ABSTRACT

Multiple Sclerosis (MS) lesion segmentation in brain MRI remains a challenging task for radiologists and neurologists due to the complex nature of MS lesions and the variability in their appearance. In this report, we present a state-of-the-art (SOTA) approach utilizing U-Net architectures with a deep feature extraction backbone to address this challenge.

The methodology capitalizes on the rich representation power of deep learning models to accurately segment MS lesions from brain MRI scans. We employ advanced U-Net variants that integrate deep feature extraction backbones, enabling the network to capture intricate details and spatial relationships crucial for lesion identification.

Evaluation of our approach is performed using various metrics including Dice score and Intersection over Union (IoU), which are widely accepted measures for assessing segmentation accuracy. The results demonstrate exceptional performance, surpassing previous methods in accurately delineating MS lesions.

Furthermore, we highlight the significance of our findings in clinical practice, emphasizing the critical need for precise and efficient lesion segmentation tools to aid radiologists and neurologists in diagnosing and monitoring MS progression. By providing reliable segmentation results, our method aims to alleviate the burden on clinicians and enhance patient care in the management of Multiple Sclerosis.

Keywords: Backbone, Dice Score, Feature Extraction, IoU, Lesions, MRI, Multiple Sclerosis, U-Net

CHAPTER 1: INTRODUCTION

The chapter discusses in brief about Multiple Sclerosis, it's causes, symptoms, modalities used in order for diagnosis and how clinical decision support systems can help doctors in assisted

1.1 INTRODUCTION

Multiple Sclerosis (MS) is a persistent autoimmune disorder that specifically impacts the Central Nervous System (CNS), mainly the brain and spinal cord. Multiple sclerosis (MS) is characterized by an autoimmune response in which the immune system erroneously targets the myelin sheath, a protective layer surrounding nerve fibers. This results in inflammation, scarring (sclerosis), and interference with the transmission of nerve signals. This leads to a diverse array of symptoms, such as exhaustion, debility, loss of sensation or tingling, impaired mobility, visual disturbances, and cognitive decline. Multiple sclerosis (MS) has an unpredictable nature and displays significant variations in both its severity and course across different individuals. Typically, it is diagnosed in young adults aged between 20 and 45 years, with a higher incidence observed in females [1]. Although a cure for MS is not yet available, therapies focus on symptom management, slowing the course of the illness, and enhancing the quality of life for sufferers. Typical symptoms of the condition include eye sores, reduced sensitivity, pain, weakness, loss of sensation in the arms and legs, fatigue, cognitive problems, and sadness [2].

1.2 CADe/CADx

Yoon and Kim in [3] describe the primary goals of using artificial intelligence (AI) to assist image interpretation as follows:

1) Computer-aided detection (CADe), also known as automated lesion detection, focuses on identifying worrisome anomalies in an image.

2) Computer-aided diagnosis (CADx), the process of describing anomalies found by either the radiologist or the computer.

The interpreting radiologist decides the clinical significance of the discovered aberration and whether it requires additional study based on the CADe/CADx analysis.

Kim et. al. in [4] report that early investigations noted that computer-aided detection (CAD), which functions as an automated second reader by flagging potentially problematic regions for radiologists to analyse, may increase the sensitivity of initial screening.

1.2.1 Decision Support System

Decision support systems (DSS) are invaluable tools in medical and clinical settings, leveraging advanced algorithms and data analytics to assist healthcare professionals

in various aspects of patient care. From diagnosis support to treatment planning, clinical decision making, disease management, and workflow optimization, DSS enhance clinical workflows, improve patient outcomes, and streamline healthcare delivery. By analysing patient data and providing evidence-based recommendations, DSS empower clinicians to make informed decisions, personalize treatment plans, and efficiently manage patient care. In today's healthcare landscape, DSS play a crucial role in enhancing the quality, efficiency, and effectiveness of healthcare delivery, ultimately benefiting both patients and healthcare providers.

1.3 CHOICE OF TECHNIQUE

U-Net architecture was primarily looked upon and was introduced by *Ronneberger et. al. in* [5] and was specifically designed for biomedical segmentation but as time passed and the complexity of modalities grew, the demand for better and deeper models grew with the desire to achieve higher accuracies. Moreover, the scarcity of data highlighted the problem of gradient diminishing through the end of the training process. The semantic gap between the encoder and decoder features, meaning that the decoder may struggle to reconstruct fine-grained details and produce accurate segmentation which is again a significant problem. Models introduced later in the time like that of U-Net++ introduced in 2018 in paper [6] by *Zhou et. al.* introduced advanced concepts like skip connections in order to resolve problems like dying gradient, semantic gap.

1.3.1 Feature Extraction Backbones

Backbones are foundational components in U-Net architectures, crucial for feature extraction and hierarchical representation learning. They employ convolutional layers to capture low-level and high-level features from input data, facilitating information fusion and adaptability. By leveraging pre-trained backbones, U-Net models can benefit from transfer learning, improving performance, and convergence speed, especially in scenarios with limited data. Overall, backbones play a pivotal role in enhancing the effectiveness and efficiency of U-Net models across diverse medical image analysis tasks.

1.4 COMPARISON OF MODELS

Dice Score and Intersection over Union (IoU) are fundamental metrics used in evaluating the performance of segmentation algorithms, particularly in tasks like identifying lesions in medical imaging such as brain MRI scans.

1.4.1 Dice Similarity Coefficient

The Dice Similarity Coefficient (DSC), commonly known as the Dice Score, is a key metric used in evaluating segmentation algorithms, especially in medical imaging. It measures the agreement between segmented regions produced by an algorithm and ground truth annotations. Calculated by comparing twice the intersection of predicted and ground truth regions to their sum, the Dice Score ranges from 0 to 1. A score of 1 signifies perfect overlap, while 0 indicates no overlap. Its simplicity and intuitive

interpretation make it widely favored. In medical imaging, the Dice Score aids in objectively assessing algorithm performance, guiding improvements, and ensuring accurate diagnoses. The DSC is formulated as:

$$DSC = \frac{2 \times |Prediction \cap Ground\ Truth|}{|Prediction| + |Ground\ Truth|} \qquad (1)$$

1.4.2 Intersection over Union

The Intersection over Union (IoU), also known as the Jaccard Index, is a crucial metric for evaluating segmentation algorithms, particularly in tasks like object detection and image segmentation. It quantifies the overlap between predicted and ground truth regions by measuring the ratio of their intersection area to their union area. With IoU values ranging from 0 to 1, a higher score indicates better agreement between the predicted and ground truth segmentations. IoU provides a balanced assessment of segmentation performance by considering both overlapping and non-overlapping areas. It is widely used in various fields, including medical imaging, to objectively evaluate algorithm accuracy and guide improvements in segmentation models.

$$IoU = \frac{|Prediction \cap Ground\ Truth|}{|Prediction \cup Ground\ Truth|} \qquad (2)$$

1.5 TOOLS AND REQUIREMENTS

1. **ANACONDA**: The Anaconda platform facilitates the development and deployment of deep learning models predicated on neural networks.

2. **Google Colab**: This platform afforded us a complimentary 15 GB Graphics Processing Unit (GPU) infrastructure for the training of our architectural models.

3. **Kaggle Notebooks**: The platform offers a wide range of GPUs which are available to use depending on the requirement and task like memory heavy codes can be run using T4x2 while for speed P100 and Tensor Processing Unit (TPUs) can be used. Nvidia A100 GPU was accessed as well.

4. Diverse Libraries such as Keras, TensorFlow, Matplotlib, Keras U-Net collection and Keras Tuner were employed in the study.

Report Outline

The document starts with the Introduction, followed by a chapter on Multiple Sclerosis then a brief Literature Review follows, Data is in Chapter 4, Chapters 5 and 6 are on Training and Results and Discussion, Conclusions are drawn in Chapter 6 with the Future Scope in the last chapter. The References, Appendix I and II are attached in the end.

CHAPTER 2: MULTIPLE SCLROSIS

The chapter discusses about Multiple Sclerosis (MS) as a disease, complications while diagnosis, various modalities involved in the diagnosis along with the role of CADx/CADe.

2.1 DISEASE

Multiple Sclerosis (MS) is a persistent autoimmune disorder that specifically impacts the central nervous system (CNS), including the brain, spinal cord, and optic nerves. Multiple sclerosis (MS) is characterized by an autoimmune response in which the immune system erroneously targets the myelin sheath, a protective layer around nerve fibers, resulting in inflammation, scarring (sclerosis), and disruptions in nerve signal transmission as shown in Figure 1. This results in a wide range of symptoms that can vary greatly among individuals, including fatigue, weakness, numbness or tingling, vision problems, difficulty walking, muscle stiffness or spasms, problems with coordination and balance, bladder or bowel dysfunction, and cognitive impairment. The course of MS is unpredictable and varies from person to person, with periods of relapse and remission or progressive worsening over time. While the exact cause of MS is unknown, a combination of genetic, environmental, and immune system factors is believed to play a role in its development. There is currently no cure for MS, but treatments are available to manage symptoms, slow disease progression, and improve quality of life for affected individuals.

As Multiple Sclerosis progresses, the cerebral cortex, which is the outer layer of the brain, experiences a decrease in size, a condition known as cortical atrophy. Multiple sclerosis refers to the distinct areas of scar tissue, sometimes called sclerosis, plaques, or lesions, that form as a result of the immune system's assault on myelin. These plaques are observable through magnetic resonance imaging (MRI), varying in size from as small as a pinhead to as large as a golf ball [7].
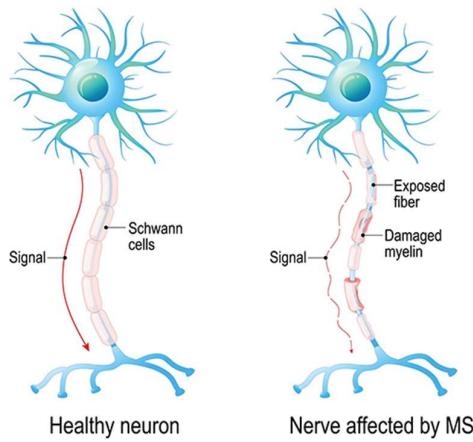

Figure 2.1: Neuron with Myelin Degraded

2.1.1 Symptoms

Patwardhan et. al. in [8] state clearly the approximate lifetime cost per patient suffering from MS is around about $2.5 million dollars in United States of America and hence MS diagnosis at an early stage is highly crucial given the condition cannot be reverted but the progression can be slowed down. According to Jhon Hopkins University in [9] the common initial symptoms can be quite confusing from being very basic ones like trouble walking, loss of vision because of optic neuritis (swelling of the optic nerve), hearing loss, bowel problems, depression etc. The symptoms are then grouped in advanced fashion namely: Primary, Secondary and Tertiary depending on the

destruction of myelin as primary symptoms show up as a direct result of destruction of myelin like paralysis whereas secondary symptoms occur as a result of primary symptoms like bedsore caused due to paralysis and the tertiary symptoms are related to social, job and psychological problems.

2.2 DETECTION

Detection of MS is highly difficult given its rarity and the mild lesions it causes in the random areas of brain and the symptoms are often very confusing, hence a targeted diagnosis is difficult to think of at first only.

2.2.1 Lumbar Puncture

Lumbar Puncture, also known as Spinal Tap or Cerebrospinal fluid study, involves the examination and diagnosis of the fluid present in the spinal column. This examination seeks to identify chemical and cellular irregularities that are associated with multiple sclerosis.

2.2.2 Electroencephalogram

Electroencephalogram (EEG) scans measure the brain's electrical response to auditory, visual, and sensory stimuli. These tests ascertain the presence of a deceleration in the transmission of information within certain areas of the brain.

2.2.3 Magnetic Resonance Imaging

Magnetic Imaging technique eliminates the need for X-rays by creating images of the body's organs and structures using a huge magnets. It can locate MS-related scars or plaques. Typically, an isolated incident combined with certain patterns of alterations in brain tissue observed during a contrast-enhanced MRI scan can indicate multiple sclerosis.

2.3 PROBLEMS

Multiple sclerosis (MS) detection is difficult for several reasons including:

1. The unpredictability of symptoms and the lack of a definitive diagnostic test are chief among the difficulties in detecting Multiple Sclerosis (MS).

2. The non-specific nature of MS symptoms often overlaps with those of other neurological disorders, leading to delayed or incorrect diagnoses.

3. The absence of any particular biomarkers for MS makes it challenging to provide a conclusive diagnosis.

4. MS can manifest in various forms, including progressive-relapsing, relapsing-remitting, primary progressive, and secondary progressive, each with distinct clinical presentations.

5. Establishing a definitive diagnosis of MS may require repeated imaging procedures and clinical examinations over time due to the diversity of its presentations.

6. Clinical assessment and imaging tests like MRIs may have limitations for diagnosis, including resource availability, interpretation discrepancies among medical professionals, and the potential for false-positive or false-negative results.

7. While MRI can reveal typical brain lesions associated with MS, these findings are not specific to the disease and can also be present in other disorders.

8. The lack of widespread acceptance or awareness of MS, particularly in its early stages when symptoms may be subtle, poses a significant obstacle to its identification.

9. This can lead to delays in seeking medical attention or reluctance to pursue further diagnostic evaluation, resulting in missed opportunities for early intervention and treatment.

2.4 BENEFITS OF CADe AND SEGMENTATION

Computer-aided detection and segmentation systems have the potential to greatly transform the early identification and diagnosis of Multiple Sclerosis (MS) by offering automated tools that can aid healthcare practitioners, even when MS is not their primary focus. The following is an explanation of how these systems might be beneficial:

1. Early Detection: Computer-aided methods may analyse medical imaging data, like as MRI scans, to identify small irregularities that suggest the presence of MS, even before symptoms become noticeable to the patient or healthcare practitioner.

By recognizing distinctive lesions associated with multiple sclerosis (MS), these systems can aid in the early identification and intervention, resulting in enhanced outcomes for patients.

2. Objective Analysis: These methods offer an impartial and consistent assessment of medical pictures, minimizing the unpredictability and subjectivity commonly linked to human interpretation. Computer-aided systems can provide useful information for clinical decision-making by measuring the features and distribution of lesions, which helps in understanding how a disease progresses and how it responds to therapy.

3. Alert Mechanisms: Computer-aided systems can include alert mechanisms that promptly inform healthcare practitioners when imaging results indicate a possible diagnosis of MS. These warnings operate as reminders for further assessment and can trigger rapid actions, such as extra diagnostic testing or referral to an expert, which help in early detection and intervention.

4. Efficiency and Workflow Optimization: Computer-aided systems can improve the efficiency of radiology operations and alleviate the workload of healthcare personnel by automating the identification and segmentation of problems linked to multiple sclerosis (MS). This enables radiologists and clinicians to dedicate their time and skills to more intricate cases or patient care tasks, hence enhancing the overall efficiency of their workflow.

5. Population screening: It can utilize computer-aided methods to identify persons who are at risk of acquiring MS or those who have early-stage illness and may

benefit from early intervention techniques. Through the analysis of extensive amounts of imaging data, these systems have the capability to detect individuals with modest or asymptomatic signs of illness, allowing for focused screening and intervention measures.

Summary

To summarize, the detection of Multiple Sclerosis (MS) poses challenges because of its nonspecific symptoms, the absence of conclusive testing, and the many ways in which the illness manifests. In order to tackle this issue, it is imperative to do continuous research to achieve more accurate diagnosis and enhance awareness. Computer-aided detection systems have the potential to greatly improve early diagnosis of multiple sclerosis by automating the detection of small anomalies in medical imaging. These technologies can enhance early intervention and improve patient outcomes by offering unbiased analysis and optimizing workflow. By utilizing several imaging modalities in conjunction with these technologies, the detection and accuracy of diagnosing multiple sclerosis can be significantly improved.

CHAPTER 3: LITERATURE REVIEW

The chapter deals with various benchmark results already in publication in various national and international reputed journals and conferences. The overall aim of the chapter is to review the established studies and provide an ablation study.

3.1 SEGMENATION MODELS

The current section of the paper discusses various segmentation techniques published in accordance with biomedical imaging techniques. In [10] *Ronneberger et. al.* gives the most basic U-Net architecture, and the name is derived from the "U" shape of the network comprising of encoder and decoder blocks as shown in Figure 2. The architecture described consists of a contracting path on the left side and an expansive path on the right side, resembling a convolutional network.
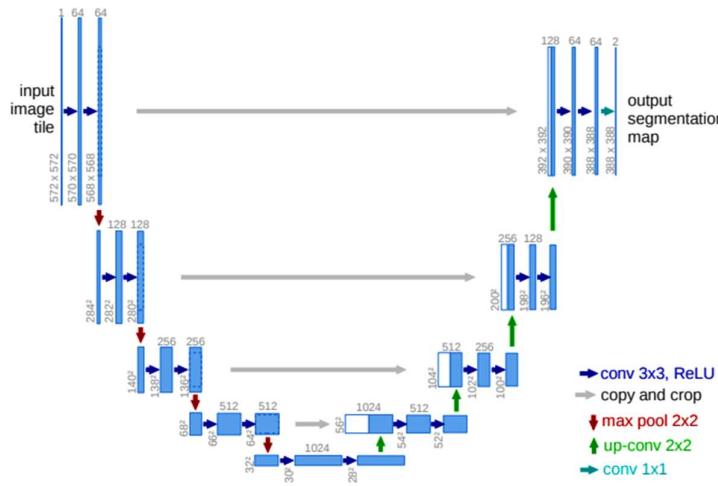
Figure 3.1: U-Net Architecture as described in [10]

The contraction route consists of a series of 3x3 convolutions, followed by ReLU activations and 2x2 max pooling for downsampling. Additionally, the feature

channels are doubled at each step. The expanding approach involves increasing the resolution of the input by upsampling, then reducing the number of feature channels by applying a 2x2 convolution. This is followed by concatenating the resulting feature map with the equivalent cropped feature map from the contracting path, then applying further convolutions. The last layer employs a 1x1 convolution to transform feature vectors into the intended classes. The network consists of 23 convolutional layers, which enables the output segmentation map to be tiled seamlessly. This is achieved by ensuring that all max-pooling operations are done to layers with even x- and y-sizes. After the introduction of U-Net in [10] there was extensive research in the domain of medical segmentation and the prominent problem of gradient decay along with other problems like improper upsampling of images due to poor fine grading and over smoothing was observed and was termed as semantic gap where fine details in segmentation masks could be lost due to excessive down-sampling and to cater to these problems *Zhou et. al.* in [11] introduced a U-Net++ architecture with nested architecture as shown in Figure 2.2, where each stage of the network progressively
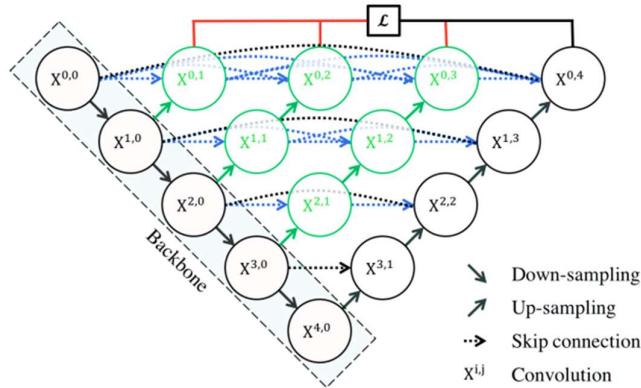
Figure 3.2: Architecture of UNet++ described by *Zhou el. al.*

refines feature representations. This hierarchical feature extraction enables better feature reuse and representation learning, leading to more robust segmentation performance. In UNet++ as described in [11] authors use skip connection as shown in Figure 3.3, these skip connections bridge encoder and decoder blocks, facilitating multi-scale information exchange. They preserve fine details and incorporate high-level semantics, enhancing feature representation. Additionally, they stabilize training by ensuring gradient flow. This architecture refinement enables UNet++ to produce accurate semantic segmentation, ideal for various image segmentation tasks.

$$x^{0,1} = H[x^{0,0}, U(x^{1,0})] \quad x^{0,2} = H[x^{0,0}, x^{0,1}, U(x^{1,1})] \quad x^{0,3} = H[x^{0,0}, x^{0,1}, x^{0,2}, U(x^{1,2})]$$

$$x^{0,4} = H[x^{0,0}, x^{0,1}, x^{0,2}, x^{0,3}, U(x^{1,3})]$$
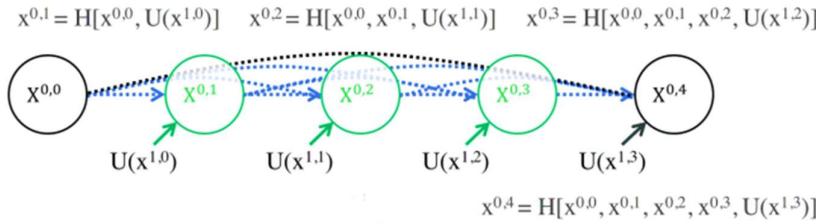
Figure 3.3: Comprehensive examination of the initial skip route of UNet++

Over the years more authors gradually worked on the same architecture and *Huang et. al.* in [12] introduced UNet 3+ a full scale connected UNet, the proposed full-scale skip connections enhance both the interconnection between the encoder and decoder and the intraconnection among decoder sub-networks. While UNet and UNet++ utilize different types of connections, they do not fully exploit information from all scales, resulting in a lack of explicit learning of organ position and boundaries. To address this limitation, each decoder layer in UNet 3+ integrates feature maps from both smaller and same-scale layers from the encoder, as well as larger-scale layers from the decoder. This approach captures fine-grained details and coarse-grained

semantics across all scales, improving overall performance. The full scale skip

connections can be visualised in Figure 3.4
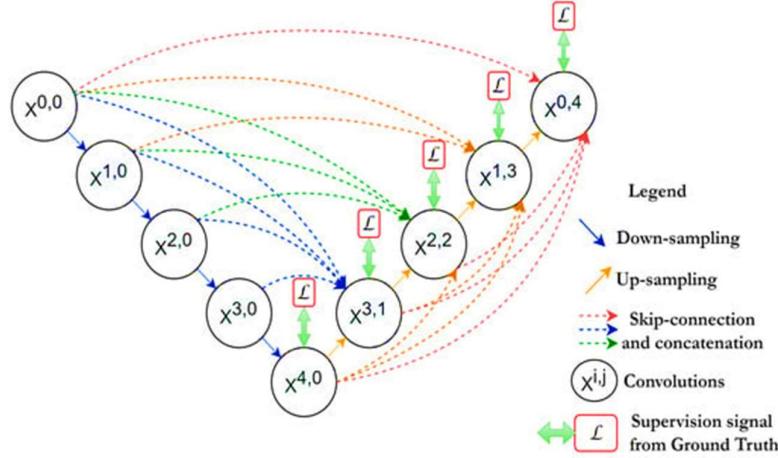
Figure 3.4: UNet3+ with full scale skip connections.

Diakogiannis et. al. in [13] introduced ResUNet-a a complete deep learning framework

which incorporates the capabilities of ResNet to avoid the dying gradient along with

semantic segmentation capabilities of UNet. The architecture of the ResUNet is shown

in Figure 3.5.
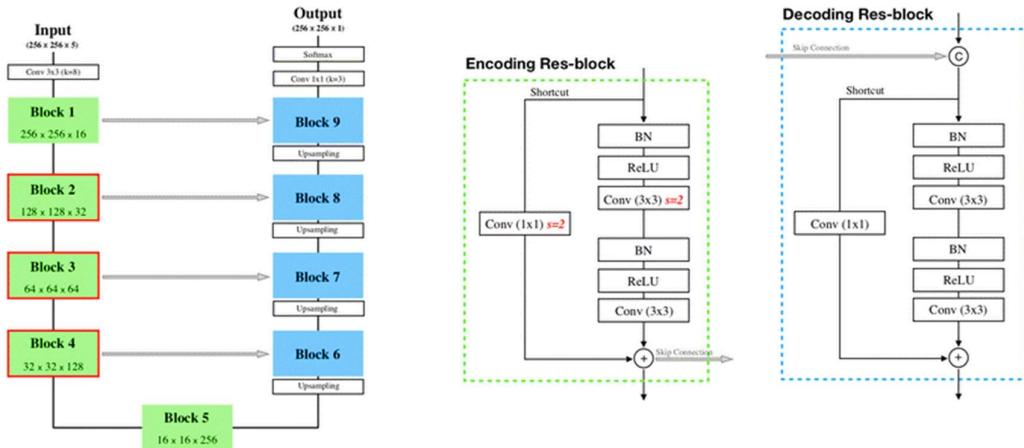
Figure 3.5: ResUnet-a architecture with Encoding and Decoding Residual blocks

Oktay et. al. in [14] introduced Attention U-Net as shown in Figure 3.6 where the

tradition U-Net architecture is incorporated with attention mechanism using an

17
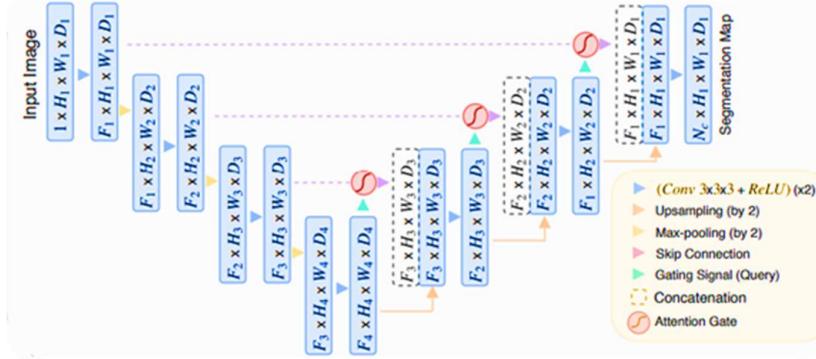
Figure 3.6: Attention U-Net with Attention Gate

Attention Gate (AG) as shown in Figure 3.7. The Attention UNet incorporates attention modules in both the encoder and decoder pathways. These modules adaptively modify the weights of features at various spatial locations according to their significance for the segmentation task. This allows the network to allocate greater resources to informative regions while diminishing the impact of distracting or less pertinent areas.
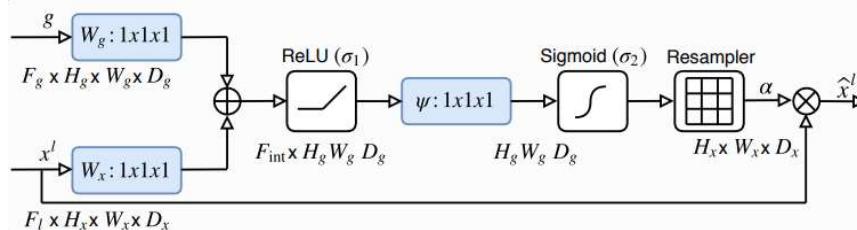
Figure 3.7: Attention Gate mechanism.

3.2 MACHINE LEARNING AND MULTIPLE SCLEROSIS

The author in [15] provides an ablation study of multiple research publications differentiated in four primary classes namely "Automated diagnosis of MS", "Prediction of MS disease progression", "Differentiation of MS stages" and "Differentiation of MS from similar disorders" the author lays ground for exploration in domain of immunology to address MS in generalized public screening which in

turn helps the masses of people in general. *Narayan et. al.* in [16] uses non contrast based T1-weighted, T-2 weighted, FLAIR MRI scans to prove that machine learning techniques like custom neural network with a 2D VGG 16 network cascaded over some of the result layers to classify the images as MS and Non MS with a satisfactory result of the convolutional neural network (CNN) achieving a sensitivity of 78% and a specificity of 73% in detecting enhancing lesions per slice. However, when considering predictions on a participant level, the results marginally decrease to 72% for sensitivity and 70% for specificity. The mean values of the areas under the receiver operating characteristic curve (AUC-ROC) were 0.82 for predictions made on a per-slice basis and 0.75 for predictions made on a per-participant basis. The results highlight the CNN's efficacy in detecting enhancing lesions, especially when examining individual slices, as seen by superior sensitivity, specificity, and AUC-ROC values compared to assessments made on a participant level. *Seccia et. al.* in [17] published results and concluded that there is no usable model to detect MS as such however given the increase in knowledge base the future seems promising. The author tries to establish prognostic results while working up on established clinical data. The author in [17] also works up on various methodologies to figure out the probable path for selecting suitable machine learning methodologies using multiple validation approaches. *Aslam et. al.* in [18] review multiple papers and conclude that most of the authors i.e., 51 researchers out of 90 prefer MRI data for their studies. *Zhao et. al.* in [19] use Expanded Disability Status Scores (EDSS) to determine patient status after a 5-year baseline visit and used Support Vector Machine (SVM) classifier and Logistic

Regression (LR) on demographic, clinical and MRI data. The author concluded that in most situations, the performance of LR was consistently worse than that of SVM. Variables such as race, family history of multiple sclerosis (MS), and brain parenchymal fraction were shown to be important predictors for the group that did not experience worsening of the disease. On the other hand, brain T2 lesion volume was identified as a powerful predictor for the group that did experience worsening of the disease. *Pinto el. al.* in [20] first classifies data into three classes based on severity of disease and further after using pre-processing techniques, normalization, etc. training classifiers like Decision Trees, K-Nearest Neighbours etc. The author in [20] reports that the 2-year validation model are the best for SP Development/Not Development, Disease Severity in 6th year and for the Disease Severity in 10th year the 5- year model performed the best where SP is the Secondary Progressive which the course of the ailment and the onset is the called as the Primary Progressive (PP).

Summary

The chapter reviews in detail the various research publications established for both segmentation models and role of computers in MS diagnosis. However, it was noteworthy that no significant work was found for automated segmentation lesions.

CHAPTER 4: DATA

The chapter in detail talks about data sourcing, preprocessing techniques etc. The chapter tries to establish lay all the pre-requisites before the entire training process.

4.1 DATA SOURCING

The MRI data being seriously private and expensive is hard to come by in public domain and for such a rare disease is even harder to come by. *Muslim et. al.* in [21] provide an open-sourced data collected from 20 different centres, the patient's MRI scans were obtained using a 1.5 Tesla MRI, each using different MRI sequence settings. All the images were stored in NIfTI image format in .nii format which was created by Neuroimaging Informatics Technology Initiative and is commonly used to store MRI data. The images were captured using all T1-weighted, T2-weighted, and Fluid-Attenuated Inversion Recovery (FLAIR) where T1-weighted, T2-weighted, and Fluid-Attenuated Inversion Recovery (FLAIR) are different types of magnetic resonance imaging (MRI) sequences used to visualize various tissues and pathologies within the body. T1-weighted MRI emphasizes differences in tissue relaxation times, with shorter times appearing bright and longer times appearing darker. T2-weighted MRI highlights contrasts in tissue relaxation times, with longer times appearing bright and shorter times appearing darker. FLAIR MRI is a specialized sequence that suppresses fluid signals while enhancing tissue contrasts, particularly useful for detecting brain lesions and abnormalities near fluid-filled spaces. The data comprised

3 of these images for all 60 patients overall in 3D image format. Nibabel library was used to open .nii files. The 3D view of MRI can be seen in Figure 4.1.

(a)   (b)

(c)

(d)
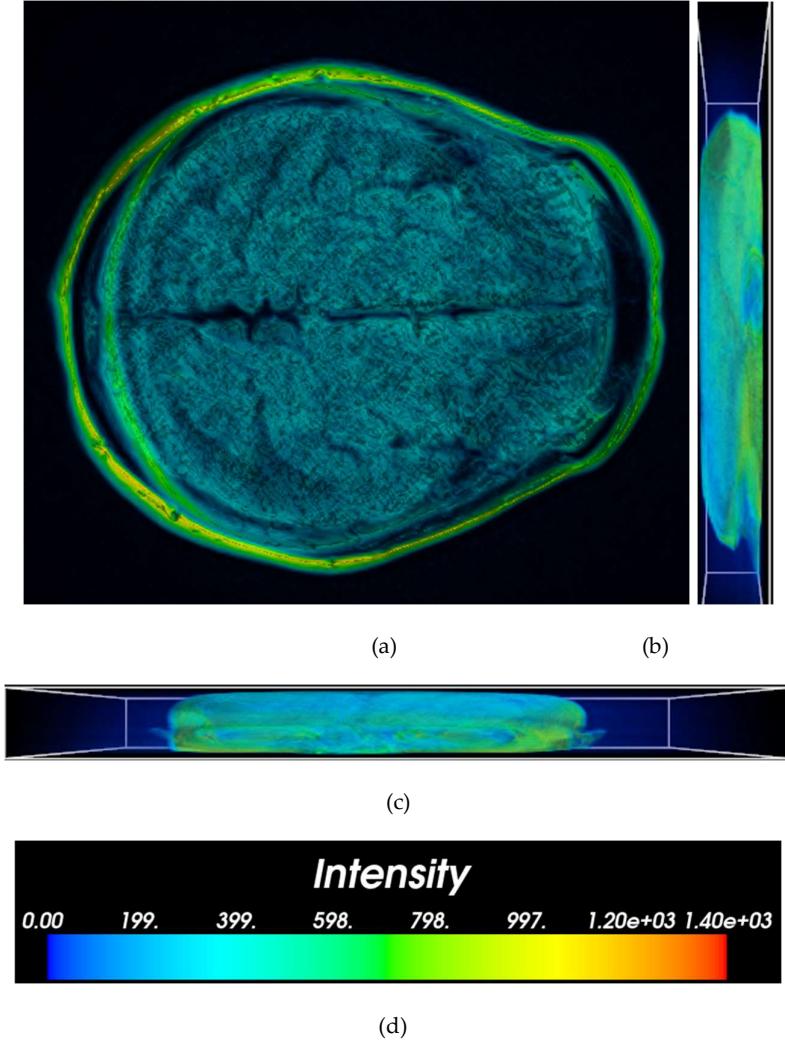
Figure 4.1: 3D view of MRI using Mayavi

In Figure 4.1 (a) represents Z axis or top view, (b) is Y axis or side view, (c) is X axis or front view, (d) is volumetric intensity legend and from the image it can be clearly seen that the X and Y axis do not have any significant information and instead hold the density due to number of slices. The dataset in [21] also provide a folder of consecutive masks for each of the masks and that too was presented in .nii format and was

visualized similarly using nibalbel and mayavi libraries. The Z axis information for the masks corresponding to the MRI scans can be seen in Figure 4.2
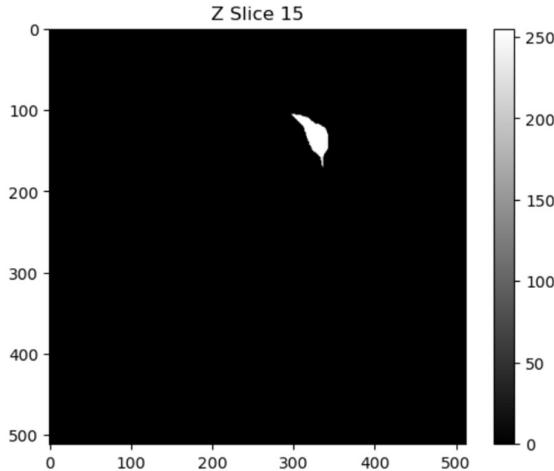
Figure 4.2: Z axis slice of a mask

4.1.1 Slice Extraction

To work on the MRI data the easiest way to work is extraction of slices from the 3D volumetric data which can be determined from the .nii header or the image shape in general as in seen in Figure 4.3.

Figure 4.3: Shape of Image and Mask

The format of the image shape in Figure 4.3 can be used to conclude that the image and mask are both of shape (512, 512) while both have 19 slices each which shows that image and mask are very well coherent.

Due to huge size of the MRI slices it was crucial to keep only those slides which have some pixels in the corresponding masks because the other information is all redundant. The same can be seen in Figure 4.4 where both a slice with corresponding mask completely empty and an MRI with mask which has some significant pixels can be seen with their corresponding slice number at the bottom.
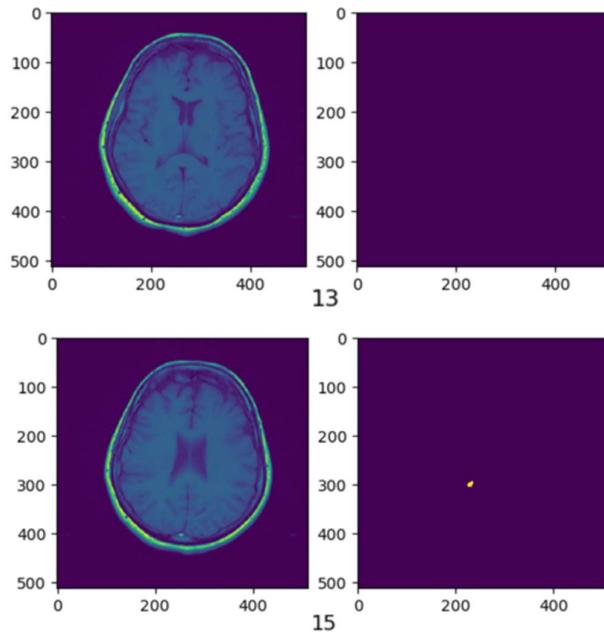
Figure 4.4: MRI scans with corresponding masks

Now to remove this redundant information it was important to remove the slices with no information is given in Figure 4.5 while to store the information of images and masks in form of NumPy arrays for easier input to the models as the models train effectively using np arrays format and the format is convertible to image format in .png, .jpg or whatever format suits the best and the code for the same can be seen in Figure 4.6. the Figure 4.6 also illustrates that the images saved as NumPy arrays have an extension of

.npy as a more suitable option. The conversion of images to NumPy arrays also makes it

easy to normalize and augment.

```python
import nibabel as nib
import matplotlib.pyplot as plt
import numpy as np

# Load the NIfTI files for image and mask
image_nifti = nib.load(r'C:\Users\DELL\Downloads\1-T1.nii\1-T1.nii')
mask_nifti = nib.load(r'C:\Users\DELL\Downloads\1-LesionSeg-T1.nii\1-LesionSeg-T1.nii')

# Get the image and mask data arrays
image_data = image_nifti.get_fdata()
mask_data = mask_nifti.get_fdata()

# Iterate through the slices
for slice_idx in range(image_data.shape[2]):
    # Extract the 2D slice and corresponding mask slice
    image_slice = image_data[:, :, slice_idx]
    mask_slice = mask_data[:, :, slice_idx]

    # Check if the mask slice has any non-zero values
    if np.any(mask_slice):
        # Plot and save the slice
        plt.imshow(image_slice, cmap='gray')
        plt.axis('off')  # Turn off axis
        plt.savefig(f'slice_{slice_idx}.png', bbox_inches='tight', pad_inches=0)
        plt.close()
        plt.imshow(mask_slice, cmap='gray')
        plt.axis('off')  # Turn off axis
        plt.savefig(f'mask_slice_{slice_idx}.png', bbox_inches='tight', pad_inches=0)
        plt.close()
```

Figure 4.5: Selecting images whose mask have some information

```python
import cv2
import numpy as np
import os

# Define the path to the folder containing JPG images
jpg_folder = r'C:\Users\DELL\Downloads\Brain MRI Dataset of Multiple Sclerosis with Consensus Manual Lesion Segmentation

# Define the path to the folder where you want to save the NumPy arrays
numpy_array_folder = r'C:\Users\DELL\Downloads\Brain MRI Dataset of Multiple Sclerosis NP\Masks'

# Create the folder if it doesn't exist
os.makedirs(numpy_array_folder, exist_ok=True)

# Iterate over each JPG file in the folder
for filename in os.listdir(jpg_folder):
    if filename.endswith('.jpg'):
        # Read the JPG image using OpenCV
        image_path = os.path.join(jpg_folder, filename)
        image = cv2.imread(image_path)

        # Convert the image to a NumPy array
        image_array = np.array(image)

        # Save the NumPy array
        npy_filename = os.path.splitext(filename)[0] + '.npy'  # Change extension to .npy
        npy_path = os.path.join(numpy_array_folder, npy_filename)
        np.save(npy_path, image_array)

        print(f"Saved {npy_path}")

print("Conversion and saving completed.")
```

Figure 4.6: Conversion of images to NumPy arrays in .npy extension

4.1.2 Data Splitting

The data was split into 90:10 ratio for training and testing initially before any methods were applied to data in order to receive better performance analysis. Data splitting also mitigates the problem of over fitting. Overfitting is the result of a model learning to closely match the training data, including noise and irrelevant patterns that do not apply well to fresh data. Through the process of assessing the model's performance on an independent testing dataset, we are able to identify and address the issue of overfitting. If the model exhibits high performance on the training set but low performance on the testing set, it is probable that the model is overfitting. When developing a model, it is typical to fine-tune hyperparameters such as the learning rate and regularization strength in order to maximize performance. The testing set serves to validate these hyperparameter selections by offering an autonomous dataset for examination. This mitigates the potential bias in hyperparameter setting towards the training data. Also, when comparing various models as in this case, the train-test split allows for equitable comparisons. Every model may undergo training and evaluation using identical training and testing sets, enabling unbiased comparisons of their performance.

4.2 PREPROCESSING

Data preprocessing is an essential stage in machine learning and data analysis. It comprises the cleaning, converting, and arranging of data prior to analysis or

modeling. Due to reservations in the medical fraternity data preprocessing is looked down upon by doctors and clinicians. Reasons for the same are:-

1. **Information Loss:** Medical professionals stress the precision and comprehensiveness of medical data. Preprocessing procedures, such as data imputation or outlier removal, may include the elimination or alteration of data points, which might result in the loss of crucial information. Medical experts may have concerns regarding the potential consequences of data loss on the dependability and authenticity of the dataset.

2. **Biases and assumptions**: Which can be introduced into the data through preprocessing techniques such as normalization or standardization. Physicians depend on impartial and objective data analysis to make well-informed clinical judgments. The users can be reluctant to utilize preprocessed data if it includes unintentional biases or assumptions that could compromise the accuracy or validity of their evaluations.

3. **Interpretability**: Medical practitioners place a high importance on the interpretability and openness of data analysis outcomes when considering complexity. Advanced preprocessing procedures or modifications might enhance the intricacy of the data, posing difficulties for doctors in interpreting and comprehending the analytic results. Physicians may favor less complex and more transparent preparation methods that preserve the interpretability of the data.

4. **Ethical and legal consideration:** Doctors comply with ethical standards and regulatory rules that control the privacy and confidentiality of patient data. To safeguard patient privacy, it may be essential to employ certain preprocessing methods, such as de-identification or anonymization. Nevertheless, physicians may refuse to use preprocessing techniques that jeopardize patient confidentiality or infringe upon ethical principles.

5. **Insufficient Clinical Validation**: The medical community may not always approve or endorse preprocessing approaches. Physicians depend on evidence-based procedures and established clinical guidelines to guide their decision-making. They could refuse to accept preprocessing procedures that do not have validation or agreement among medical experts, instead favoring ways supported by clinical data and research that has been evaluated by peers.

Due to similar reasons data augmentation techniques which are frequently employed strategy in machine learning that includes artificially expanding the size of a dataset by performing different transformations to existing data samples, such as rotation, translation, scaling, and flipping are also not supported by doctors as it might give arise to data and images depicting diseases which do not exist at all, as the use of augmented data has the potential to introduce mistakes or distortions that may have an impact on patient safety. Physicians value the well-being and safety of their patients above all other considerations, and they may hesitate to utilize enhanced data if it presents any potential harm to the health or overall welfare of the patient.

Given the reasons not to implement data preprocessing and data augmentation it is crucial to at least resize the images to a standard size so as to maintain a uniformity for the same reason the image size was set to '256 X 256' so that model can be tuned to an input size of 256 X 256.

Summary

The chapter talks in detail about data sourcing and various techniques applied to convert 3D .nii NIfTI files to 2D images and subsequently to NumPy arrays for easier storage and usage. The later sections discuss the role of preprocessing and augmentation techniques for bio-medical image data.

CHAPTER 5: TRAINING

The chapter revolves around various models adopted and various backbones in use for the segmentation models to train and for feature extraction.

5.1 BACKBONE

ResNet, VGG, and other similar backbones have a vital function in UNet segmentation models since they operate as feature extractors. The pre-trained convolutional neural networks (CNNs) have undergone training on extensive image classification tasks and have acquired the ability to efficiently extract high-level features from pictures. When included into UNet structures, these foundational components offer a robust framework for extracting features at various sizes and degrees of abstraction.

The main function of backbones in UNet segmentation models is to extract significant characteristics from input pictures. ResNet, VGG, and analogous architectures have acquired the ability to identify diverse patterns and structures in pictures by employing many layers of convolutional processes. By utilizing pre-trained backbones, UNet segmentation models may exploit these acquired characteristics, augmenting their capacity to gather pertinent information for segmentation tasks.

Backbones like as ResNet and VGG are specifically built to have a hierarchical structure of feature representations. In this structure, the lower levels are responsible for capturing low-level characteristics such as edges and textures, while the upper layers collect more abstract and sophisticated features such as forms and objects.

When included into UNet topologies, these hierarchical representations allow the model to capture both intricate features and overarching semantic information, which are crucial for precise segmentation.

For evaluation of various backbones, a standard Unet model was chosen using sm library by Qubvel [23] the quantive analysis show that VGG [23] proved out to be the best backbone in terms of overall Unet models however in the standard Unet model MobileNet [24] proved to be the best. Figure 5.19 (a) shows the VGG blocks for feature extraction while Figure 5.1 (b) shows the MobileNet architecture.
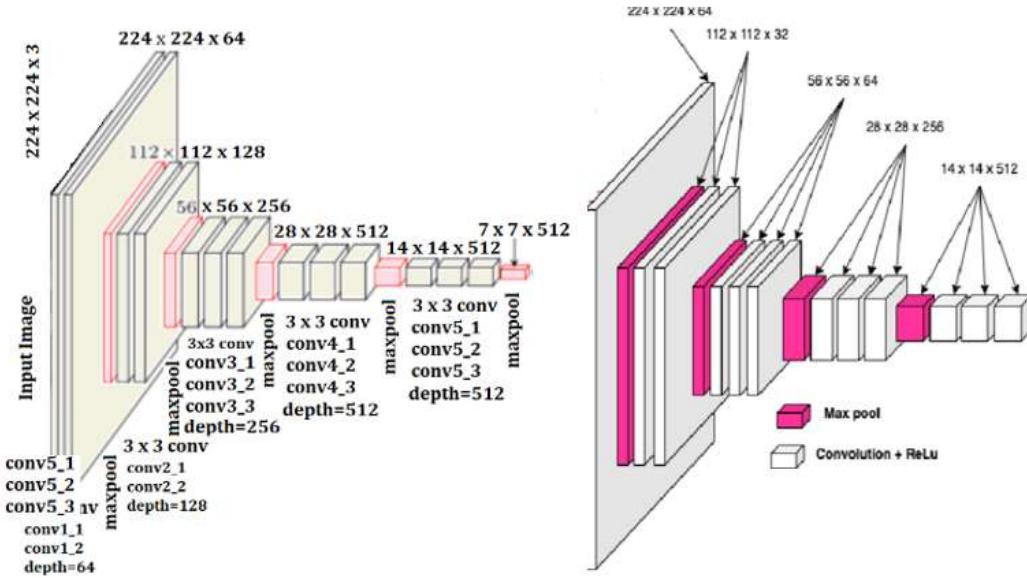
Figure 5.1: (a) VGG Feature Extractor, (b) MobileNet Feature Extractor

5.1.1 VGG Over MobileNet

There are various reasons why VGG is commonly chosen over MobileNet as the backbone for UNet in semantic segmentation tasks. The reasons for the same is given as follows in the points given below:

1. **Feature Extraction Capacity:** VGG has a greater feature extraction capacity than MobileNet due to its deeper design, enabling it to collect more intricate and complicated information from input photos. Precisely identifying high-level features is essential in semantic segmentation to precisely delineate object boundaries and capture semantic information. The deeper layers of VGG let it to capture more abstract representations, resulting in enhanced segmentation performance.

2. **Availability of pre-trained weights:** Pre-trained VGG models are readily accessible and often employed in diverse computer vision applications, such as picture categorization. The pre-trained weights may be easily transferred to the UNet backbone, which simplifies transfer learning and enhances segmentation performance, particularly in cases when there is a scarcity of labeled training data. Although MobileNet also offers pre-trained models, the widespread availability and broad usage of VGG make it a more favored option in several situations.

3. **Trade-off Between Speed and Accuracy:** MobileNet is specifically engineered to have a low computational burden and be highly efficient, making it ideal for mobile and embedded applications that prioritize speed and have limited resources. However, in semantic segmentation jobs where precision is of utmost importance and there is an abundance of processing resources, the focus switches towards reaching a greater level of accuracy in segmentation rather than prioritizing speed optimization. The deeper design of VGG places

a higher emphasis on the quality of features rather than computational performance, making it more suitable for semantic segmentation tasks in this particular situation.

4. **Dataset:** The decision to use either VGG or MobileNet as a UNet backbone may also be influenced by the level of difficulty of the segmentation job and the specific attributes of the dataset. If the job requires precise delineation of object borders and various semantic information, or if the dataset consists of high-resolution photos with many features, VGG's more complicated architecture may be more adept at capturing these intricacies efficiently as in our case the lesions are small and complex in varied regions of brain hence, VGG stands a better chance at adapting to the problem.

5.2 MISCELLANEOUS

5.2.1 Number of Epochs

Due to complexity of the task at hand the number of epochs were also checked using a standard P100 GPU at various levels starting from 25 till 200 at increments of 25. However, it was noteworthy that the time consumed for one the model with 200 epochs was approximately more than 4 hours and the progression was slow and was not cost effective after 75 epochs, also the GPU quota of the Kaggle Notebook was curtailed at 30 hours hence more usage was restricted.

5.2.2 Learning Rate

Hyperparameter tuning especially the learning rate is crucial to learning the sophisticated features of the images and lesion region. If the learning rate is very high, the optimization process may surpass the minimum of the loss function and be unable to reach convergence. This results in training behaviour that is unstable, characterized by oscillations or divergence of the loss function. High learning rates might lead to the model bypassing the optimal solution or converging to an inferior one. Consequently, the model exhibits inadequate ability to accurately predict outcomes on new and unfamiliar data. Nevertheless, higher learning rates can accelerate the training process by making bigger strides towards the minimum of the loss function. Occasionally, they can expedite the model's departure from local minima and saddle points.

Whereas low learning rates result in a decelerated convergence process during training. The optimization technique employs smaller increments towards the minimum of the loss function, leading to an extended duration for training. Although slower, smaller learning rates frequently lead to more consistent training behaviour. The model converges smoothly towards the ideal solution without any oscillations or divergence. Lower learning rates have a reduced probability of failing to find the best answer and demonstrate improved generalization to new, unknown data. They offer more streamlined optimization paths and more precise modifications to model parameters. Hence, it is highly crucial to find the optimal learning rate as in

my case starting with 0.1, we scheduled to 0.01, 0.001, 0.0001 but the best cost to

performance ratio was at almost 0.001. The Figure 5.2 shows the role of learning rate

in process of model training.
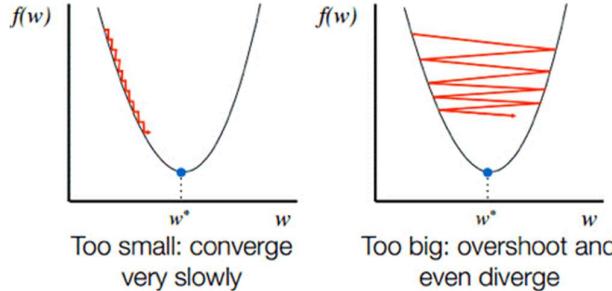
Figure 5.2: Role of Learning Rate

In case of contour plots we can see Figure 5.3 that depicts the contour 3D plots for a

function where (a) depicts small learning rate reaching global minimum very slowly

and (b) depicts large learning rate getting stuck in local minimums.
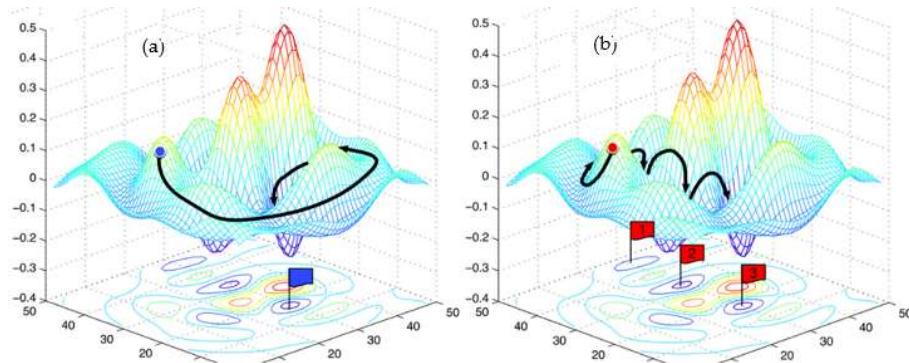
Figure 5.3: Learning Rate in 3D contours

5.2.3 Filters and Activation

Filters are selected starting from 64 multiplied by the powers of two as shown in

Figure 5.4 with an input size of 256 X 256 X 3 which is a standard for UNet collection

library by Keras by *Yingkaisha* [26] the activation is set to ReLu activation formulated as shown in equations (5) and is plotted in Figure 5.5.

$$f(x) = max(0, x) \tag{5}$$
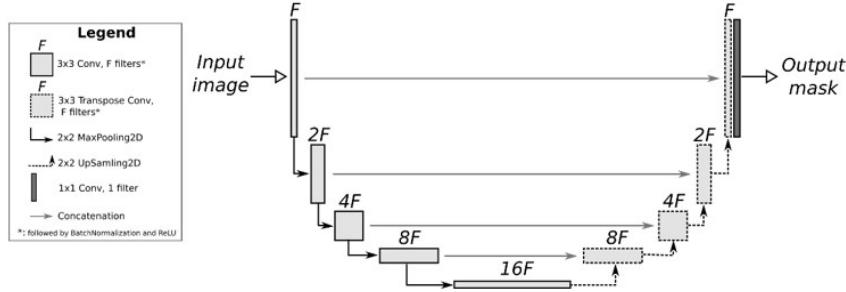
Figure 5.4: UNet Architecture (filters) used

Output activation was set to sigmoidal activation formulated in (6). The difference between ReLU (Rectified Linear Unit) and Sigmoidal Function is depicted in Figure 5.5. ReLU effectively mitigates the issue of vanishing gradient by maintaining a consistent gradient of 1 for all positive input values. The retrograde flow of gradients during backpropagation is facilitated, leading to enhanced training efficacy.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{6}$$
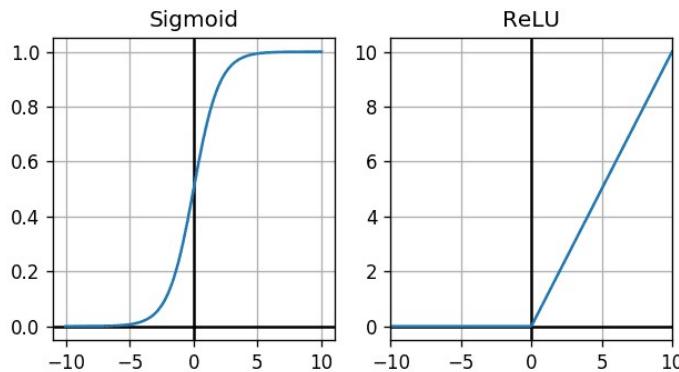
Figure 5.5: ReLU vs Sigmoidal function

5.2.4 Miscellaneous

This section focuses on the setup of a UNet model for training, specifically highlighting important elements like the loss function, optimizer, and evaluation measures.

Loss Function

The selected loss function for this UNet model is binary cross-entropy. The binary cross-entropy loss function is commonly employed for binary classification tasks and is particularly suitable for situations where the model predicts binary outputs, such as segmentation tasks.

Optimization algorithm

To improve the process, the Adam optimizer is utilized with a learning rate of 1e-3. Adam is a very efficient and effective optimization technique for training neural networks that is specifically designed to adaptively adjust the learning rate. Adam optimizes training by dynamically modifying the learning rate according to the gradients, resulting in accelerated convergence and improved performance.

Evaluation Metrics

During the training process, the model's performance is assessed using two metrics:

1. Accuracy which is a commonly used statistic in classification tasks. It quantifies the proportion of samples that are successfully categorized. However, it is not a trustworthy metric given that the segmentation masks are mostly black and hence the accuracy is not a standard because the imbalance is significantly high.

2. The Dice Coefficient is a widely used statistic in segmentation tasks that measures the degree of overlap between expected and ground truth segmentation masks. The Dice coefficient offers a measure of the model's capacity to precisely define the borders of objects and is formulated in (1).

To achieve successful segmentation of target objects in the input photos, I trained models by configuring the UNet model with binary cross-entropy loss, Adam optimizer with a learning rate of 1e-3 and evaluating its performance using accuracy and the Dice coefficient on the validation set

Summary

The chapter entails a list of parameters, backbone feature extraction tools which were chosen after quantitative analysis on a standard model as to run various moderner variations and versions based on standard UNet model.

CHAPTER 6: RESULTS AND DISCUSSION

The chapter in detail talks about the performance of various State Of The Art (SOTA) UNet architecture most of them are described in Chapter 3 of the report and with the parameters discussed in Chapter 5.

6.1 PERFORMANCE METRICS

The following tables discuss various performance metrics and how these affect the final score. Table 6.1 gives number of parameters for all 6 models.

Table 6.1: Number of Parameters

S No.	Model name	Trainable	Non-Trainable	Total
1.	UNet with VGG Backbone and ImageNet weights	14,722,368 (56.16 MB)	16,450,049 (62.75 MB)	31,172,417 (118.91 MB)
2.	UNet+ 2D with VGG Backbone and ImageNet weights	13,316,609 (50.80 MB)	14,722,368 (56.16 MB)	28,038,977 (106.96 MB)
3.	Attention UNet with VGG Backbone and ImageNet weights	13,664,293 (52.13 MB)	14,720,448 (56.15 MB)	28,384,741 (108.28 MB)
4.	UNet with VGG Backbone trained from scratch	34,520,193 (131.68 MB)	13,696 (53.50 KB)	34,533,889 (131.74 MB)
5.	R2 UNet with VGG Backbone trained from scratch	101,942,017 (388.88 MB)	37,248 (145.50 KB)	101,979,265 (389.02 MB)
6.	Attention UNet with VGG Backbone trained from scratch	31908517 (121.72 MB)	11,776 (46.00 KB)	31920293 (121.77 MB)

Table 6.2 defines Dice Coefficient for both training and validation set.

Table 6.2: Dice Scores

S. No.	Model name	Training	Validation
1.	UNet with VGG Backbone and ImageNet weights	0.7853	0.4626
2.	UNet+ 2D with VGG Backbone and ImageNet weights	0.7178	0.1482
3.	Attention UNet with VGG Backbone and ImageNet weights	0.7733	0.3373
4.	UNet with VGG Backbone trained from scratch	0.6384	0.4054
5.	R2 UNet with VGG Backbone trained from scratch	0.6930	0.43367
6.	Attention UNet with VGG Backbone trained from scratch	0.5955	0.3922

From the table above it is clearly evident that the validation dice score lesser than training scores, but the Figures 6.1 to 6.6 depict that the model tries to converge as we move further into the training process which was stopped at 50 epochs for the most models due to limited compute power but models which were trained from scratch perform much better that model using pretrained weights. Also, UNet+ 2D performed severely poorly and hence it was replaced by R2 UNet [26] in second training phase. Figure 6.1 to Figure 6.6 have plots for the Dice Score for all of the 6 models that were

used overall for the ablation study of all the models. The Appendix I of the report carries images of Accuracy vs Loss plots and associated tables just to prove that the models were in fact not overfitting.
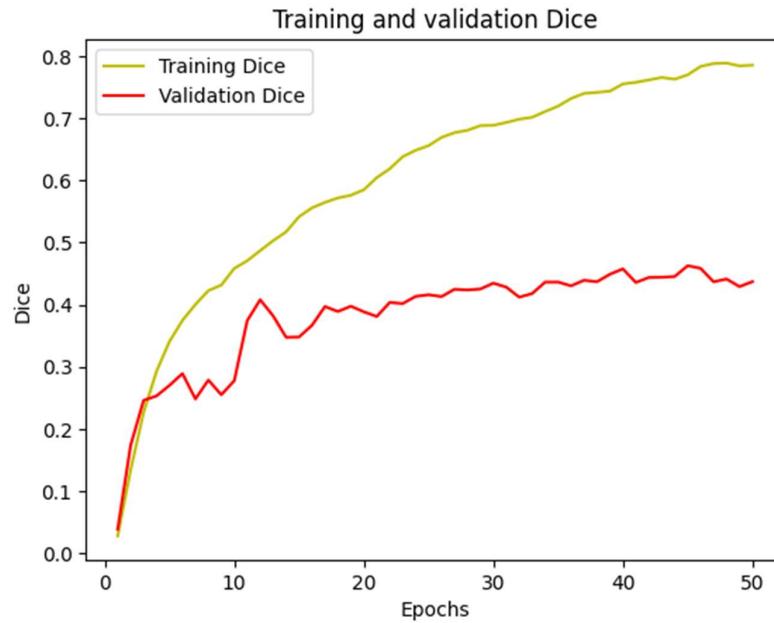
Figure 6.1: Dice Score for UNet using ImageNet
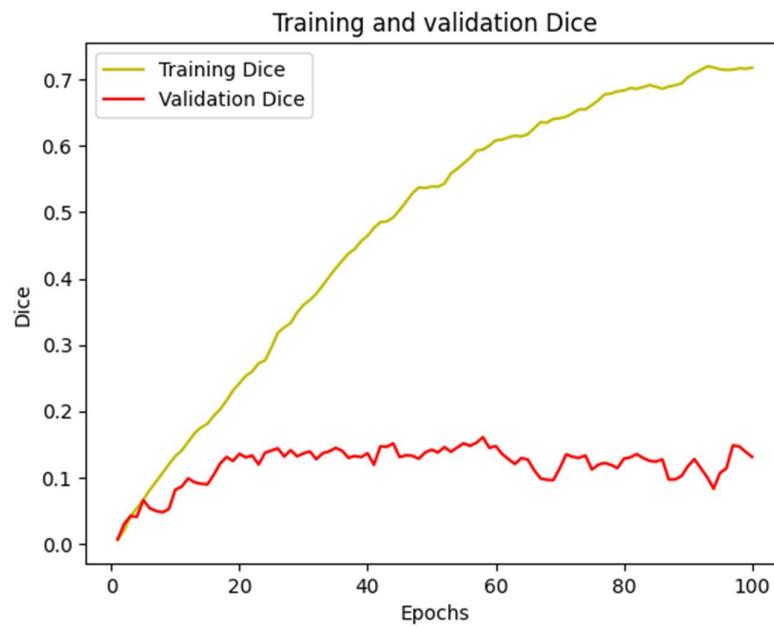
Figure 6.2: Dice Score for UNet+ 2D using ImageNet

41
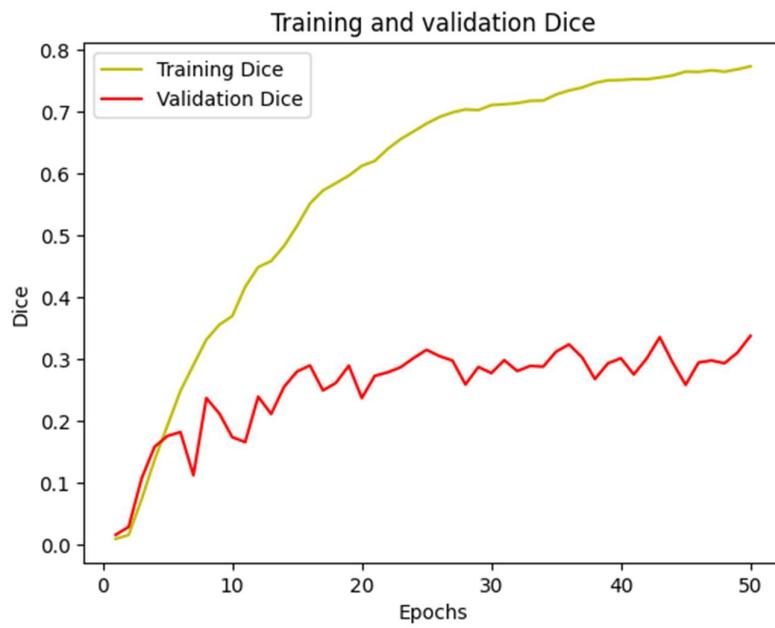
Figure 6.3: Dice Score for Attention UNet using ImageNet
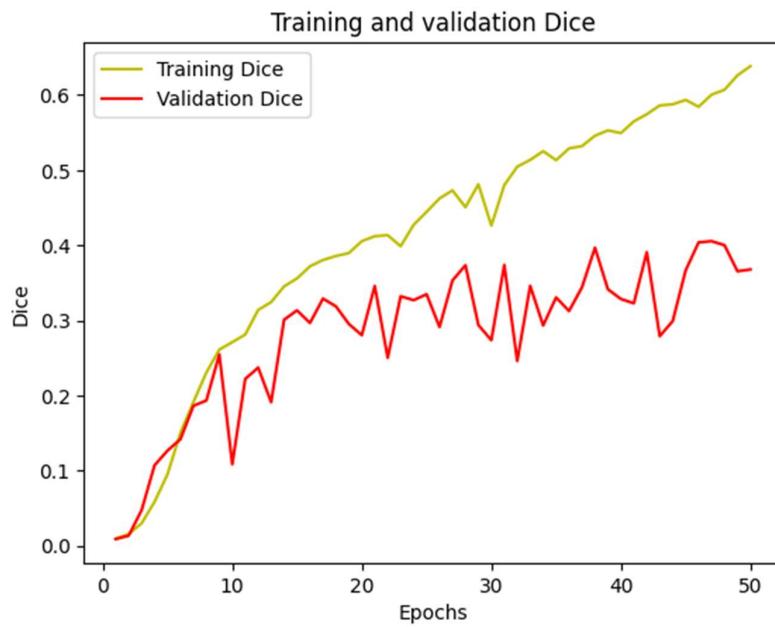
Figure 6.4: UNet trained from scratch
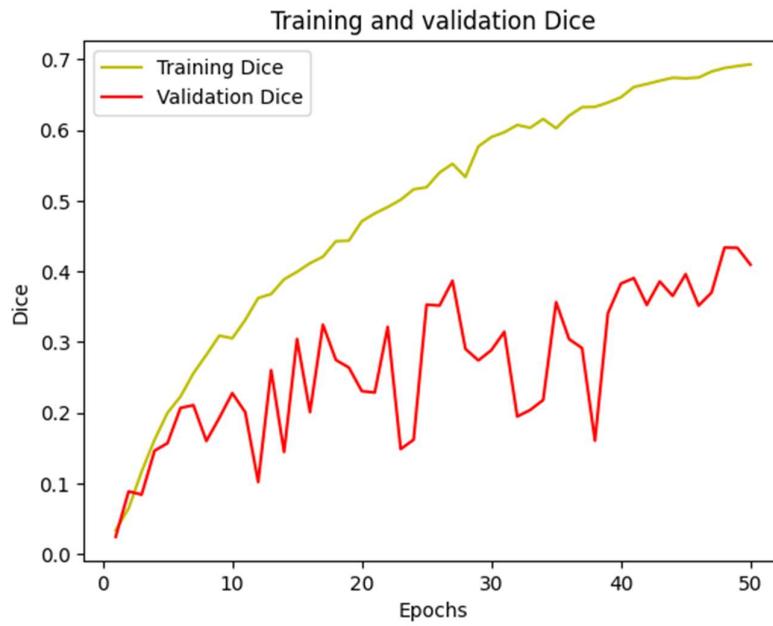
Figure 6.5: : R2 UNet trained from scratch
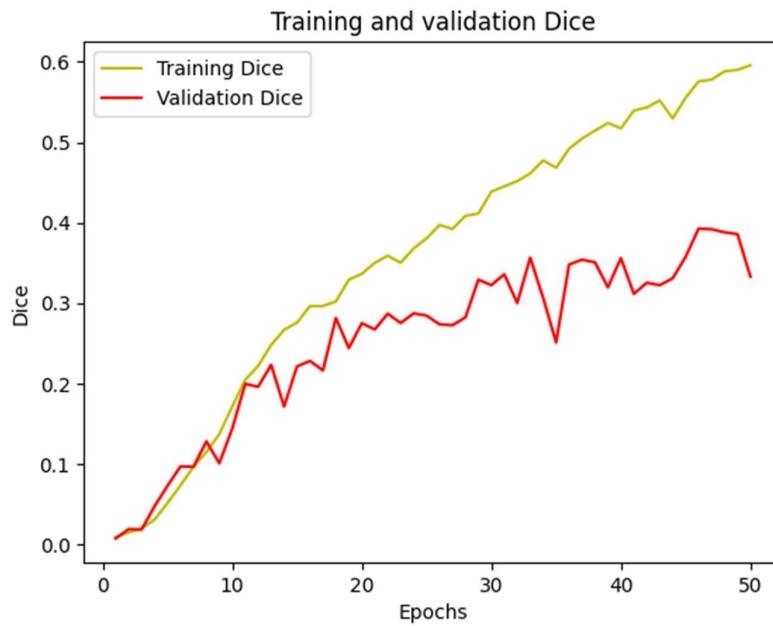
Figure 6.6: Attention UNet trained from scratch

6.2 RESULTS

Based on the results received from various models the resultant masks were generated

and compared original masks side by side. The results for each of the models were

characterised as good, average, and bad and are compiled in Appendix 2 of the report.

Figure 6.7 gives a glimpse of dataset while Table 6.3 gives average IoU testing values.
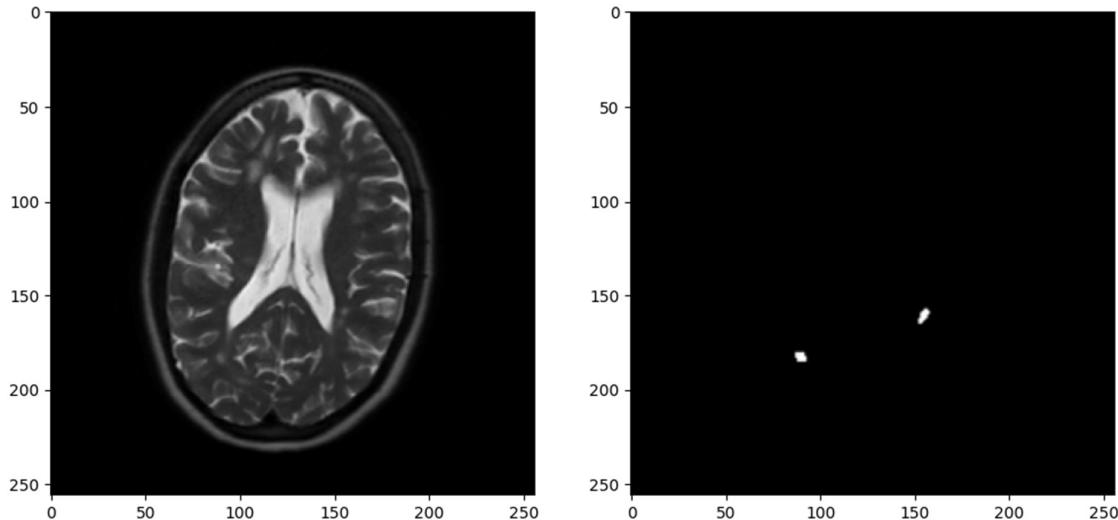
Figure 6.7: Dataset Glimpse

Table 6.3: IoU values

S. No.	Model name	IoU values
1.	UNet with VGG Backbone and ImageNet weights	0.6127
2.	UNet+ 2D with VGG Backbone and ImageNet weights	0.5126
3.	Attention UNet with VGG Backbone and ImageNet weights	0.5695
4.	UNet with VGG Backbone trained from scratch	0.6020
5.	R2 UNet with VGG Backbone trained from scratch	0.5990
6.	Attention UNet with VGG Backbone trained from scratch	0.5858

Summary

The chapter presents results and performance metrics checked for various models.

CHAPTER 7: CONCLUSION

The chapter finally draws conclusions at the end encompassing whole of the report in nutshell as a final chapter

7.1 CONCLUSION

When trying to improve the process of identifying Multiple Sclerosis (MS) lesions in brain MRI scans, the selection of model initialization weights becomes an important factor to consider. Utilizing pre-learned weights from models trained on extensive datasets such as ImageNet has been a common practice in the deep learning field for a significant period of time. Nevertheless, our thorough examination into the appropriateness of ImageNet model weights for MS lesion segmentation tasks has produced enlightening discoveries that question this practice.

During our extensive testing, we carefully examined the performance of several U-Net architectures with deep feature extraction backbones. I especially analysed the effectiveness of models that were started using ImageNet weights compared to those that were fine-tuned specifically for MS lesion segmentation. The findings regularly showed that fine-tuned models outperformed their ImageNet-initialized counterparts in terms of segmentation accuracy and resilience, despite the broad and diversified nature of the ImageNet training data. It was also noteworthy that the models with skip connections perform poorly

This unforeseen result implies that although ImageNet-pretrained models provide a strong foundation for many computer vision tasks, they may not be well suited for the subtle complexities of MS lesion segmentation. A more customized approach to model initialization is required due to the distinct features of MS lesions, such as their diverse forms, sizes, and intensities.

Fine-tuning, a technique where the model starts with pre-trained weights and then further refined using data unique to MS lesion segmentation, has proven to be a more effective method. This technique enables the model to adjust its acquired features to the complexities of MS lesion characteristics, leading to improved performance on segmentation tasks.

The significance of our discoveries goes beyond the segmentation of MS lesions, emphasizing wider aspects in the processing of medical images. The text emphasizes the significance of optimizing tasks specifically and highlights the constraints of using general-purpose pre-training in medical imaging applications.

The project proposes a fundamental change in the way models are initialized in the medical imaging field. My findings highlight the need of fine-tuning segmentation techniques for specific tasks, which can lead to more precise, dependable, and clinically applicable methods for diagnosing and planning therapy for multiple sclerosis (MS).

The work adds to the continuous development of deep learning approaches in medical image analysis by questioning commonly accepted beliefs and offering practical insights. This study establishes the foundation for future research efforts

focused on utilizing deep learning to enhance patient treatment and results in the field of neurological illnesses such as Multiple Sclerosis.

Summary

The chapter finally concludes the work and brings the essence of the work together in the last to bring it forward to summarize what was learned and what was gained during the experimentation.

CHAPTER 8: FUTURE SCOPE

The chapter discusses about the future probabilities and possibilities in the domain of Automated MS lesion segmentation and classification.

8.1 VISION AHEAD

Utilizing sophisticated U-Net architectures with deep feature extraction backbones for Multiple Sclerosis (MS) lesion segmentation offers promising prospects for future study and development. As we consider the future, several encouraging paths become apparent, providing chances to improve the precision, effectiveness, and practicality of MS lesion segmentation algorithms.

An area with significant potential for progress is computing power. Due to the swift advancement of hardware technologies such as GPUs and TPUs, there is a greater potential for increased computing resources. This can lead to faster model training and inference, enabling the investigation of more intricate architectures and bigger datasets. The enhanced computing capability has the potential to open up new possibilities in MS lesion segmentation. This means that more advanced models may be used to accurately capture the subtle details of lesion form and distribution. Moreover, the future of MS lesion segmentation is closely connected to the caliber and amount of accessible datasets. Although current datasets have provided a strong basis for study, it is crucial to make deliberate efforts to gather larger, more varied, and higher-quality information. These datasets, which contain annotated MS lesion data

from different imaging methods and patient demographics, might offer useful knowledge on the diversity of MS lesions and improve the ability of segmentation algorithms to make generalizations.

Progress in model design also holds the potential to influence the future of MS lesion segmentation. The SWIN transformer architecture is an innovative technology that brings about a fundamental change in image processing. It has the potential to enhance the ability to capture long-range connections and spatial correlations within MRI images. The use of these sophisticated models into MS lesion segmentation processes has the potential to enhance the accuracy and reliability of segmentation.

Furthermore, the idea of ongoing learning and adaptable modeling offers fascinating prospects for the future of MS lesion segmentation. Models that have the ability to be trained online and receive input from physicians and radiologists have the potential to be continuously improved and adjusted in real-time to changing clinical situations. This feedback loop guarantees that segmentation models stay pertinent and efficient in response to evolving patient demographics, illness features, and diagnostic techniques.

Summary

The future of MS lesion segmentation is marked by a coming together of progress in processing power, dataset quality, model design, and adaptive learning approaches. By capitalizing on these collaborative effects, researchers and practitioners can further advance the limits of precision and effectiveness in the segmentation of MS lesions.

REFERENCES

[1] Pinto, M.F., Oliveira, H., Batista, S. et al. Prediction of disease progression and outcomes in multiple sclerosis with machine learning. Sci Rep 10, 21038 (2020). https://doi.org/10.1038/s41598-020-78212-6

[2] N. Koch-Henriksen and P. S. Sørensen, "The changing demographic pattern of multiple sclerosis epidemiology," The Lancet Neurology, vol. 9, no. 5, pp. 520–532, May 2010. doi:10.1016/s1474-4422(10)70064-8

[3] J. H. Yoon and E.-K. Kim, "Deep learning-based Artificial Intelligence for Mammography," Korean Journal of Radiology, vol. 22, no. 8, p. 1225, 2021.

[4] E.-K. Kim, H.-E. Kim, K. Han, B. J. Kang, Y.-M. Sohn, O. H. Woo, and C. W. Lee, "Applying data-driven imaging biomarker in mammography for breast cancer

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Lecture Notes in Computer Science, pp. 234–241, 2015. doi:10.1007/978-3-319-24574-4_28

[6] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net Architecture for Medical Image segmentation," Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11, 2018. doi:10.1007/978-3-030-00889-5_1

[7] "Multiple sclerosis," National Institute of Neurological Disorders and Stroke, https://www.ninds.nih.gov/health-information/disorders/multiple-sclerosis (accessed May 3, 2024).

[8] MB Patwardhan, DB Matchar, GP Samsa, DC McCrory, RG Williams, and TT Li. Cost of multiple sclerosis by level of disability: a review of literature. Multiple Sclerosis Journal, 11(2):232–239, 2005.

[9] "Multiple sclerosis (MS)," Johns Hopkins Medicine, https://www.hopkinsmedicine.org/health/conditions-and-diseases/multiple-sclerosis-

ms#:~:text=Key%20points%20about%20multiple%20sclerosis&text=Others%20may%20lose%20the%20ability,and%20bowel%20and%20bladder%20incontinence. (accessed May 5, 2024).

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Lecture Notes in Computer Science, pp. 234–241, 2015. doi:10.1007/978-3-319-24574-4_28

[11] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net Architecture for Medical Image segmentation," Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11, 2018. doi:10.1007/978-3-030-00889-5_1

[12] H. Huang et al., "UNet 3+: A full-scale connected unet for medical image segmentation," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2020. doi:10.1109/icassp40776.2020.9053405

[13] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-A: A deep learning framework for semantic segmentation of remotely sensed data," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 162, pp. 94–114, Apr. 2020. doi:10.1016/j.isprsjprs.2020.01.013

[14] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa,K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, andD. Rueckert, ''Attention U-Net: Learning where to look for the pancreas,''2018, arXiv:1804.03999

[15] F. Moazami, A. Lefevre-Utile, C. Papaloukas, and V. Soumelis, "Machine learning approaches in study of multiple sclerosis disease through Magnetic Resonance Images," Frontiers in Immunology, vol. 12, Aug. 2021. doi:10.3389/fimmu.2021.700582

[16] P. A. Narayana et al., "Deep learning for predicting enhancing lesions in multiple sclerosis from Noncontrast MRI," Radiology, vol. 294, no. 2, pp. 398–404, Feb. 2020. doi:10.1148/radiol.2019191061

[17] R. Seccia et al., "Machine learning use for prognostic purposes in multiple sclerosis," Life, vol. 11, no. 2, p. 122, Feb. 2021. doi:10.3390/life11020122
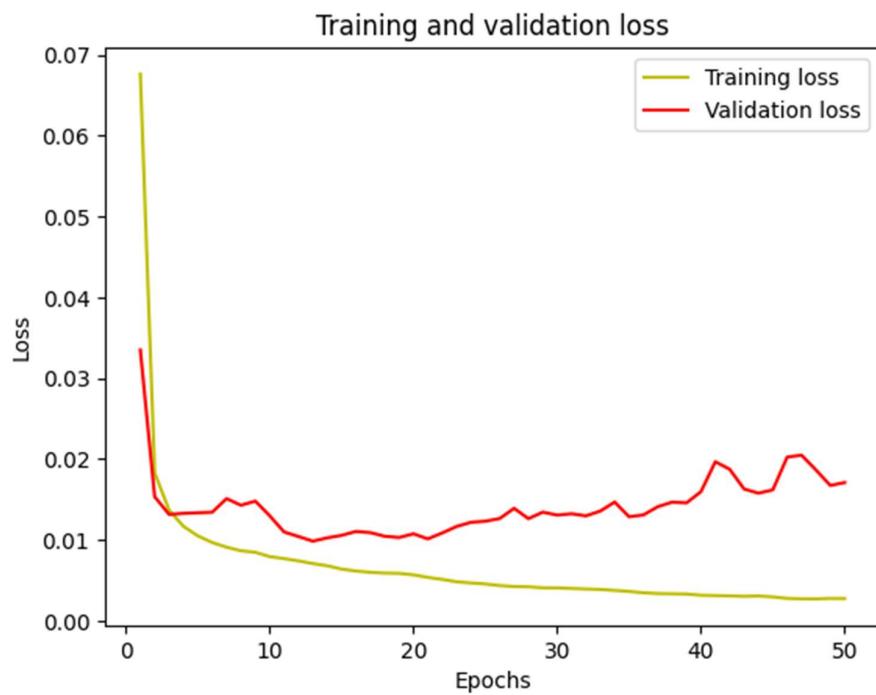
[18] N. Aslam et al., "Multiple sclerosis diagnosis using machine learning and Deep learning: Challenges and opportunities," Sensors, vol. 22, no. 20, p. 7856, Oct. 2022. doi:10.3390/s22207856

[19] Y. Zhao et al., "Exploration of machine learning techniques in predicting multiple sclerosis disease course," PLOS ONE, vol. 12, no. 4, Apr. 2017. doi:10.1371/journal.pone.0174866

[20] M. F. Pinto et al., "Prediction of disease progression and outcomes in multiple sclerosis with Machine Learning," Scientific Reports, vol. 10, no. 1, Dec. 2020. doi:10.1038/s41598-020-78212-6

[21] A. M. Muslim et al., "Brain MRI dataset of multiple sclerosis with consensus manual lesion segmentation and patient meta information," Data in Brief, vol. 42, p. 108139, Jun. 2022. doi:10.1016/j.dib.2022.108139

[22] P. Iakubovskii, 'Segmentation Models', GitHub repository. GitHub, 2019.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv.org, 10-Apr-2015. [Online]. Available:https://arxiv.org/abs/1409.1556.

[24] A. G. Howard et al., 'MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications', arXiv [cs.CV]. 2017.

[25] Y. Sha, 'Keras-unet-collection', GitHub repository. GitHub, 2021.

[26] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, 'Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation', CoRR, vol. abs/1802.06955, 2018.

[27] Z. Liu *et al.*, 'Swin Transformer: Hierarchical Vision Transformer using Shifted Windows', in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

APPENDIX I
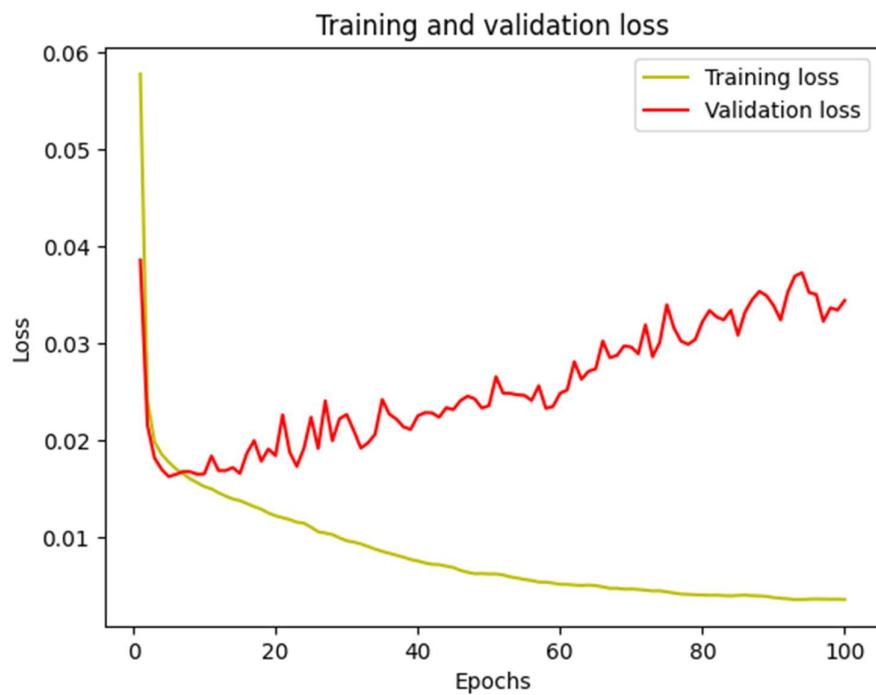
The following are the figures and table referred in Chapter 6.

Appendix 1 Table 1: Accuracy vs Loss
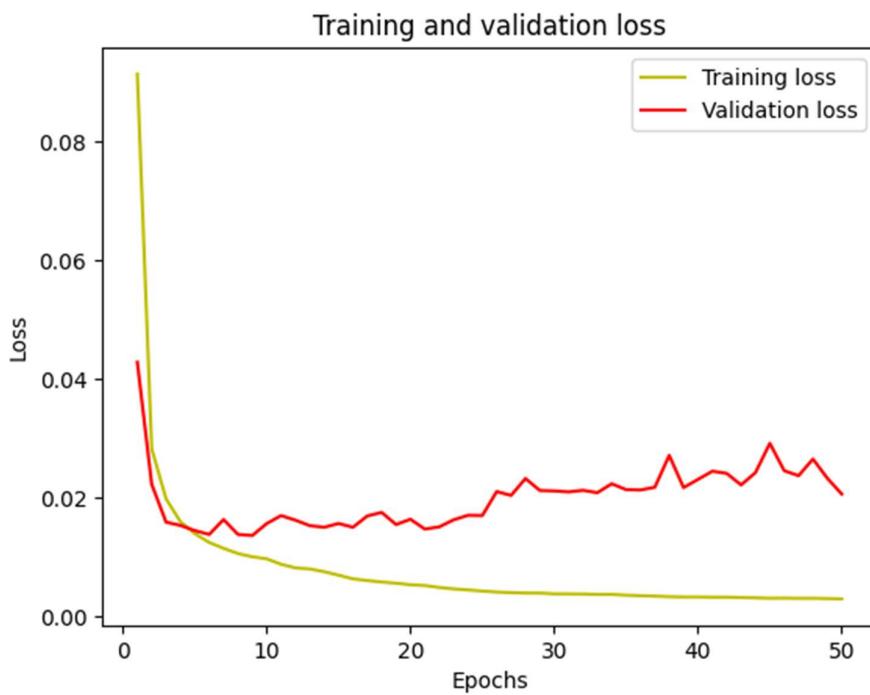
S No.	Model name	Accuracy		Loss	
		Training	Validation	Training	Validation
1.	UNet with VGG Backbone and ImageNet weights	99.66%	99.53	0.27%	1.17%
2.	UNet+ 2D with VGG Backbone and ImageNet weights	99.64%	99.42%	0.36%	3.44%
3.	Attention UNet with VGG Backbone and ImageNet weights	99.65%	99.49%	0.9%	1.3%
4.	UNet with VGG Backbone trained from scratch	99.60%	99.55%	0.47%	1.0%
5.	R2 UNet with VGG Backbone trained from scratch	99.62	99.54	0.4%	1.1%
6.	Attention UNet with VGG Backbone trained from scratch	99.58%	99.54%	0.05%	0.1%
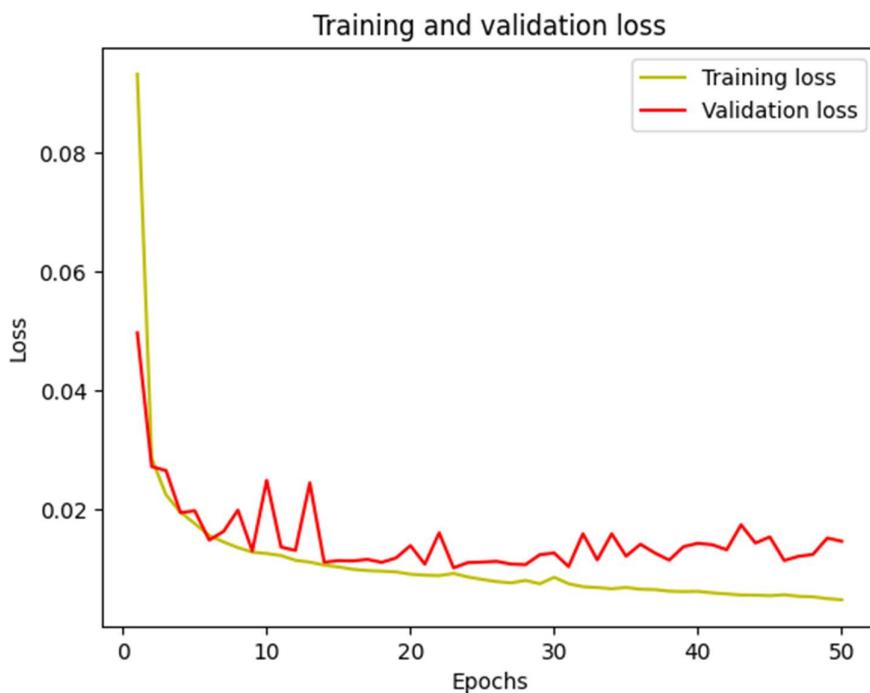
Appendix 1 Figure 1: Accuracy vs Loss for UNet on ImageNet
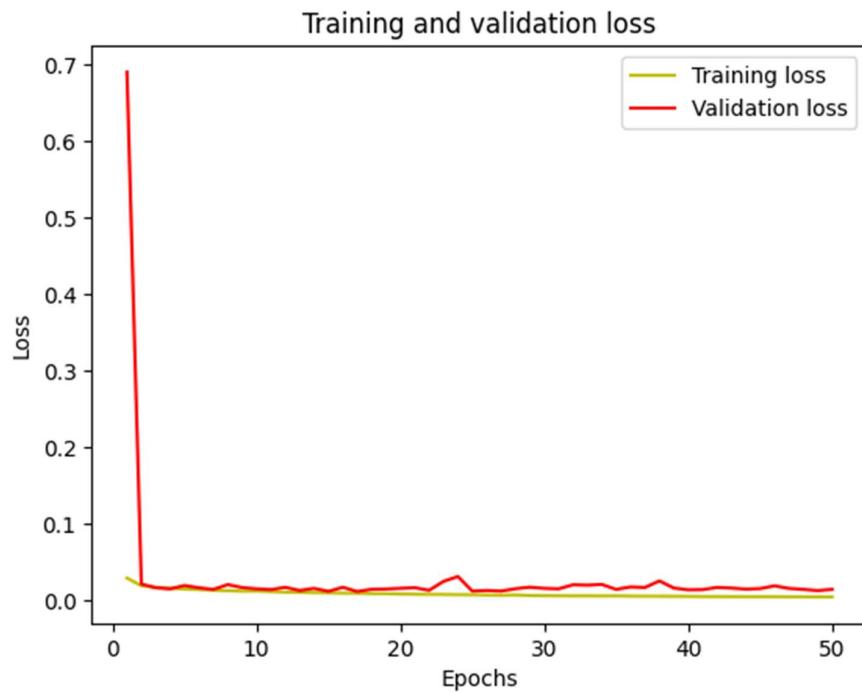
Appendix 1 Figure 2: Accuracy vs Loss for UNet+ 2D on ImageNet
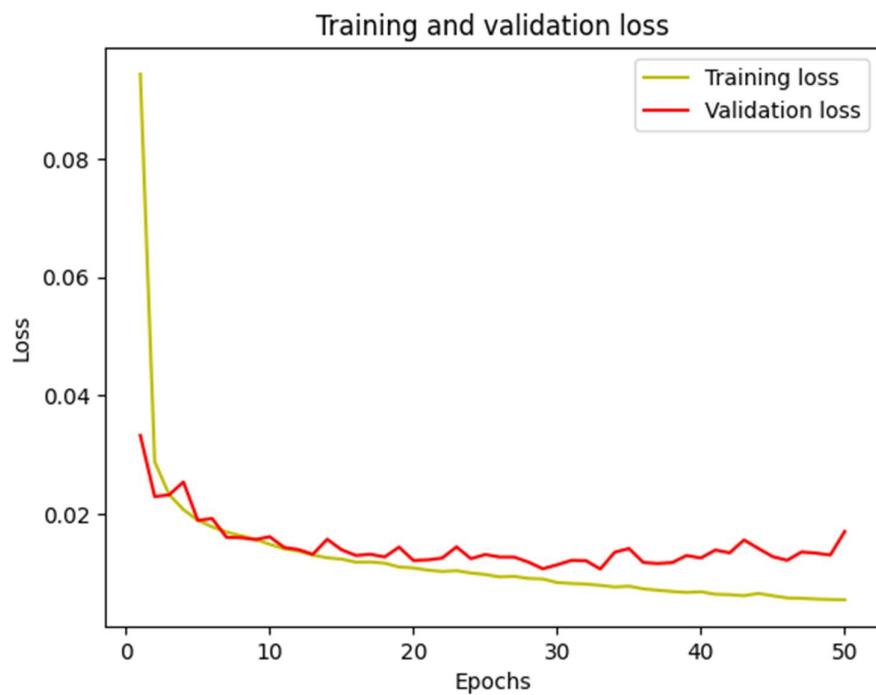
Appendix 1 Figure 3: Accuracy vs Loss for Attention UNet on ImageNet

Appendix 1 Figure 4: Accuracy vs Loss for UNet from scratch
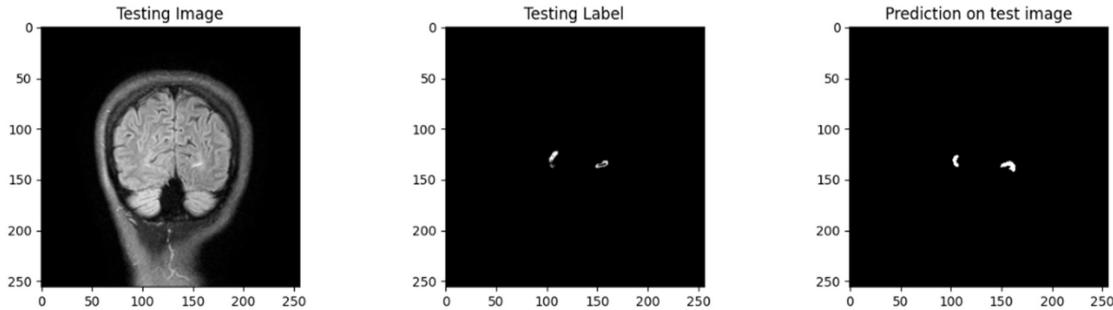
Appendix 1 Figure 5: Accuracy vs Loss for R2 UNet from scratch
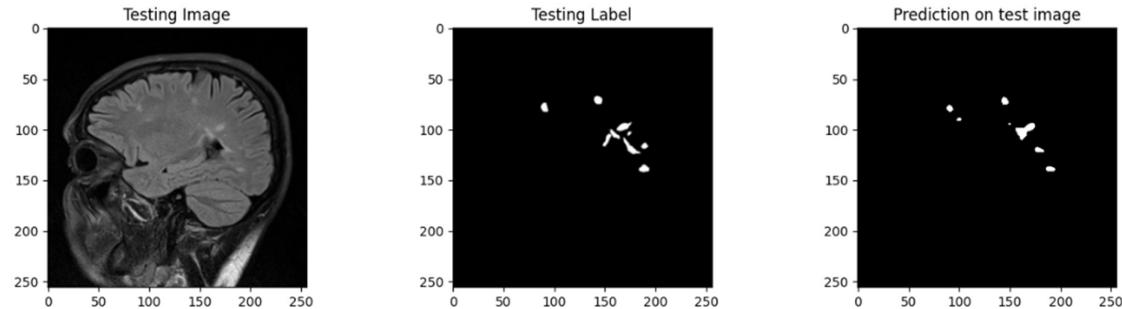
Appendix 1 Figure 6: Accuracy vs Loss for R2 UNet from scratch

APPENDIX II
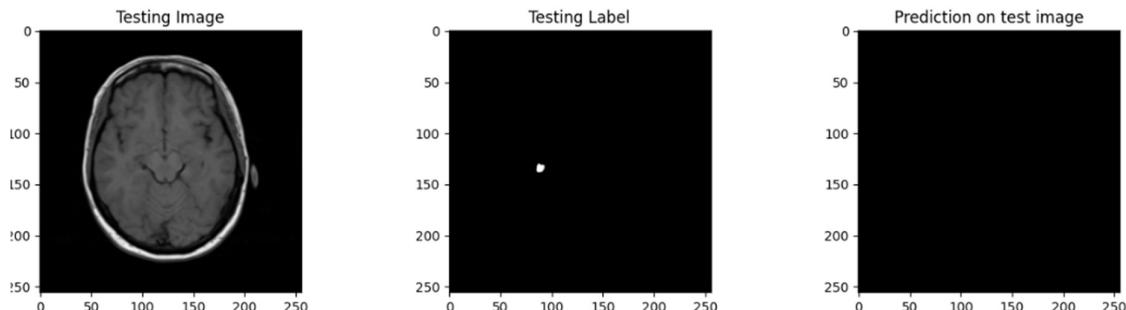
The Appendix contains all the images of generated masks from all 6 models.
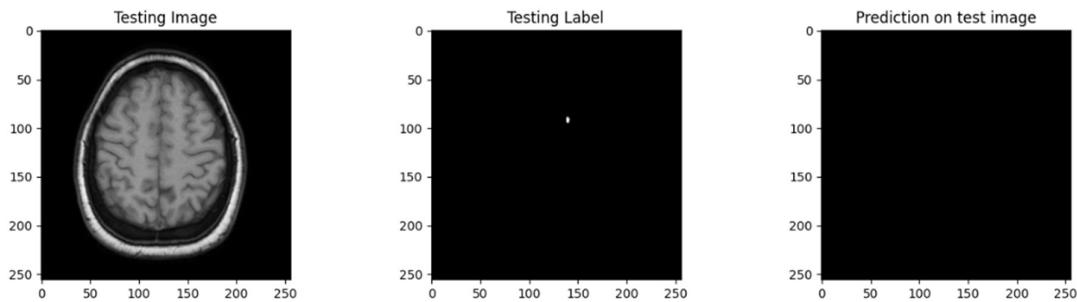
Appendix 2 Figure 1: Good quality mask from UNet using ImageNet Weights
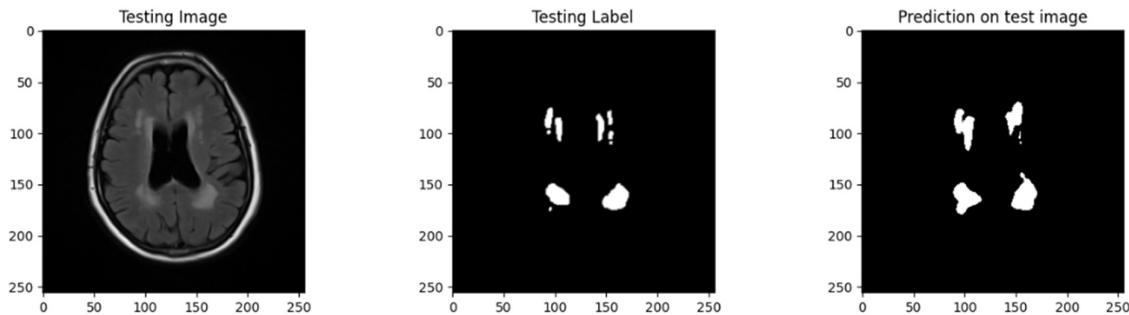
Appendix 2 Figure 2: Average quality mask from UNet using ImageNet
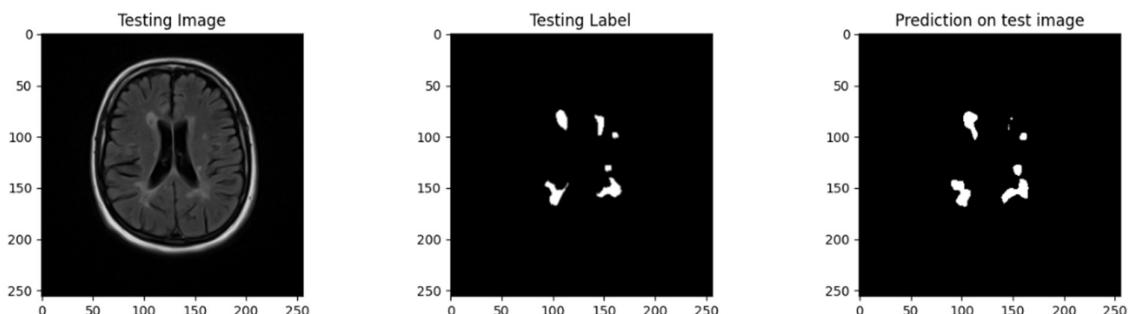
Appendix 2 Figure 3: Bad quality mask from UNet using ImageNet
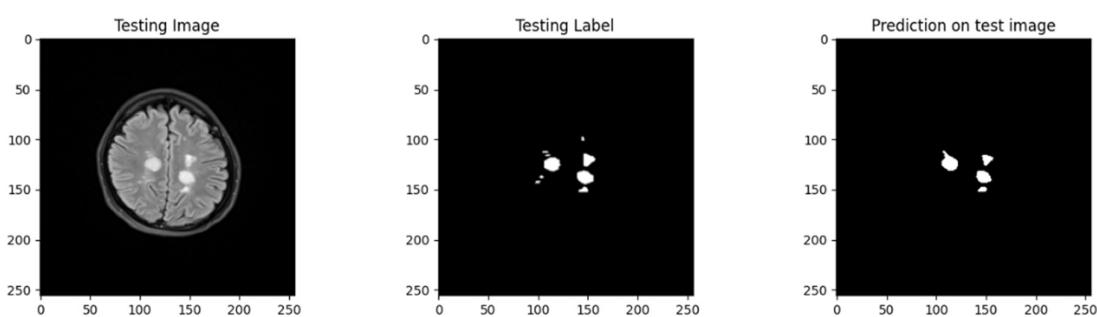
Appendix 2 Figure 4: Good quality mask from UNet trained from scratch

Appendix 2 Figure 5: Average quality mask from UNet trained from scratch
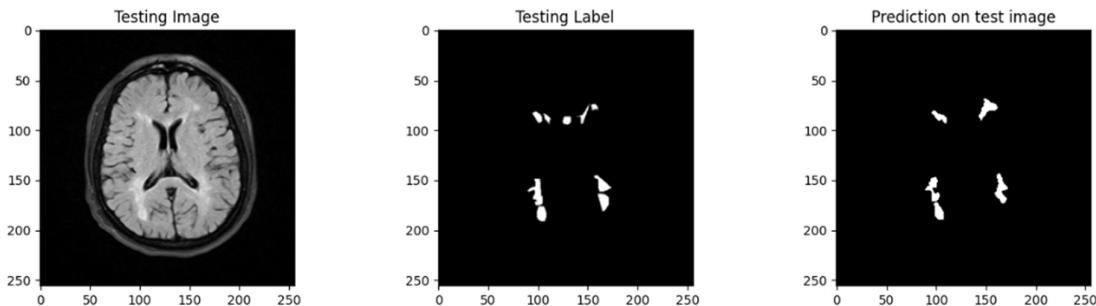
Appendix 2 Figure 6: Bad quality mask from UNet trained from scratch
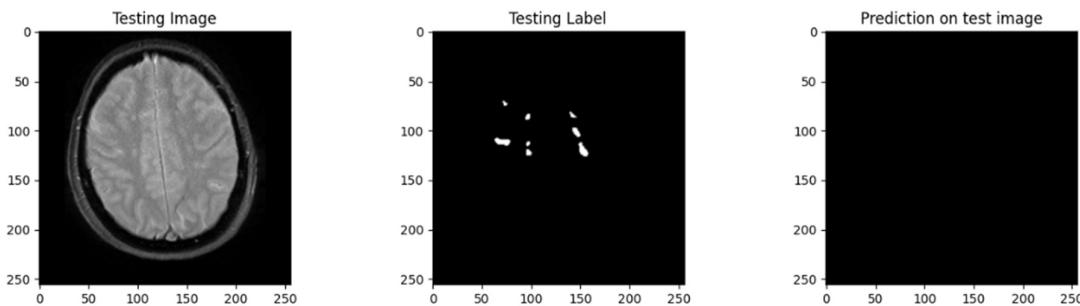
Appendix 2 Figure 7: Good quality mask from UNet trained from scratch
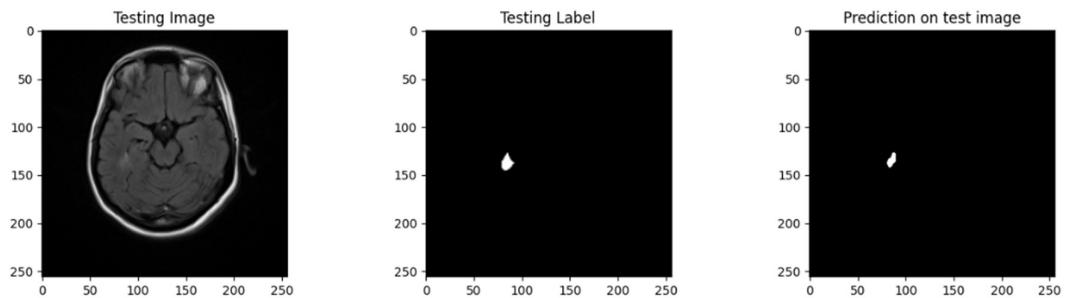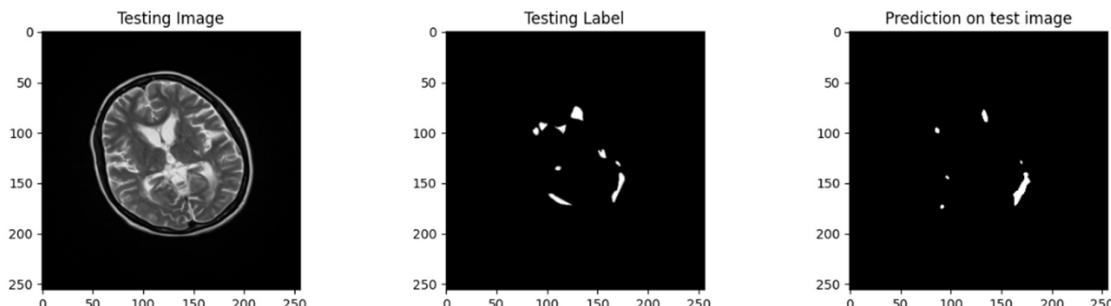
Appendix 2 Figure 8: Average quality mask from UNet trained from scratch

Appendix 2 Figure 9: Bad quality mask from UNet trained from scratch
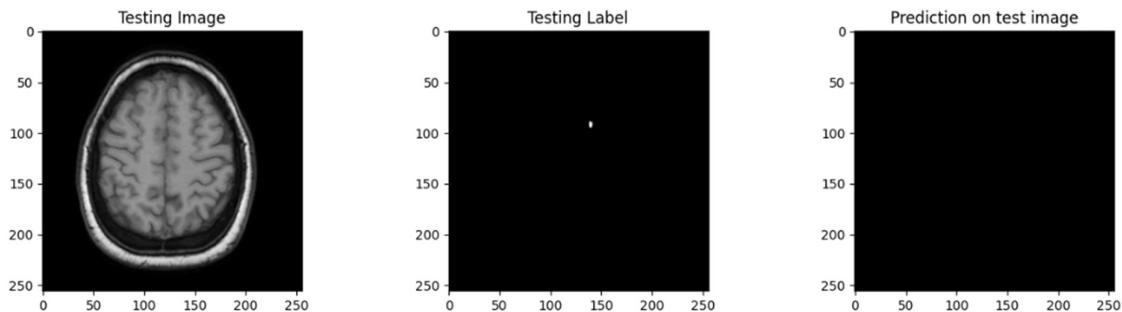
Appendix 2 Figure 10: Good quality mask from Attention UNet trained from scratch

Appendix 2 Figure 11: Average quality mask from Attention UNet trained from scratch

Appendix 2 Figure 12: Bad quality mask from Attention UNet trained from scratch