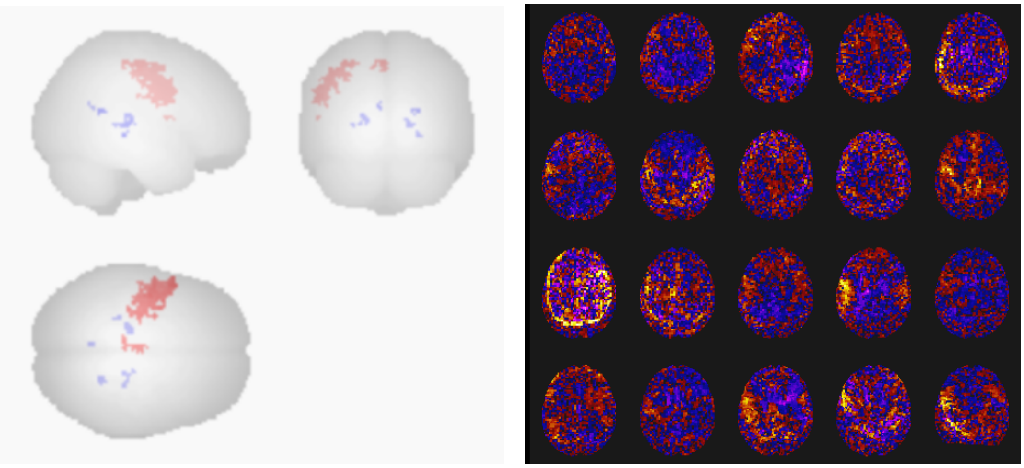


Figure 35: The cluster generated from statistical analysis in the left precentral gyrus, red showing positive and blue showing negative connectivity: beta = 0.65, T-value = 6.25, p-one-sided = 0.00003 (left); The QA montage of subjects selected for SBC second level analysis in the left precentral gyrus (right)

The single voxel of coordinate [-54, +02, +26] was selected for giving the highest T-value. The connectivity values at the single voxel were collected for each ROI.

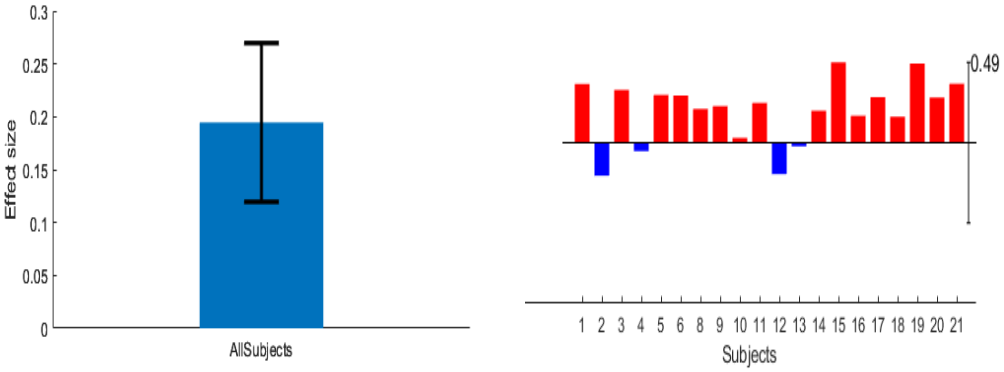


Figure 36: Box plot of effect size (T-value) from arbitrary subject (left); and connectivity value of the corresponding subject (right)

6.1.3 ROI-to-ROI Connectivity matrices

ROI-to-ROI connectivity metrics measure the connectivity between two or more ROIs from a set of brain regions. These metrics follow identical orientation and properties as the SBC measures but prevent the SBC asymmetry between seeds (ROIs) and targets (voxels). It is widely used when we are interested in the entire networks of connections. The equation is:

$$r(i,j) = \frac{\int R_i(t) R_j(t)\, dt}{\left(\int R_i^2(t)\, dt \int R_j^2(t)\, dt\right)^{\frac{1}{2}}}$$
$$Z(i,j) = \tanh^{-1}(r(i,j))$$

where R is the BOLD time-series within each ROI, r is a matrix of correlation coefficients, and Z is the matrix of Fisher-transformed correlation coefficients.

Functional connectivity measure of each region is obtained from CONN toolkit first level analysis. Following the second-level analysis, ROI to ROI (region of interest to region of interest) analysis can be drawn from the functional connectivity network matrix display where the functional connectivity between each cluster, a bundle of or individual functional brain regions, is visualized in a circular ring. According to the T-value, the regions of interest have been identified. We selected the four highest T-value connectivities; between precuneus cortex and posterior cingulate cortex, between contralesional intraparietal sulcus (IPS), and finally between each ipsilesional IPS and supramargianl gyrus (SMG). The regions identified from the seed-to-voxel analysis have also been selected for comparison.

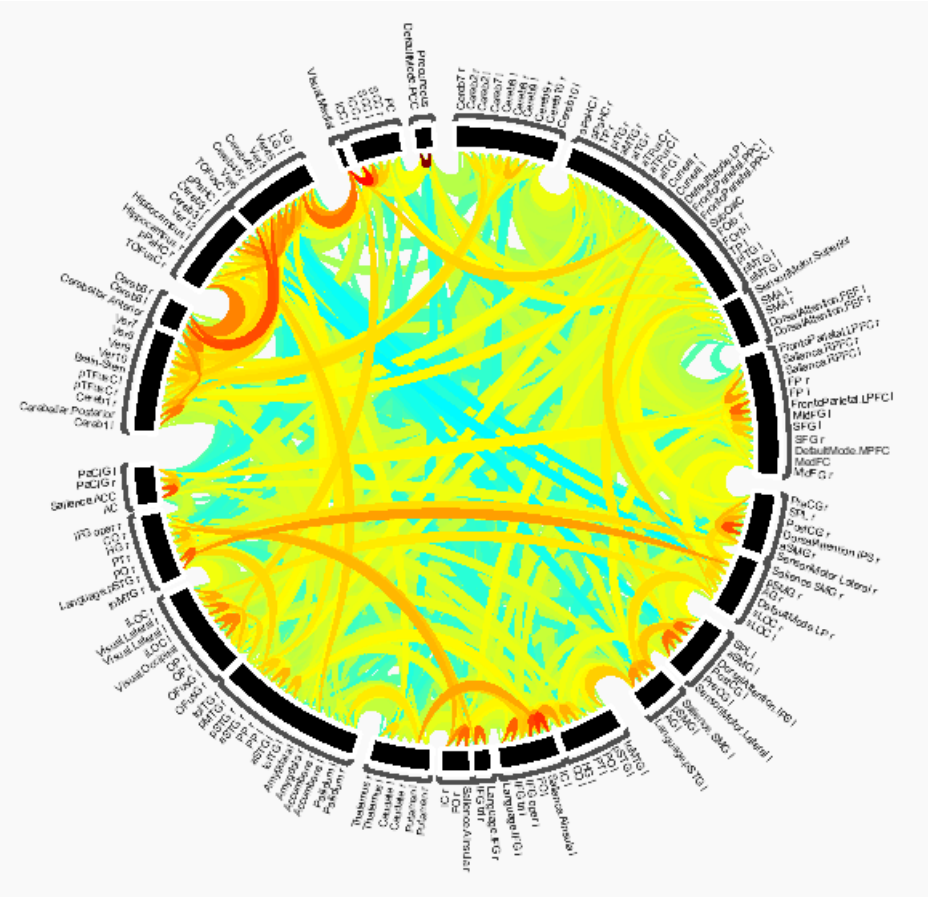


Figure 37: ROI to ROI Connectivity matrix showing relationship between 132 pre-defined ROIs

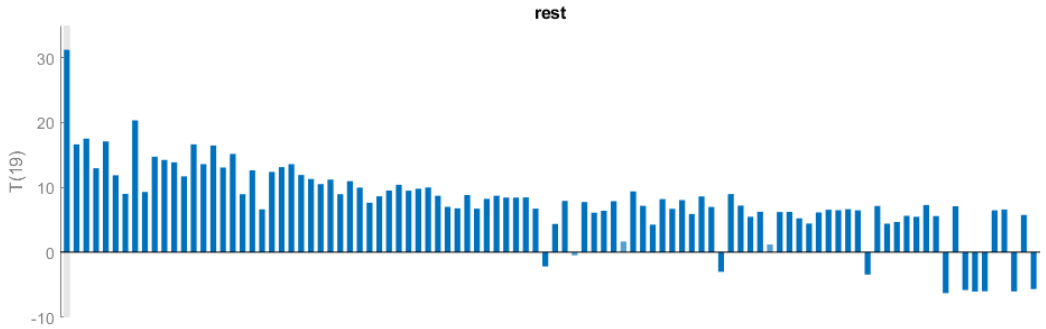


Figure 38: Effect size (T-value) against all ROIs

In addition, to analyse the difference between the ipsilesional and contralesional connectivity measures, all lesions have been transferred to the right side of the brain, and corresponding functional MRI images have been processed to match the changes accordingly.

6.2 Task based fMRI

Unlike resting-state fMRI which measures spontaneous fluctuations in brain activity during rest, task-based functional MRI is used to investigate brain activity associated with specific cognitive tasks or experimental stimuli which is performed by presenting participants with structured tasks while measuring changes in BOLD signals in the brain [100]. Specifically, a large series of images is acquired rapidly with changes in brain activity between two or more pre-defined tasks, and by correlating the signal time course in each voxel of the slices with the known time course of the task, it is possible to identify the voxels that show significant changes associated with the brain function in which we are interested [101].

As task-based fMRI requires specific tasks to be performed, its logical and sophisticated task design is crucial in delivering the desired analysis. We obtained 12 images from different patients taken from the Auckland hospital and distributed by the research group at the University of Auckland. The set of processes entitled to hand-grip tasks is outlined below:

1. Have the patient lie down on the MRI (Siemens 1.5T)
2. place a dynamometer on each hand and secure the affected hand with a bandage
3. Zero the system and set the maximum for the affected hand by asking the patient to produce maximum grip force and hold while it is recorded.
4. Begin recording the MRI and the patient is asked to perform a four-phase task for each trial
    - **Plus Only**: The resting state before the trial begins with relaxed hands
    - **Target**: The patient grips the dynamometer with the affected hand to turn the arrow on the screen to green. The arrow turns red when gripped with the unaffected hand which immediately skips to the relaxation phase
    - **Hold**: The phase of applying the grip for a few seconds until the instructor says to let go
    - **Relax**: Relaxation phase of removing force from the hands to get ready for the next trial
5. Consecutively repeat the trial 45 times

These data were then preprocessed following the same steps as discussed in the resting state fMRI section. In the CONN toolbox, the test conditions were constructed regarding the four tasks for each subject unlike the resting state fMRI in which no test conditions were outlined.

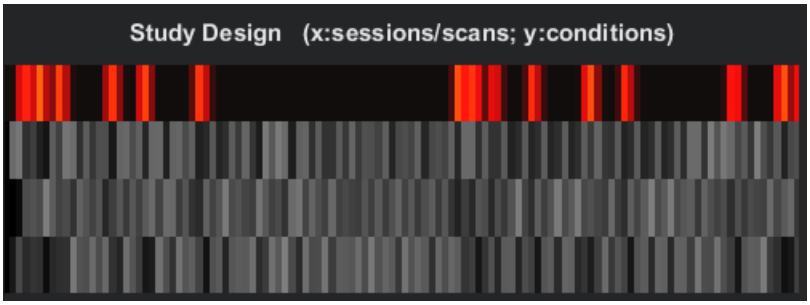


Figure 39: Test condition of tasks represented by each row. From the top; Hold, Plus only, Relax and Target

6.2.1 Generalized Psycho-Physiological Interaction matrices

Generalized psycho-physiological interaction, also known as gPPI, measures the level of task-regulated connectivity between two ROIs, such as changes in functional association strength that vary with the external factors. gPPI is computed using a separate multiple regression model for each target ROI timeseries. The predictors of each model are represented by several classes:

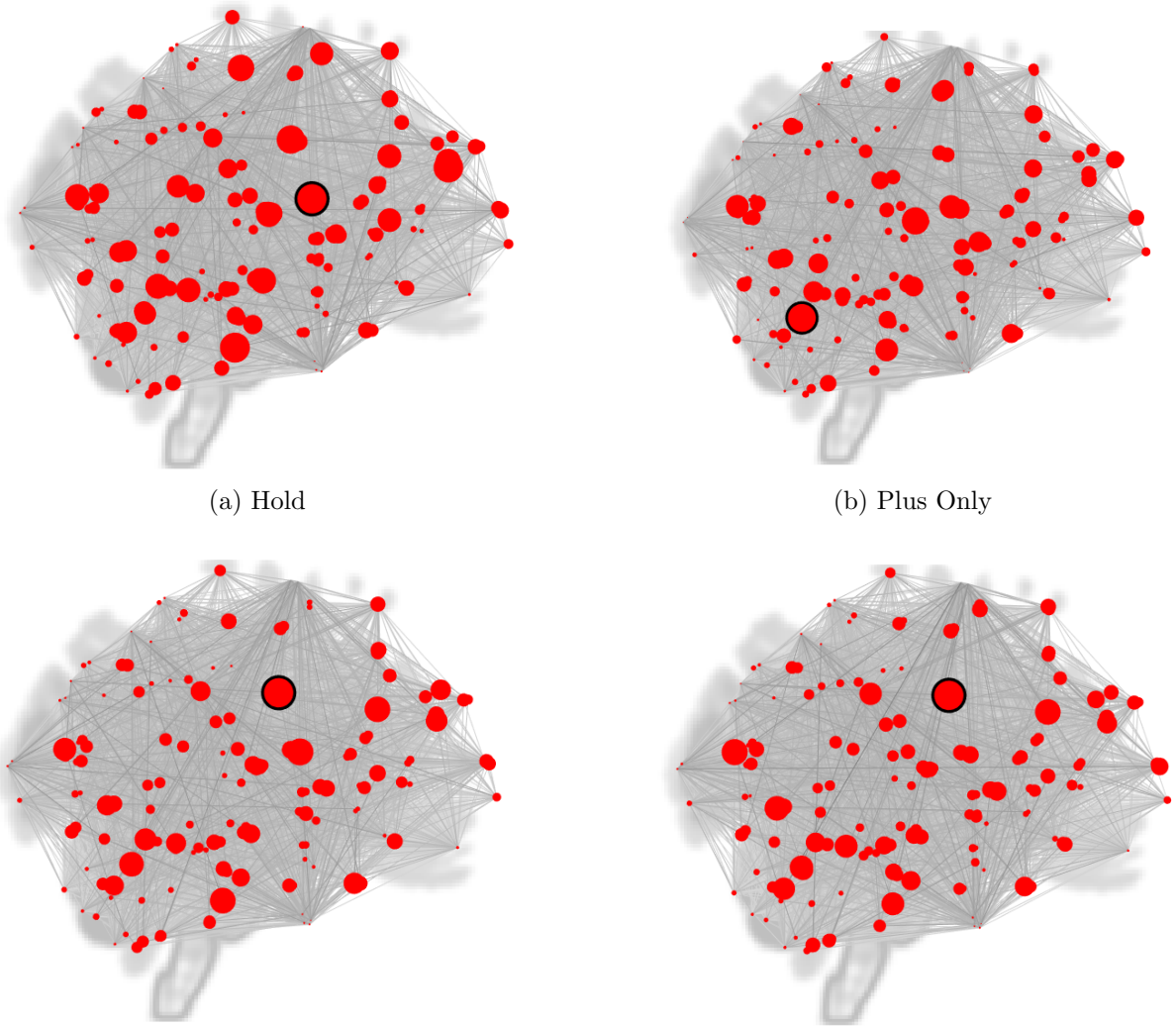
- All of the task strengths convolved with a canonical hemodynamic response function. This is a major psychological factor in interaction terminology
- BOLD time series of each seeded ROI which is a major physiological factor
- The product of the two factors that provides generalized combination of the psycho-physiological factors.

As a result, these models are represented by the regression coefficients that define the interaction terms described above.

$$
\begin{aligned}
R_j(t) &= \left(\beta(i,j) + \sum_k \gamma_k(i,j) h_k^*(t)\right) R_i(t) \quad + \sum_k \alpha_k(i,j) h_k^*(t) + \epsilon_{ij}(t) \\
h_k^*(t) &= h_k(t) * f(t) \\
\alpha_k(i,j), \beta(i,j), \gamma_k(i,j) &| \min \int \epsilon_{ij}^2(t)\, dt
\end{aligned}
$$

where R is the BOLD time series at each ROI, h is the task time series convolved with a canonical hemodynamic response function f. Gamma represents the gPPI matrix of regression coefficients for each condition, estimated simultaneously with alpha and beta parameters using an Ordinary Least Squares (OLS) solution to the linear model described above.

Since gPPI is a more data-driven approach that does not rely on pre-defined ROIs much like conventional ROI to ROI analysis in resting state fMRI, it is effective in examining task-modulated connectivity across the whole brain to allow for a comprehensive exploration of connectivity patterns. Following this aspect, graph theory analysis was utilized.



(a) Hold

(b) Plus Only

(c) Relax

(d) Target

Figure 40: Graph theory analysis displays on the entire dataset, nodes represented as red dots and edges as grey lines: correlation coefficient < 0.05, one-side test positive direction

The ROI of the largest T-value was selected for each task and the centrality information was recorded. Betweenness centrality identifies nodes that serve as bridges or intermediaries, closeness centrality identifies nodes with short average distances to all other nodes, and finally, eigenvector centrality that identify nodes that are connected to other important nodes in the network. All three centralities have been selected as a feature for prediction modeling.

6.3 Prediction model

In creating a prediction machine learning model, random forest (RF) and support vector machines (SVM) have been used. Random forest is a widely used ensemble learning method for classification and regression and uses decision tree based algorithms [102]. It is composed of multiple decision tree that repeatedly partition the feature space into smaller regions and thus make predictions based on the average value for each region. It has been widely adopted in detecting ischemic stroke lesions using neuroimaging data and estimating the outcome [103]. Although SVM is also involved in classification and regression tasks much like random forest, it is particularly effective in handling both linear and non-linear classification tasks by mapping the input into high-dimensional spaces to find the optimal hyperplane that separates data into two or more classes [104]. Both RF and SVM have been chosen for our classification for their robustness and flexibility.

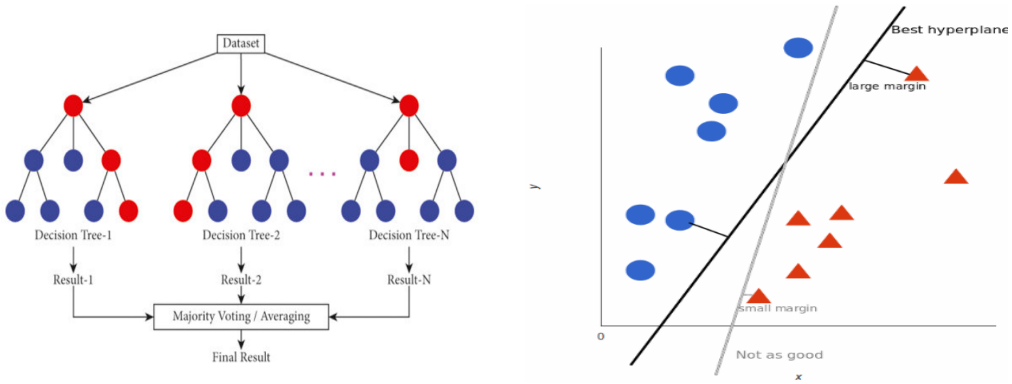


Figure 41: Diagram of random forest (left); and support vector machines (right)

For predicting the recovery of the patient, action research arm test (ARAT) score has been used with the first recording taken a week after the stroke onset referred to as baseline and the second recording taken 6 months later. ARAT is a test that measures the patient's capabilities to grasp, reach, release and manipulate objects faced during daily activities [105]. The test comprises 19 items scored on a grading system from 0 to 3 where score of 3 suggests the patient can perform instructed action with normal movement. Hence, a total of 57 points indicate a sufficient recovery of the affected limb regarding movement control with less than 50 points indicating the presence of notable motor impairment [106]. The ARAT score is well-known for quantifying the degree of upper-extremity control for its strong psychometric properties such as change in sensitivity of people with stroke [107]. It is also widely recognized as a reliable measurement in upper-extremity rehabilitation [108]. The ARAT score was divided into mild (class 2; score above 41), moderate (class 1; score between 41 and 21) and severe (class 0; score below 21) [109]

To boost and compare the prediction outcomes, we used the results we collected as much data as possible and separated them into three parts.

- **Structural MRI data**: The prediction results collected from structural T1 MRI

data using segmentation from chapter 1.

  - **lesion volume**: lesion volume calculated in voxels
  - **Lesion overlap**: an overlap between the predicted lesion and the brain regions of interest;precentral gyrus and postcentral gyrus. Calculated as dice score (percentage) of overlap.

- **Functional MRI data**: functional connectivity intensity data through various connectivity measurements.

  - **Seed-to-voxel based values**: brain regions related to primary motor cortex; precentral gyrus, postcentral gyrus and sensorimotor cortex.
  - **ROI to ROI connectivity**: Selected regions with high T-values; IPS, PCC, aSMG and the primary motor cortex regions.
  - **Graph theory results**: specific to task fMRI with centrality values; betweenness, closeness and Eigen.

- **Clinical data**: Clinical data collected throughout the fMRI tests

  - **Generic demographics**: Age of stroke onset, gender, hand dominance, lateralization (location of stroke lesion left or right) and type of stroke (ischemic or hemorrhagic)
  - **Adult assisting hand assessment (Ad-AHA)**: a Rasch based performance scale that measures the effectivity of the affected hand while performing bimanual tasks with varying scores between 0 to 100 with a score of 100 indicating high bimanual ability under normal patterns [110]. This has been acquired 6 months after the stroke onset
  - **ABILHAND questionnaire**: scores are judged by the patient on the scale of 0 point (impossible), 1 point (diffcult), and 2 point (easy), and these raw scores are converted to logit-scores to capture any clinically distinguishable differences [111]. This has been acquired 6 months after the stroke onset

6.3.1 Metrics

Metrics for measuring the performance of the prediction model are as follows:

- **Accuracy**: Measures the proportion of correctly classified instances out of the total instances in the dataset and calculated as the ratio of the number of correct predictions to the total number of predictions. It is worthwhile to note that the accuracy parameter may not be suitable for imbalanced datasets where one class dominates the others [112].

- **Precision**: Measures the proportion of true positive predictions out of all instances predicted as positive by the model and is mainly used for quantifying the model's ability to avoid false positives. Precision is particularly important in cases where the cost of false positives is high.

- **Recall**: Also known as sensitivity or true positive rate and measures the proportion of true positive predictions out of all actual positive instances in the dataset. It quantifies the model's ability to minimize false negatives and is usually assessed alongside precision under scenarios where the balance between false positives and false negatives is critical such as classification tasks [113].

- **F1-Score**: A harmonic mean of precision and recall that provides a single metric that balances both precision and recall. It is particularly useful when we want to compare the performance of different models or in cases where an uneven class distribution is present.

6.3.2 Resting state fMRI results

6.3.2.1 Contralesional vs Ipsilesional

We used the ROI to ROI connectivity values on the primary motor cortex to study the contralesional and ipsilesional contributions to ARAT score recovery prediction. Both RF and SVM models were performed with 0.4 test to train split ratio with a linear kernel used for SVM model.

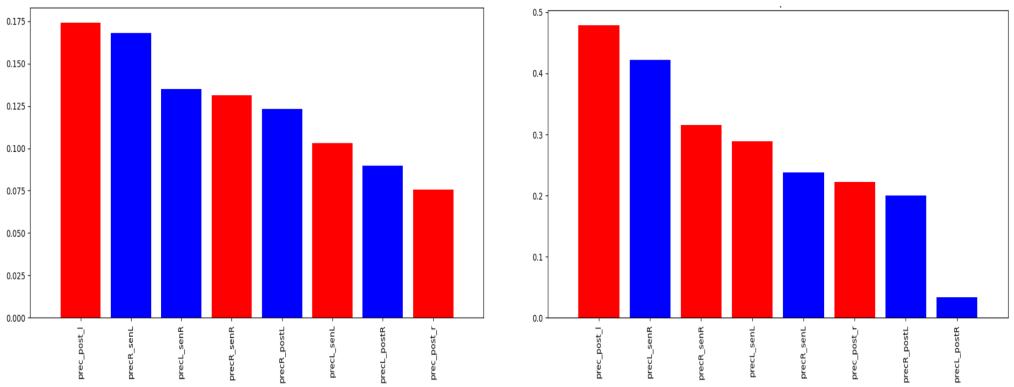


Figure 42: Feature importance diagram of random forest (left); and support vector machines (right) of the primary motor cortex and sensorimotor cortex

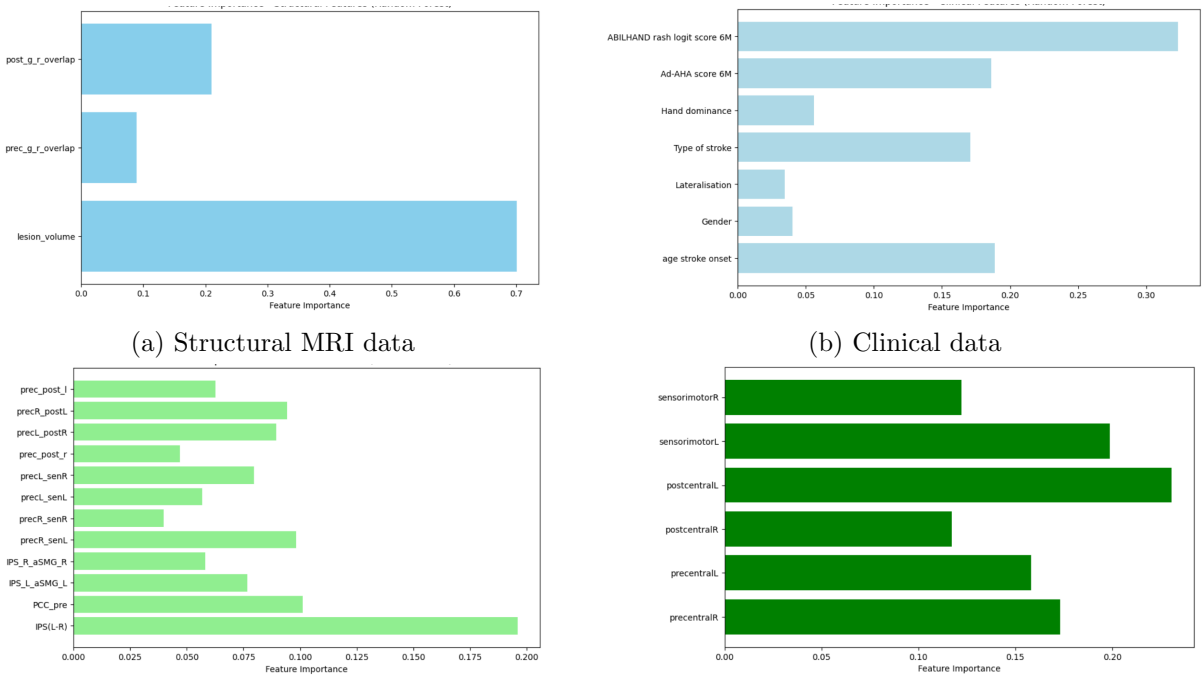
The red bar plots represent the ipsilesional contributions to prediction and the blue bar plots represent the contralesional contributions. From both the RF and SVM analysis, it shows that the connectivity between the precentral gyrus and postcentral gyrus on the left side scored the highest.

Model	Class	Accuracy	Precision	Recall	F1-Score
Random Forest	0	0.67	0.33	0.50	0.40
Random Forest	1	0.67	0.00	0.00	0.00
Random Forest	2	0.67	0.83	0.83	0.83
Support Vector Machine	0	0.56	0.33	1.00	0.50
Support Vector Machine	1	0.56	0.00	0.00	0.00
Support Vector Machine	2	0.56	1.00	0.50	0.67

Table 6: RF and SVM results in predicting with primary motor cortex and sensorimotor cortex connectivity values

It is to note that for both models class 1 (moderate) showed zero values throughout all metrics meaning no prediction has been made.

6.3.2.2 Overall outcome



(a) Structural MRI data

(b) Clinical data

(c) Functional MRI data (ROI)

(d) Functional MRI data (SBC)

Figure 43: Feature importance analysis using Random forest on structural MRI data, functional MRI (ROI and SBC) data and clinical data

From the figure above, we selected the highest contributors from each dataset. Lesion volume, IPS (L-R), left postcentral gyrus and ABILHAND at 6 months have been chosen from structural MRI, functional MRI (ROI), functional MRI (SBC) and clinical datasets respectively as a primary feature in predicting the recovery after stroke, especially for the SVM analysis.

	Accuracy	Precision	Recall	F1-Score
Structural	0.67	0.53	0.67	0.59
Functional (ROI)	0.56	0.74	0.56	0.56
Functional (SBC)	0.56	0.42	0.56	0.48
Functional (combined)	0.67	0.50	0.67	0.57
Clinical	0.89	0.79	0.89	0.84
Structural + Clinical	1.00	1.00	1.00	1.00
Functional + Clinical	0.78	0.78	0.78	0.75
Structural + Functional + Clinical	0.78	0.72	0.78	0.72

Table 7: RF results of various combinations showing performance of each

	Accuracy	Precision	Recall	F1-Score
Structural	0.67	0.44	0.67	0.53
Functional (ROI)	0.67	0.44	0.67	0.53
Functional (SBC)	0.67	0.44	0.67	0.53
Functional (combined)	0.67	0.44	0.67	0.53
Clinical	0.78	0.83	0.78	0.79
Structural + Clinical	0.78	0.70	0.78	0.73
Functional + Clinical	0.56	0.72	0.56	0.61
Structural + Functional + Clinical	0.78	0.83	0.78	0.79

Table 8: SVM results of various combinations showing performance of each

To note that for both SVM and RF analysis, the combination of structural and clinical scored the highest.

6.3.3 Task based fMRI results

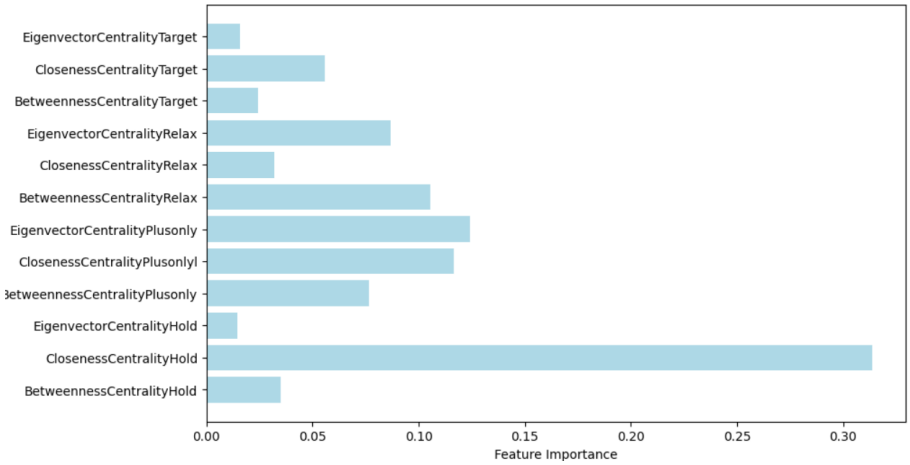


Figure 44: Feature importance of the centrality values obtained from the graph theory analysis

Upon obtaining the accurate centrality values for each task, we chose the highest T-value region to extract these values; Hold = right central opercular cortex, Plusonly = right cerebellum, Relax and Target = left lateral sensorimotor cortex. The highest feature importance for the closeness centrality at Hold position has been observed and subsequently used for primary feature in SVM analysis. The RF and SVM analysis has been performed keeping the same setup as resting state analysis.

	Accuracy	**Precision**	**Recall**	**F1-Score**
Structural	0.60	0.60	1.00	0.75
Functional	0.60	0.60	1.00	0.75
Clinical	0.60	0.60	1.00	0.75
Structural + Clinical	0.80	0.80	1.00	1.00
Functional + Clinical	0.60	0.60	1.00	0.75
Structural + Functional + Clinical	0.60	0.60	0.78	0.79

Table 9: RF and SVM results of various combinations showing performance of each for task fMRI data

In addition to RF and SVM, we performed principal component analysis (PCA) to reduce the dimensionality. we found that only the combination of structural and clinical results increased the accuracy similar to the findings in resting state analysis.

6.4 Discussion

From the resting state analysis in table 5 and 6, a combination of data rather than singular data showed higher accuracy for both RF and SVM. This could mean that combining the two types of information provides complementary information that allows the model to capture a broader range of factors that influence stroke recovery, or a possibility of creating a more robust model compared to using singular data. Hence, this indicates that considering multiple aspects of stroke pathology and patient characteristics leads to better predictions of recovery. For both tables, Clinical data scored the highest even on singular data, and this could be majorly influenced by the ABILHAND score as indicated by the feature importance analysis. This strong relationship could be derived from both taking measurements of physical motor function although ABILHAND focuses on the overall capacity to use the hands in functional tasks required for daily living whereas ARAT focuses on motor control and coordination of the upper extremities [105] [111]. Since ABILHAND provides a wider assessment of manual motor function, it is intuitive to suggest that it creates a firm complement with the ARAT score boosting the prediction.

From both resting-state and task-based analysis, the highest accuracy was observed in a combination of structural and clinical data. This is notable because the lesion volume was previously found to be a less reliable predictor of stroke recovery as the lesion location and the amount of damage of parts of hand movement as opposed to lesion volume itself was considered to be a more dependent parameter assessed with structural imaging [114].

However, from our feature importance analysis, lesion volume had a higher influence in prediction compared to the lesion location evaluated by the overlap between the lesion and the components of the primary motor cortex and sensorimotor cortex. This directly challenges the idea that the lesion located near the motor cortex would influence the motor impairment at the upper extremities more than the size of the lesion. Although arguments such as the use of a small dataset can be raised, cautious proposal of the strength of lesion volume in predicting ARAT score after 6 months can be made.

The analysis for comparing between contralesional and ipsilesional connectivity showed an uncorrelated manner although ipsilesional precentral-postcentral gyrus connectivity depicted the largest influence in ARAT prediction for both RF and SVM analysis. This is intuitive as direct damage to a brain may disrupt the functional connectivity between the affected lesion and the other networks and will be more severe if the area is associated with motor function or other cognitive domains. Hence, the degree of damage at the primary cortex on the same side of the lesion will be directly proportional to the severity of the motor impairment shown by ARAT score. To note, our findings suggest the correlation between the precentral and postcentral gyrus is higher than that of other regions because we can also observe from figure 42 the precentral gyrus and postcentral gyrus on the contralesional side ranked second which may suggest that the contralesional side is compensating the work required by the ipsilesional side and no other connections take over.

Some limitations were raised during the analysis mainly from the data we used. Firstly, our data only had the initial measurement and did not contain fMRI recordings after 6 months which is the prediction target period. A study involving a direct comparison between the stroke onset and 6 months after, known as a longitudinal study, was unavailable. Longitudinal studies involve data such as lesion volume, where the difference between initial and final lesion volume or overlap between the seed and the prediction obtained from segmentation is taken as a valuable input [115]. Relying heavily only on the initial data hence decreased the prediction power significantly. Data bias was another issue that hindered accurate prediction. The ARAT score had a bias toward mild (class 2) motor impairment for task-based fMRI with 10 out of 12 patients showing mild motor impairment. This explains why both RF and SVM on this dataset gave the same values for nearly every combination reducing the reliability significantly.

Another concern raised regarding task-based functional MRI was an unclear test procedure. Following the complexity, the data recorded may contain faulty data that affects the connectivity values. The test was recorded 45 times consecutively to reduce human error as much as possible but there is always a possibility of subtle artifacts. This inherent drawback may be a reason why task-based fMRI performs much poorer than resting-state fMRI, at least in our case. A large dataset or simplified test will be key to closing this gap.

7 Future Improvements

Although this study provided valuable insights into stroke lesion segmentation and recovery prediction using deep learning and machine learning techniques, there are several improvements and explorations to be made:

- **Incorporation Multi-Modal Imaging:** Integrating additional imaging modalities such as diffusion-weighted imaging (DWI), CT, FLAIR, and T2 could enhance the accuracy of lesion segmentation and provide richer information for recovery prediction.
- **Use of other learning methods:** Creating segmentation models using unsupervised or semi-supervised will provide a reliable solution when no lesion labels are provided.
- **Access to larger dataset:** Gaining access to larger fMRI dataset with longitudinal recordings will deliver much accurate and reliable recovery prediction
- **Computer aided diagnosis software:** Creating fully automatic segmentation software using Kubernetes and Amazon S3 as a cloud service with a friendly user interface will further assist clinicians with their diagnosis

8 Conclusion

In conclusion, this thesis has contributed to advancing our understanding of stroke lesion segmentation and recovery prediction through the integration of deep learning and machine learning techniques with functional magnetic resonance imaging (fMRI) data. After developing and evaluating novel methodologies, several key findings and implications have emerged:

- **Accurate Stroke Lesion Segmentation:** Our developed deep learning model, Enhanced ensemble method, has demonstrated promising performance comparable to the state-of-the-art method, Strengthening the reliability of the deep learning models for diagnosis.
- **Effective Recovery Prediction:** Machine learning algorithms, including random forest and support vector machine (SVM), have shown competence in predicting stroke recovery outcomes based on functional MRI and clinical data to suggested that lesion volume can be a powerful factor in prediction. It will assist in facilitating early prognosis and treatment planning.
- **Limitation:** Some minor limitations were highlighted that affected analysis of our predictions such as lack of dataset for longitudianl study or unclear test procedure. These issue can however be resolved as more data is acquired by our institute with clear guideline, and something that can be addressed and resolved in the near future under supervision.
- **Future Directions:** While this thesis has made valuable contributions, several improvements exist such as multi-modal imaging integration, creating a model with other methods, access to abundant dataset and making a computer aided diagnosis software

In the end, this thesis underpins the importance of interdisciplinary collaboration in the fields of medical imaging where machine learning, deep learning and clinical neuroscience are incorporated together in order to advance the research and improve patient care. By harnessing the power of computational methods and imaging technologies, we hope to strive towards delivering more precise, personalized, and effective diagnosis for stroke patients and ultimately enhancing their quality of life.

References

[1] Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, and Shi-Min Hu. Visual attention network. *Computational Visual Media*, 9(4):733–752, 2023.

[2] Junde Wu, Rao Fu, Huihui Fang, Yu Zhang, Yehui Yang, Haoyi Xiong, Huiying Liu, and Yanwu Xu. Medsegdiff: Medical image segmentation with diffusion probabilistic model. *arXiv preprint arXiv:2211.00611*, 2022.

[3] Ganesh Chauhan and Stéphanie Debette. Genetic risk factors for ischemic and hemorrhagic stroke. *Current cardiology reports*, 18(12):124, 2016.

[4] Elaine L Miller, Laura Murray, Lorie Richards, Richard D Zorowitz, Tamilyn Bakas, Patricia Clark, and Sandra A Billinger. Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient: a scientific statement from the american heart association. *Stroke*, 41(10):2402–2448, 2010.

[5] Tal Geva. Magnetic resonance imaging: historical perspective. *Journal of cardiovascular magnetic resonance*, 8(4):573–580, 2006.

[6] Corentin Provost, Marc Soudant, Laurence Legrand, Wagih Ben Hassen, Yu Xie, Sébastien Soize, Romain Bourcier, Joseph Benzakoun, Myriam Edjlali, Grégoire Boulouis, et al. Magnetic resonance imaging or computed tomography before treatment in acute ischemic stroke: effect on workflow and functional outcome. *Stroke*, 50(3):659–664, 2019.

[7] Ethem Alpaydin. *Machine learning*. MIT press, 2021.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[9] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, 2019.

[10] Chengjia Wang, Tom MacGillivray, Gillian Macnaught, Guang Yang, and David Newby. A two-stage 3d unet framework for multi-class segmentation on full resolution image. *arXiv preprint arXiv:1804.04341*, 2018.

[11] J Devlin, MW Chang, K Lee, and KB Toutanova. Pre-training of deep bidirectional transformers for language understanding in: Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers). *Minneapolis, MN: Association for Computational Linguistics*, pages 4171–86, 2019.

[12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[13] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International conference on machine learning*, pages 10347–10357. PMLR, 2021.

[14] Cheukkai Hui, Yu Xiang Zhou, and Ponnada Narayana. Fast algorithm for calculation of inhomogeneity gradient in magnetic resonance imaging data. *Journal of Magnetic Resonance Imaging*, 32(5):1197–1208, 2010.

[15] Jianzhong Wang, Jun Kong, Yinghua Lu, Miao Qi, and Baoxue Zhang. A modified fcm algorithm for mri brain image segmentation using both local and non-local spatial constraints. *Computerized medical imaging and graphics*, 32(8):685–698, 2008.

[16] S Ramathilagam, R Pandiyarajan, A Sathya, R Devi, and SR Kannan. Modified fuzzy c-means algorithm for segmentation of t1–t2-weighted brain mri. *Journal of Computational and Applied Mathematics*, 235(6):1578–1586, 2011.

[17] Ali Khazaee, Ata Ebrahimzadeh, and Abbas Babajani-Feremi. Identifying patients with alzheimer's disease using resting-state fmri and graph theory. *Clinical Neurophysiology*, 126(11):2132–2141, 2015.

[18] Joshua Kahan, Maren Urner, Rosalyn Moran, Guillaume Flandin, Andre Marreiros, Laura Mancini, Mark White, John Thornton, Tarek Yousry, Ludvic Zrinzo, et al. Resting state functional mri in parkinson's disease: the impact of deep brain stimulation on 'effective'connectivity. *Brain*, 137(4):1130–1144, 2014.

[19] Gabriela Pérez, Aura Conci, Ana Belén Moreno, and Juan Antonio Hernandez-Tamames. Rician noise attenuation in the wavelet packet transformed domain for brain mri. *Integrated Computer-Aided Engineering*, 21(2):163–175, 2014.

[20] Kostas Michalopoulos and Nikolaos Bourbakis. Combining eeg microstates with fmri structural features for modeling brain activity. *International Journal of Neural Systems*, 25(08):1550041, 2015.

[21] Nikos K Logothetis. The neural basis of the blood–oxygen–level–dependent functional magnetic resonance imaging signal. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1424):1003–1037, 2002.

[22] Oscar Esteban, Rastko Ciric, Karolina Finc, Ross W Blair, Christopher J Markiewicz, Craig A Moodie, James D Kent, Mathias Goncalves, Elizabeth DuPre, Daniel EP Gomez, et al. Analysis of task-based functional mri data preprocessed with fmriprep. *Nature protocols*, 15(7):2186–2202, 2020.

[23] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[24] Weiyi Yu, Zhizhong Huang, Junping Zhang, and Hongming Shan. San-net: Learning generalization to unseen sites for stroke lesion segmentation with self-adaptive normalization. *Computers in Biology and Medicine*, 156:106717, 2023.

[25] Yanglan Ou, Ye Yuan, Xiaolei Huang, Kelvin Wong, John Volpi, James Z Wang, and Stephen TC Wong. Lambdaunet: 2.5 d stroke lesion segmentation of diffusion-weighted mr images. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 731–741. Springer, 2021.

[26] Nabil Ibtehaz and M Sohel Rahman. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural networks*, 121:74–87, 2020.

[27] Mohsen Soltanpour, Russ Greiner, Pierre Boulanger, and Brian Buck. Improvement of automatic ischemic stroke lesion segmentation in ct perfusion maps using a learned deep neural network. *Computers in Biology and Medicine*, 137:104849, 2021.

[28] Azhar Tursynova and Batyrkhan Omarov. 3d u-net for brain stroke lesion segmentation on isles 2018 dataset. In *2021 16th International Conference on Electronics Computer and Computation (ICECCO)*, pages 1–4. IEEE, 2021.

[29] Bin Zhao, Zhiyang Liu, Guohua Liu, Chen Cao, Song Jin, Hong Wu, and Shuxue Ding. Deep learning-based acute ischemic stroke lesion segmentation method on multimodal mr images using a few fully labeled subjects. *Computational and Mathematical Methods in Medicine*, 2021, 2021.

[30] Batyrkhan Omarov, Azhar Tursynova, Octavian Postolache, Khaled Gamry, Aidar Batyrbekov, Sapargali Aldeshov, Zhanar Azhibekova, Marat Nurtas, Akbayan Aliyeva, and Kadrzhan Shiyapov. Modified unet model for brain stroke lesion segmentation on computed tomography images. *Computers, Materials & Continua*, 71(3), 2022.

[31] Shih-Yen Lin, Pi-Ling Chiang, Peng-Wen Chen, Li-Hsin Cheng, Meng-Hsiang Chen, Pei-Chun Chang, Wei-Che Lin, and Yong-Sheng Chen. Toward automated segmentation for acute ischemic stroke using non-contrast computed tomography. *International journal of computer assisted radiology and surgery*, 17(4):661–671, 2022.

[32] Khushboo Verma, Satwant Kumar, and David Paydarfar. Automatic segmentation and quantitative assessment of stroke lesions on mr images. *Diagnostics*, 12(9):2055, 2022.

[33] Alex Wong, Allison Chen, Yangchao Wu, Safa Cicek, Alexandre Tiard, Byung-Woo Hong, and Stefano Soatto. Small lesion segmentation in brain mris with

subpixel embedding. In *International MICCAI Brainlesion Workshop*, pages 75–87. Springer, 2021.

[34] Amish Kumar, Palash Ghosal, Soumya Snigdha Kundu, Amritendu Mukherjee, and Debashis Nandi. A lightweight asymmetric u-net framework for acute ischemic stroke lesion segmentation in ct and ctp images. *Computer Methods and Programs in Biomedicine*, 226:107157, 2022.

[35] Kelvin K Wong, Jonathon S Cummock, Guihua Li, Rahul Ghosh, Pingyi Xu, John J Volpi, and Stephen TC Wong. Automatic segmentation in acute ischemic stroke: Prognostic significance of topological stroke volumes on stroke outcome. *Stroke*, 53(9):2896–2905, 2022.

[36] Hae Sol Moon, Lindsay Heffron, Ali Mahzarnia, Barnabas Obeng-Gyasi, Matthew Holbrook, Cristian T Badea, Wuwei Feng, and Alexandra Badea. Automated multimodal segmentation of acute ischemic stroke lesions on clinical mr images. *Magnetic resonance imaging*, 92:45–57, 2022.

[37] Xiuquan Du, Kunpeng Ma, and Yuhui Song. Agmr-net: Attention-guided multiscale recovery framework for stroke segmentation. *Computerized Medical Imaging and Graphics*, 101:102120, 2022.

[38] Abdelmajid Bousselham, Omar Bouattane, Mohamed Youssfi, Abdelhadi Raihani, et al. Towards an accurate mri acute ischemic stroke lesion segmentation based on bioheat equation and u-net model. *International Journal of Biomedical Imaging*, 2022, 2022.

[39] Jiyeon Lee, Minho Lee, Jongseung Lee, Regina EY Kim, Seong Hoon Lim, and Donghyeon Kim. Fine-grained brain tissue segmentation for brain modeling of stroke patient. *Computers in Biology and Medicine*, 153:106472, 2023.

[40] Hani Alquhayz, Hafiz Zahid Tufail, and Basit Raza. The multi-level classification network (mcn) with modified residual u-net for ischemic stroke lesions segmentation from atlas. *Computers in Biology and Medicine*, 151:106332, 2022.

[41] Shaoquan Li, Jianye Zheng, and Dongjiao Li. Precise segmentation of non-enhanced computed tomography in patients with ischemic stroke based on multi-scale u-net deep network model. *Computer Methods and Programs in Biomedicine*, 208:106278, 2021.

[42] Lei Li, Kunpeng Ma, Yuhui Song, and Xiuquan Du. Tsrl-net: Target-aware supervision residual learning for stroke segmentation. *Computers in Biology and Medicine*, 159:106840, 2023.

[43] Valeriia Abramova, Albert Clerigues, Ana Quiles, Deysi Garcia Figueredo, Yolanda Silva, Salvador Pedraza, Arnau Oliver, and Xavier Lladó. Hemorrhagic stroke lesion segmentation using a 3d u-net with squeeze-and-excitation blocks. *Computerized Medical Imaging and Graphics*, 90:101908, 2021.

[44] R Karthik, R Menaka, M Hariharan, and Daehan Won. Ischemic lesion segmentation using ensemble of multi-scale region aligned cnn. *Computer Methods and Programs in Biomedicine*, 200:105831, 2021.

[45] Tianyu Shi, Huiyan Jiang, and Bin Zheng. C 2 ma-net: Cross-modal cross-attention network for acute ischemic stroke lesion segmentation based on ct perfusion scans. *IEEE Transactions on Biomedical Engineering*, 69(1):108–118, 2021.

[46] Haisheng Hui, Xueying Zhang, Zelin Wu, and Fenlian Li. Dual-path attention compensation u-net for stroke lesion segmentation. *Computational Intelligence and Neuroscience*, 2021, 2021.

[47] Boyu Huang, Guanru Tan, Haowen Dou, Zhihan Cui, Youyi Song, and Teng Zhou. Mutual gain adaptive network for segmenting brain stroke lesions. *Applied Soft Computing*, 129:109568, 2022.

[48] Hyunkwang Shin, Rockson Agyeman, Muhammad Rafiq, Min Cheol Chang, and Gyu Sang Choi. Automated segmentation of chronic stroke lesion using efficient u-net architecture. *Biocybernetics and Biomedical Engineering*, 42(1):285–294, 2022.

[49] Qiqi Bao, Shiyu Mi, Bowen Gang, Wenming Yang, Jie Chen, and Qingmin Liao. Mdan: mirror difference aware network for brain stroke lesion segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(4):1628–1639, 2021.

[50] Manjin Sheng, Wenjie Xu, Jane Yang, and Zhongjie Chen. Cross-attention and deep supervision unet for lesion segmentation of chronic stroke. *Frontiers in Neuroscience*, 16:836412, 2022.

[51] Pan Feng, Bo Ni, Xiantao Cai, and Yutao Xie. Utransnet: Transformer within u-net for stroke lesion segmentation. In *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 359–364. IEEE, 2022.

[52] Yeonghyeon Gu, Zhegao Piao, and Seong Joon Yoo. Sthardnet: Swin transformer with hardnet for mri segmentation. *Applied Sciences*, 12(1):468, 2022.

[53] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[54] Chun Luo, Jing Zhang, Xinglin Chen, Yinhao Tang, Xiechuan Weng, and Fan Xu. Ucatr: Based on cnn and transformer encoding and cross-attention decoding for lesion segmentation of acute ischemic stroke in non-contrast computed tomography images. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3565–3568. IEEE, 2021.

[55] Liangliang Liu, Ying Wang, Jing Chang, Pei Zhang, Gongbo Liang, and Hui Zhang. Llrhnet: multiple lesions segmentation using local-long range features. *Frontiers in Neuroinformatics*, 16:859973, 2022.

[56] Zelin Wu, Xueying Zhang, Fenglian Li, Suzhe Wang, Lixia Huang, and Jiaying Li. W-net: A boundary-enhanced segmentation network for stroke lesions. *Expert Systems with Applications*, page 120637, 2023.

[57] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[58] Zelin Wu, Xueying Zhang, Fenglian Li, Suzhe Wang, and Lixia Huang. Multi-scale long-range interactive and regional attention network for stroke lesion segmentation. *Computers and Electrical Engineering*, 103:108345, 2022.

[59] Moritz Platscher, Jonathan Zopes, and Christian Federau. Image translation for medical image generation: Ischemic stroke lesion segmentation. *Biomedical Signal Processing and Control*, 72:103283, 2022.

[60] H Van Voorst, PR Konduri, LM van Poppel, W van der Steen, PM van der Sluijs, EMH Slot, BJ Emmer, WH van Zwam, YBWEM Roos, CBLM Majoie, et al. Unsupervised deep learning for stroke lesion segmentation on follow-up ct based on generative adversarial networks. *American Journal of Neuroradiology*, 43(8):1107–1114, 2022.

[61] Shuqiang Wang, Zhuo Chen, Senrong You, Bingchuan Wang, Yanyan Shen, and Baiying Lei. Brain stroke lesion segmentation using consistent perception generative adversarial network. *Neural Computing and Applications*, 34(11):8657–8669, 2022.

[62] Rawan Ghnemat, Ashwaq Khalil, and Qasem Abu Al-Haija. Ischemic stroke lesion segmentation using mutation model and generative adversarial network. *Electronics*, 12(3):590, 2023.

[63] Suzhe Wang, Xueying Zhang, Haisheng Hui, Fenglian Li, and Zelin Wu. Multimodal ct image synthesis using unsupervised deep generative adversarial networks for stroke lesion segmentation. *Electronics*, 11(16):2612, 2022.

[64] Gaoxiang Chen, Jintao Ru, Yilin Zhou, Islem Rekik, Zhifang Pan, Xiaoming Liu, Yezhi Lin, Beichen Lu, and Jialin Shi. Mtans: multi-scale mean teacher combined adversarial network with shape-aware embedding for semi-supervised brain lesion segmentation. *NeuroImage*, 244:118568, 2021.

[65] Abang Mohd Arif Anaqi Abang Isa, Kuryati Kipli, Ahmad Tirmizi Jobli, Muhammad Hamdi Mahmood, Siti Kudnie Sahari, Aditya Tri Hernowo, and Sinin Hamdan. Pseudo-colour with k-means clustering algorithm for acute ischemic stroke lesion segmentation in brain mri. *Pertanika Journal of Science & Technology*, 29(2), 2021.

[66] Senthil Kumar Thiyagarajan and Kalpana Murugan. Arithmetic optimization-based k means algorithm for segmentation of ischemic stroke lesion. *Soft Computing*, pages 1–13, 2023.

[67] Jacob C Reinhold, Blake E Dewey, Aaron Carass, and Jerry L Prince. Evaluating the impact of intensity normalization on mr image synthesis. In *Medical Imaging 2019: Image Processing*, volume 10949, pages 890–898. SPIE, 2019.

[68] SGOPAL Patro and Kishore Kumar Sahu. Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*, 2015.

[69] Jason S Moser, Jonathan D Huppert, Edna B Foa, and Robert F Simons. Interpretation of ambiguous social scenarios in social phobia and depression: Evidence from event-related brain potentials. *Biological psychology*, 89(2):387–397, 2012.

[70] Cullen Schaffer. Selecting a classification method by cross-validation. *Machine learning*, 13:135–143, 1993.

[71] Andrew Vince. A framework for the greedy algorithm. *Discrete Applied Mathematics*, 121(1-3):247–260, 2002.

[72] Hyunjin Kwon, Jinhyeok Park, and Youngho Lee. Stacking ensemble technique for classifying breast cancer. *Healthcare informatics research*, 25(4):283–288, 2019.

[73] John D Kelleher. *Deep learning*. MIT press, 2019.

[74] Parvez Ahmad, Hai Jin, Roobaea Alroobaea, Saqib Qamar, Ran Zheng, Fady Alnajjar, and Fathia Aboudi. Mh unet: A multi-scale hierarchical based architecture for medical image segmentation. *IEEE Access*, 9:148384–148408, 2021.

[75] Chenxi Liu, Liang-Chieh Chen, Florian Schroff, Hartwig Adam, Wei Hua, Alan L Yuille, and Li Fei-Fei. Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 82–92, 2019.

[76] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[77] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16293–16303, 2022.

[78] Lina Pedraza, Carlos Vargas, Fabián Narváez, Oscar Durán, Emma Muñoz, and Eduardo Romero. An open access thyroid ultrasound image database. In *10th International symposium on medical information processing and analysis*, volume 9287, pages 188–193. SPIE, 2015.

[79] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural networks*, 107:3–11, 2018.

[80] David Opitz and Richard Maclin. Popular ensemble methods: An empirical study. *Journal of artificial intelligence research*, 11:169–198, 1999.

[81] MA Fernandez-Seara and FW Wehrli. Postprocessing technique to correct for background gradients in image-based r* 2 measurements. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 44(3):358–366, 2000.

[82] CP Behrenbruch, S Petroudi, S Bond, JD Declerck, FJ Leong, and JM Brady. Image filtering techniques for medical image post-processing: an overview. *The British journal of radiology*, 77(suppl_2):S126–S132, 2004.

[83] K Bhargavi and S Jyothi. A survey on threshold based segmentation technique in image processing. *International Journal of Innovative Research and Development*, 3(12):234–239, 2014.

[84] Diya Chudasama, Tanvi Patel, Shubham Joshi, and Ghanshyam I Prajapati. Image segmentation using morphological operations. *International Journal of Computer Applications*, 117(18), 2015.

[85] Mary L Comer and Edward J Delp III. Morphological operations for color image processing. *Journal of electronic imaging*, 8(3):279–289, 1999.

[86] Alle Meije Wink and Jos BTM Roerdink. Denoising functional mr images: a comparison of wavelet denoising and gaussian smoothing. *IEEE transactions on medical imaging*, 23(3):374–387, 2004.

[87] Fabian Isensee, Paul F Jäger, Peter M Full, Philipp Vollmuth, and Klaus H Maier-Hein. nnu-net for brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6*, pages 118–132. Springer, 2021.

[88] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30:449–459, 2017.

[89] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[90] Emmanuel L Barbier, Laurent Lamalle, and Michel Décorps. Methodology of brain perfusion imaging. *Journal of Magnetic Resonance Imaging: An official journal of the international society for magnetic resonance in medicine*, 13(4):496–520, 2001.

[91] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[92] Alex R Carter, Gordon L Shulman, and Maurizio Corbetta. Why use a connectivity-based approach to study stroke and recovery of function? *Neuroimage*, 62(4):2271–2280, 2012.

[93] Stephen M Smith, Christian F Beckmann, Jesper Andersson, Edward J Auerbach, Janine Bijsterbosch, Gwenaëlle Douaud, Eugene Duff, David A Feinberg, Ludovica Griffanti, Michael P Harms, et al. Resting-state fmri in the human connectome project. *Neuroimage*, 80:144–168, 2013.

[94] Ronald Sladky, Karl J Friston, Jasmin Tröstl, Ross Cunnington, Ewald Moser, and Christian Windischberger. Slice-timing effects and their correction in functional mri. *Neuroimage*, 58(2):588–594, 2011.

[95] Jesper LR Andersson, Chloe Hutton, John Ashburner, Robert Turner, and Karl Friston. Modeling geometric deformations in epi time series. *Neuroimage*, 13(5):903–919, 2001.

[96] Alan C Evans, D Louis Collins, SR Mills, Edward D Brown, Ryan L Kelly, and Terry M Peters. 3d statistical neuroanatomical models from 305 mri volumes. In *1993 IEEE conference record nuclear science symposium and medical imaging conference*, pages 1813–1817. IEEE, 1993.

[97] F Gregory Ashby. *Statistical analysis of fMRI data*. MIT press, 2019.

[98] Michael N Hallquist, Kai Hwang, and Beatriz Luna. The nuisance of nuisance regression: spectral misspecification in a common approach to resting-state fmri preprocessing reintroduces noise and obscures functional connectivity. *Neuroimage*, 82:208–225, 2013.

[99] Keith J Worsley, Sean Marrett, Peter Neelin, Alain C Vandal, Karl J Friston, and Alan C Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping*, 4(1):58–73, 1996.

[100] Peter A Bandettini and Eric C Wong. A hypercapnia-based normalization method for improved spatial localization of human brain activation with fmri. *NMR in Biomedicine*, 10(4-5):197–203, 1997.

[101] Paul M Matthews and Peter Jezzard. Functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(1):6–12, 2004.

[102] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[103] Hamed Asadi, Richard Dowling, Bernard Yan, and Peter Mitchell. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PloS one*, 9(2):e88225, 2014.

[104] Rémi Cuingnet, Charlotte Rosso, Marie Chupin, Stéphane Lehéricy, Didier Dormont, Habib Benali, Yves Samson, and Olivier Colliot. Spatial regularization of svm for the detection of diffusion alterations associated with stroke outcome. *Medical image analysis*, 15(5):729–737, 2011.

[105] MA Urbin, Xin Hong, Catherine E Lang, and Alex R Carter. Resting-state functional connectivity and its association with multiple domains of upper-extremity function in chronic stroke. *Neurorehabilitation and neural repair*, 28(8):761–769, 2014.

[106] Jau-Hong Lin, Miao-Ju Hsu, Ching-Fan Sheu, Tzung-Shian Wu, Ruey-Tay Lin, Chia-Hsin Chen, and Ching-Lin Hsieh. Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. *Physical therapy*, 89(8):840–850, 2009.

[107] Catherine E Lang, Joanne M Wagner, Alexander W Dromerick, and Dorothy F Edwards. Measurement of upper-extremity function early after stroke: properties of the action research arm test. *Archives of physical medicine and rehabilitation*, 87(12):1605–1610, 2006.

[108] Stephen J Page, SueAnn Sisto, Peter Levine, and Robert E McGrath. Efficacy of modified constraint-induced movement therapy in chronic stroke: a single-blinded randomized controlled trial. *Archives of physical medicine and rehabilitation*, 85(1):14–18, 2004.

[109] Anne K Rehme, Lukas J Volz, Delia-Lisa Feis, Simon B Eickhoff, Gereon R Fink, and Christian Grefkes. Individual prediction of chronic motor outcome in the acute post-stroke stage: Behavioral parameters versus functional imaging. *Human brain mapping*, 36(11):4553–4565, 2015.

[110] L Krumlinde-Sundholm, B Hoare, and B Lindkvist. Validity of the adult version of the assissting hand assessment (ad-aha), for persons with hemiplegia after stroke. In *9th COTEC Congress of Occupational Therapy*, 2012.

[111] Massimo Penta, Luigi Tesio, Carlyne Arnould, Arturo Zancan, and Jean-Louis Thonnard. The abilhand questionnaire as a measure of manual ability in chronic stroke patients: Rasch-based validation and relationship to upper limb impairment. *Stroke*, 32(7):1627–1634, 2001.

[112] Brendan Juba and Hai S Le. Precision-recall versus accuracy and the role of large data sets. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 4039–4048, 2019.

[113] Sergio A Alvarez. An exact analytical relation among recall, precision, and classification accuracy in information retrieval. *Boston College, Boston, Technical Report BCCS-02-01*, pages 1–22, 2002.

[114] Stephen J Page, Lynne V Gauthier, and Susan White. Size doesn't matter: cortical stroke lesion volume is not associated with upper extremity motor impairment and function in mild, chronic hemiparesis. *Archives of physical medicine and rehabilitation*, 94(5):817–821, 2013.

[115] Louise C Bannister, Sheila G Crewther, and Leeanne M Carey. Improvement in touch sensation after stroke is associated with resting functional connectivity changes. *Frontiers in neurology*, 6:146927, 2015.