Master in Computational Colour and Spectral Imaging (COSI)

Transfer Learning for Brain Lesion Segmentation in 3D Pre-Clinical Magnetic Resonance Images

Master Thesis Report

Presented by

Mohammad Jaber Hossain

and defended at the

Norwegian University of Science and Technology

September 2023

Academic Supervisor(s): Professor Xiao-Zhi Gao
Host Supervisor: Professor Jussi Tohka, Dr. Juan Miguel Valverde
Jury Committee:

1. Dr. Mohib Ullah, NTNU
2. Dr. Rahat Khan

Submission of the thesis: 10th August 2023
Day of the oral defense: 4th September 2023

Abstract

This work initiates a novel exploration of domain transfer learning in pre-clinical image analysis, concentrating on the segmentation of brain lesion areas of mice MR images. Two transfer learning techniques, partially fine-tuning and fully fine-tuning the model were used with the pre-trained RatLesNetV2 model. Various training samples ranging from 8 to 40 images, both with and without data augmentation, were used to further evaluate these techniques. Utilizing MONAI framework, the research also assessed the 3D UNet, UNetR, and Dynamic Unet models. When compared to the others, the 3D UNet with augmentation performed most effectively, with an average Dice Coefficient (DC) of 0.88 and 95% Hausdorff distance (HD95) of 0.39 mm. Partial fine-tuning produced segmentations with an average DC of 0.85 and HD95 of 0.71 and these segmentations were approximately similar to the ones generated by 3D UNet and better than the segmentations produced by other MONAI models, indicating the potential adaptability of partial fine-tuning. Full fine-tuning provided insightful results but did not outperform the other approaches used in this study, demonstrating the intricate interaction between overfitting and adaptability in transfer learning. This study is one of the first to examine the possibility of transfer learning in pre-clinical settings, bringing an important approach to a different field. The results provide a solid platform for understanding the underlying mechanisms that led to these findings and expanding the scope of transfer learning's applications in pre-clinical image analysis.

Dedication

I thank the Almighty Allah for the boundless blessings, strength, and the series of opportunities that made the way for the completion of this thesis. I dedicate this thesis to my loving parents, who always believed in me, and to my supportive wife, who stood by my side during this journey. I'm equally grateful to my teachers, whose wisdom and guidance have been invaluable. This achievement is a testament to their unwavering faith and encouragement.

Acknowledgment

- I am deeply grateful to my supervisors, Prof. Xiao-Zhi Gao, Prof. Jussi Tohka, and especially Dr. Juan Miguel Valverde. Their unwavering guidance, insightful feedback, and kind encouragement have been the foundation of my journey. Their generosity in sharing knowledge and patience in mentoring has truly enriched my experience, and I feel immensely privileged to have been under their benevolent supervision.
- My heartfelt thanks to my COSI classmates who always stood by me. A special mention to Austin Ryan English, Abhinav Reddy Nimma, and Ali Raza Syed for their consistent support during challenging times.
- I extend my deepest gratitude to the EU and the dedicated coordinators of COSI for their continuous support throughout this COSI journey.

Acronyms

AD- Alzheimer's disease
ALS- Amyotrophic Lateral Sclerosis
ANFIS- Adaptive Neuro Fuzzy Inference System
CADx- Computer-aided diagnosis
CBCT- Cone-Beam Computed Tomography
CNN- Convolutional neural network
CT- Computed tomography
D-UNet- Dimension Fusion UNet
DSC- Dice Similarity Coefficient
DWI- Diffusion-weighted imaging
FLAIR- Fluid Attenuated Inversion Recovery
FOV- Field of view
HD95- 95th percentile of Hausdorff Distances
HHET- Heuristic Histogram Equalization Technique
ICE- Intracardiac echocardiograph
IVC- Inferior vena cava
MCAo- Middle cerebral artery occlusion
MONAI- Medical Open Network for Artificial Intelligence
MRI- Magnetic Resonance Imaging
MUSCLE Net- Multi-Scale Convolutional Label Evaluation Net
NMRI- Nuclear Magnetic Resonance Imaging
PET- Positron emission tomography
UNETR- UNET Transformers
US- Ultrasound

Contents

1 Introduction

1.1 Machine learning

Machine learning is a key area of artificial intelligence that includes the development and deployment of methods that enable systems to learn from and make decisions based on data. In essence, it is the process through which system independently enhance their performance through learning from data, without explicit programming for each individual job. A model is designed through training on data that recognizes patterns in data and applies those patterns to forecast future events or take action. Essentially, it is the representation that the algorithm creates after learning from the data and serving as the primary engine for drawing conclusions, generating predictions, or reaching judgments based on brand-new, undiscovered knowledge. Depending on the task at hand, machine learning models may range from simple linear regressions to intricate deep neural networks.

1.1.1 Transfer learning

Transfer learning(TL) is an approach in which a pre-trained model serves as the foundation for a new task, rather than starting from scratch to train a new model. This idea assumes that a relationship between two learning activities is necessary for transfer to occur. The ability to ride a bicycle makes learning to ride a motorbike much easier, according to real-world experience. In order to effectively tackle issues with a smaller number of labeled data or diverse data distributions, it involves reusing or regulating pre-existing models that have been pre-trained on a big dataset. The term TL refers to a variety of techniques, such as feature extraction (Zhuang et al., 2020), domain adaption (Chen et al., 2020), and fine-tuning previously trained models (Swati et al., 2019). The pretrained model's weights are updated during fine-tuning using a smaller task-specific dataset. Utilizing learned representations or features from a pretrained model as input to a new model trained on the target task is known as feature extraction. Domain adaptation is concerned with transferring knowledge from one domain to another

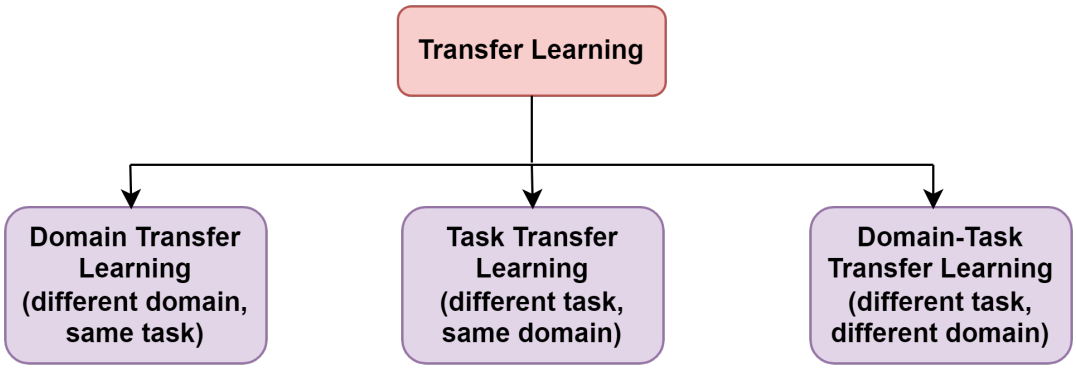

Figure 1.1: *Categories of transfer learning.*

while taking into account various data distributions. Therefore, the core idea is to use the information the pre-trained model learned from a previous task to improve the performance of the current model on an associated task. The pre-trained model often goes through training on a large dataset, usually linked to a different but related job. On the whole, traditional machine learning has achieved significant success in many real-world applications, but it has limits in certain situations. A large number of labeled training examples with distributions similar to test data are ideal for machine learning. In reality getting such data might be expensive, time-consuming, or impracticable. In order to address the aforementioned difficulties, transfer learning, seems to be a potential approach. The use of TL may be divided into three categories: different tasks, different domains, and different tasks and domains all together. When applied across various tasks within the same domain, TL takes advantage of inherent similarities for effective training; when applied across various domains but within the same task, it mitigates distribution shifts for adaptation; and in its most complex form, when applied across various tasks and different domains, it fosters a rich interplay of strategies for bridging disparities. This categorization provides a thorough understanding of the dynamic nature of TL, emphasizing its flexibility across a range of applications and inspiring further study into its underlying processes and optimization in a number of fields.

The problem related to labeled data is partly addressed by semi-supervised learning, which uses a substantial amount of unlabeled data to improve learning accuracy while minimizing the reliance on large labeled datasets. But in situations when simply collecting unlabeled data is difficult, standard models often fall short, which is a pretty common problem in medical image analysis due to patients privacy concerns and more related issues. So, in this field of medical image analysis, transfer learning has emerged as a viable method for using the information learned from previously trained models on large amounts of data (Cheplygina et al., 2019; Zhuang

et al., 2020). The approach also assists in overcoming the problem of inadequate labeled data in medical imaging applications by modifying these pre-existing models to new tasks or domains. However, prior research (Chen et al., 2018; Çiçek et al., 2016; Ronneberger et al., 2015) mainly examined segmentation accuracy rather than investigating into the specifics of transfer learning and the variations between models trained with and without it. This knowledge gap emphasizes the need for a more thorough understanding of the processes and differences in transfer learning for segmentation. Examining these elements will provide useful insights that will help deep learning models perform better and be more applicable to the analysis of medical images. Alongside, it could be worthwhile to investigate the suitability of including a certain amount or proportion of new images to optimize an existing model's ability to adapt to a new dataset.

1.2 3D Magnetic Resonance Imaging

A non-invasive radiology procedure called magnetic resonance imaging (MRI) creates comprehensive images of internal organs and bodily functions using magnetic fields and radio waves (Westbrook and Talbot, 2018). It works by detecting signals coming from hydrogen nuclei within the body while being affected by strong magnetic fields. By using unique contrasts between areas as a result of variations in molecular mobility and interaction, this technique is able to create three-dimensional photographs of live tissue. The varying intensities of these signals create distinct contrasts, allowing for the differentiation between alterations in various tissue types. By altering the pulse sequence parameters, MRI scanners may capture images with various contrasts. Importantly, unlike other imaging techniques like computed tomography (CT) or Positron emission tomography (PET), MRI does not use radiation, making it a desirable imaging technique for clinical and pre-clinical research as well as for use in healthcare. MRI provides higher contrast between soft tissues than other procedures like CT scans. The Nuclear Magnetic Resonance Imaging (NMRI) method,makes use of the body's abundant hydrogen atoms, which are mostly found in water and fat, to map their locations. MRI has grown in popularity despite concerns about patient comfort, cost-effectiveness, and over-diagnosis. This is due to the technology's flexibility in both medical and non-medical applications as well as its cutting-edge methods, such as diffusion MRI and functional MRI (Le Bihan, 2012). Besides, a state-of-the-art imaging technique called three-dimensional MRI produces extremely accurate volumetric images of the internal human body structures, particularly the brain. In contrast to its two-dimensional version, 3D MRI collects data in a large, contiguous volume that is then assembled into a 3D dataset (Gordillo et al., 2013; Fedorov et al., 2012). This approach permits the observation of complicated anatomical features from

numerous angles and in many planes (axial, sagittal, and coronal) without any loss of resolution (Ohishi et al., 2005). The ability to see detailed visualization of different properties of tissues and minor anomalies is made possible by the increased resolution. Due to its capacity to record high-contrast images of numerous brain areas, MRI is an essential technique for detecting early-stage cerebral disorders. Distinguishing between different tissue types and neuropathologies via unique grey levels is made possible by the adjustability of MRI acquisition settings (Tian and Fan, 2007). In the realm of medical image segmentation research, MRI has taken the lead because they provide better contrast than CT images (Balafar et al., 2010).

Precise images of the brain's structures provided by 3D MRI allow for the early detection and diagnosis of Alzheimer's disease (AD) as well as the identification of the illness's signature patterns of brain shrinkage (McEvoy and Brewer, 2010). In addition to Alzheimer's disease, Parkinson's disease, Huntington's disease, and Amyotrophic Lateral Sclerosis (ALS) may all be diagnosed and monitored with the use of 3D MRI. High-resolution 3D MRI scans are able to identify precise anatomical alterations in the brain associated with certain illnesses (Eskreis-Winkler et al., 2017). The diagnosis and staging of liver disorders, such as cirrhosis and hepatocellular carcinoma, as well as the evaluation of liver volume and function before and after liver transplantation, were all conducted using 3D MRI (Min et al., 2020). In oncology, 3D MRI makes it easier to find, precisely localize, stage, and monitor cancers (Haie-Meder et al., 2005), especially those that are found in the brain, breast, and prostate. This method also assists in the diagnosis of musculoskeletal conditions (Balassy and Hörmann, 2008) like osteoarthritis and sports injuries, and its growing usage in cardiology (Milano et al., 2019) enables the assessment of heart shape and function as well as the detection of conditions like congenital abnormalities and myocardial infarction.

The fluid attenuated inversion recovery (FLAIR) and T2-weighted MRI of the brain show highly intense lesions that are typically present in the elderly and are linked to ischemia or demyelinating events. These lesions are important indicators for highlighting the need of accurate quantification for both diagnosis and prognosis of these diseases (Atlason et al., 2019). A wide range of applications, including the study of brain architecture, the identification of damaged areas, the detection of tumors, and the brain regions segmentation, heavily rely on magnetic resonance imaging (MRI) data. These applications make use of the detailed and in-depth data provided by MRI scans to enable precise analysis and interpretation in the clinical and research fields (Singh and Bala, 2021; Despotović et al., 2015). Figure 1.2 shows a number of T2 MRI slices taken from the dataset used for this study's analysis.

The study of medical imaging data with image segmentation has become an essential step, influencing both pre-clinical and clinical (Pham et al., 2000)

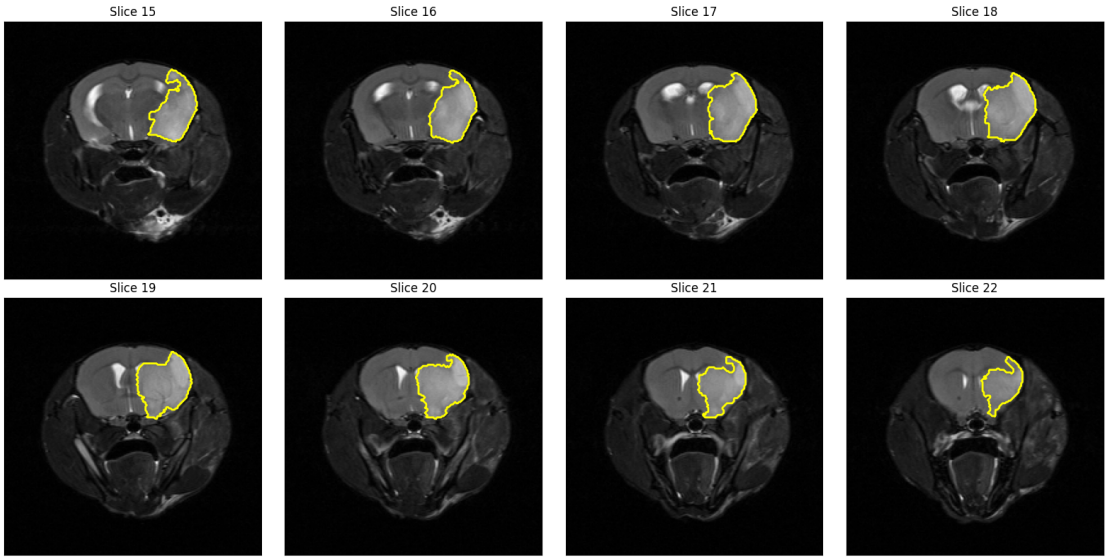

Figure 1.2: *Representative MR image from the dataset used in this study, (The ischemic stroke lesion is indicated in yellow.).*

research. In order to make the representation of an image more understandable and straightforward to analyze, it relates to the division of an image into several segments or groups of pixels. The location of objects and borders, measurement of tissue volumes, analysis of anatomical structure, diagnosis of diseases, and planning of surgical treatments all depend on this process. The importance and development of image segmentation techniques are examined in this article in the separate but related fields of clinical and pre-clinical investigations (Pham et al., 2000; LaLonde et al., 2021).

1.3 Lesion Segmentation

Medical professionals rely heavily on lesion segmentation from 3D MR images for diagnosing, monitoring, and treating a variety of neurological diseases. Clinicians and researchers may learn important information about the location, volume, and evolution of pathological alterations in the brain by isolating and measuring lesions. Such knowledge may be very helpful in determining the severity and course of the condition, guiding therapy decisions, and assessing therapeutic outcomes. Additionally, precise lesion segmentation in longitudinal investigations might point up minor changes over time that might not be immediately obvious in a clinical assessment. Lesion load and growth are often tightly correlated with the clinical symptoms and prognosis of individuals with neurodegenerative disorders, stroke,

or brain malignancies (Schwarz, 2021). In order to provide individualized patient treatment and improve the field of neuroimaging research, accurate lesion segmentation from 3D MRI data is essential. Accurate segmentation of stroke lesions enables accurate estimation of lesion volumes, assists in identifying key anatomical structures affected by stroke, and offers invaluable insights into the pathophysiology at work. It is essential for individualized treatment plans, patient care that is at the highest level, and enhanced clinical results. Additionally, when applied to large-scale investigations, this approach may advance our knowledge of disease processes, perhaps paving the way for the creation of novel treatment approaches. The methods used to segment 3D brain stroke lesions include a wide spectrum of innovations in computational methods and imaging technology. Traditional approaches include thresholding (Erdi et al., 1997), region-growing (Węgliński and Fabijańska, 2011), and segmentation based on atlases (Jyothi and Singh, 2023), but more advanced strategies use machine learning and deep learning techniques. Specifically, convolutional neural networks (CNNs) have shown promising outcomes in the automated and semi-automatic segmentation of stroke lesions by exploiting the spatial hierarchy and contextual information inside 3D MR images. The segmentation accuracy is further improved by combining structural MRI data with other types of imaging data, such as diffusion-weighted imaging (DWI). Precision, dependability, and clinical usefulness of 3D lesion segmentation in the treatment of brain strokes are continually being improved by the consistent improvement of these methods (Krishnapriya and Karuna, 2023). The need for 3D preclinical MR imaging brain stroke lesion segmentation is without any shadow of doubt important, although research in this area has been comparatively insufficient. Preclinical research, particularly those involving animal participants, sometimes need for a thorough examination of hundreds of 3D images, which presents significant segmentation issues. Understanding the pathophysiological processes of stroke is crucial for directing the development of focused therapy approaches. This requires the capacity to precisely characterize and segment brain stroke lesions. Manual segmentation, although playing a crucial role, may be prohibitively time-consuming and is often impracticable in investigations requiring huge sets of 3D images. The job is further complicated by the moderate inter and intra-rater agreement that results from the challenge of determining lesion borders. Pre-clinical 3D lesion segmentation methods are not as advanced as their clinical equivalents, and there is still room for improvement. Automation using computational techniques is crucial since conventional methods have significant speed and reliability issues. However, the use of advanced methods like machine learning and deep learning is still in its early stages in the pre-clinical setting. The difficulty is exacerbated by the particular demands of pre-clinical investigations, where specialized segmentation techniques are required because of differences in anatomy across various species.

It is not only necessary technically, but also a crucial step towards expediting pre-clinical research, opening the door for more effective therapies and a better understanding of brain stroke, to develop reliable, accurate, and automated techniques for pre-clinical 3D lesion segmentation.

1.4 Motivation

With its ability to provide non-invasive, high-resolution insights into the complex anatomy of the brain, MR imaging is a crucial technique in this context. Despite these benefits, lesion segmentation in MR images is still a difficult operation that often requires laborious manual annotation by qualified radiologists. The accuracy and dependability of diagnosis may be threatened by this process, which is vulnerable to both inter and intra-observer variability. The method of automated MRI lesion segmentation has shown encouraging results because to developments in machine learning, especially deep learning. But because of privacy issues, a lack of availability, and the exorbitant expense of human annotation, these systems often need large, annotated datasets for training. Furthermore, owing to differences in MRI collection techniques and patient groups, a problem known as domain shift, deep learning models trained on one dataset sometimes do not transfer well to other datasets. While most prior research focused on clinical MRI segmentation, our work explores the comparatively unexplored field of pre-clinical MRI segmentation.

1.5 Contributions

With the use of a domain transfer learning strategy, we conducted a subtle investigation involving two different but related species, specifically rats and mice. Despite the fact that different species employ different methods for image acquisition, this study focuses on 3D MRI scans. By using the similarities and differences between these scans, we take advantage of the underlying patterns and complexity in the MR images to accomplish precise segmentation of brain lesions. The use of RatLesNetV2 as a pre-trained model, in combination with several training sample sizes and augmentation approaches, significantly complicated and enhanced the work. The following are the contributions we made:

- We thoroughly assessed the performance of a few cutting-edge architectures created for 3D segmentation tasks. These included designs such as 3D UNet (Belue et al., 2022), UNetR (Hatamizadeh et al., 2021a) and DynUnet (Isensee et al., 2021). These models are all regarded for being reliable and efficient at processing 3D data, which makes them the perfect candidates for our study.

Our goal was to find the segmentation model that produces most precise segmentation outcomes.

- We notably address the segmentation of brain lesions in 3D mice MRIs by using the transfer learning approach with an emphasis on both partial and full fine-tuning, using pre-trained model RatLesNetV2 (Valverde et al., 2020) which was trained on Rat MR images. The main goals are to improve the study of neurological disorders using different animal models and to dive more deeply into neuroimaging diagnostics. Not only will this result in the identification of inherent patterns in MR images, it will also open the way for the use of transfer learning in pre-clinical image analysis—a field in which this technology is yet quite unexplored.

- This study aims to determine the optimal number of training images necessary for fine-tuning RatLesNetV2 (Valverde et al., 2020). We utilize transfer learning with an extra emphasis on the impacts of data augmentation to identify the best-performing model for our particular dataset. We examine the complex dynamics of augmentation in conjunction with various training sample sizes using the RatLesNetV2 model as a pre-trained model. Our study goes beyond the straightforward use of transfer learning by examining how the performance of the model is impacted by augmentation with different training sample sizes.

1.6 Thesis outline

By giving a brief summary of earlier studies that were done in a related field, Chapter 2 reviews relevant literature. An extensive review of the literature will be included to pinpoint any existing gaps, emphasize the most important results, and place the present work in the perspective of the larger field of study.

The approaches used in this research will be explained in Chapter 3 in extensive detail. This will include a full breakdown of the segmentation designs used, including their core ideas and benefits. The transfer learning approach used in this research will also be thoroughly discussed, along with the justification for its choice and how it was used in the study.

A comprehensive look at the findings from the carried-out experiments will be covered in Chapter 4. In order to evaluate and draw valuable insights from the results, this analysis will require a comprehensive evaluation of the data using the relevant performance evaluation metrics.

A thorough discussion of the findings and their implications will be included in Chapter 5, where the findings are critically examined and evaluated in light of

the study's goals and assumptions. Prospective directions for further investigation will be looked at. The chapter will end with an overview that highlights the major conclusions and contributions of the research.

2 Related Works

The idea of using deep learning-based techniques continues to gain a lot of acceptance to enhance various domain of medical imaging related research. The intrinsic capacities of deep learning algorithms to efficiently manage noticeably higher amounts of data contribute to this approach's efficacy to predict by significantly decreasing both intra and inter-observer variability. Deep learning based techniques can handle large datasets with increased stability and dependability, reducing inconsistencies in interpretation made by different human expert (Kora et al., 2022). In (Kora et al., 2022), the study claims that, human expert's performance is not static but rather changes over time and is technically referred to as intra-operator variability. The human expert's expertise is improved through constant exposure to a growing quantity of images, which therefore improves the accuracy of diagnostic and treatment results. Transfer learning has become an acknowledged method in the field of medical image segmentation, earning an increasing amount of attention recently. Transfer learning has been widely used by studies in several investigation, taking advantage of its ability to improve the precision of segmentation. These evaluations provide insightful critiques of the many methodologies, databases, and field results that have been seen (Cheplygina et al., 2019; Karimi et al., 2021; Valverde et al., 2021).(Ghafoorian et al., 2017) evaluated the impact of transfer learning in the context of MRI-based brain white matter hyperintensity segmentation when the source and target regions had different acquisition techniques. According to the study's findings, transfer learning produced better results than traditional training.

By starting with a pre-trained CNN to find suitable samples for annotation and then gradually enhancing the (fine-tuned) CNN through ongoing fine-tuning, (Zhou et al., 2021) presented a novel method by integrating active learning and transfer learning (fine-tuning) into a single framework to significantly reduce the cost of annotation. The effectiveness and reliability of transfer learning using deep CNNs for classifying breast lesions in ultrasound (US) images are examined by (Byra et al., 2019). In comparison to conventional supervised learning approaches, the study (Van Opbroek et al., 2014) substantially decreased classification mistakes by using transfer learning techniques to enhance segmentation performance, particularly

when there is not enough of easily available representative training data. The study (Ghafoorian et al., 2017) examined domain adaptation in medical image analysis, focusing on the volume of data needed for modifying an existing network and retraining previously learned model parameters. The experiments done on segmentation challenge show that a domain-adapted CNN trained with only two classes surpasses a network built from scratch on the identical cases, attaining a Dice score of 0.63.

2.1 Volumetric Image Segmentation

Due to its excellent accuracy and efficiency, the UNet architecture (Ronneberger et al., 2015), originally created for medical imaging segmentation, has grown in prominence and is now useful for a variety of image segmentation applications. The UNet architecture consists of two parts: an encoder with down convolutions after a decoder with up convolutions coming and max-pooling layers for feature extraction. Using the previously extracted features, the model creates an output mask with dimensions that match the input mask. Medical imaging often produces volumetric data that depicts three-dimensional anatomical structures, hence it is necessary to have tools that can understand this multidimensional spatial data. Although there has been significant advancement with 2D UNet, it typically analyzes images slice-by-slice, essentially converting the volumetric data into a sequence of two-dimensional images. This method may possibly lose crucial 3D context and relationships between neighboring slices, which would negatively impact segmentation performance. The main benefit of 3D UNet (Çiçek et al., 2016) is its capacity to analyze volumetric data in its original format, which allows it to better represent the three-dimensional spatial environment. The 3D convolution process in 3D UNet keeps track of depth data, enabling the model to identify patterns across different slices and provide more reliable and precise segmentation results.

The 3D UNet offers more thorough and precise segmentation by acting directly in three-dimensional space, gathering spatial information and capturing volumetric structures with more accuracy. This development has greatly enhanced the segmentation of 3D medical images, providing new opportunities for accurate diagnosis, treatment planning, and interdisciplinary study. In (Kumar et al., 2022), the automated segmentation of the liver from a computer-assisted CT scan was done using the UNet model from MONAI in order to pinpoint the areas of interest, such as tumor locations and organ borders. In (Sivakumar and Ganeshkumar, 2017), a novel application of an Adaptive Neuro Fuzzy Inference System (ANFIS) classifier was made, with the approach focused on the automated identification and segmentation of brain stroke regions. This classifier and a Heuristic Histogram Equalization Technique (HHET) to the imaging data were used to enhance the clar-

ity and visibility of the interior brain regions. This method significantly improves the quality of the brain scans, which eventually helps to more precisely define the regions afflicted by stroke. Deep learning based approach in (Zhou et al., 2019), proposed a novel 2D network framework that effectively combines a DeconvNets (EDD)-Net and a Multi-Scale Convolutional Label Evaluation Net (MUSCLE Net) within an ensemble for lesion segmentation, eradicating the need for artificial feature design inherent in traditional machine learning and achieving better performance on a large clinical dataset. The Dimension Fusion UNet (D-UNet) method, which combines the strengths of both 2D and 3D convolutions during the encoding stage, was introduced in the study (Zhou et al., 2019) to address the challenges in the field of lesion segmentation and the approach achieves an improved segmentation performance compared to 2D networks while significantly reducing computational time when compared to 3D networks. Leveraging the power of transfer learning for 3D data representation, the study in (Kolarik et al., 2021) successfully managed to segment heavily unbalanced data without selective sampling, yielding more reliable results with less training data in a single modality, demonstrating the efficacy of this approach in practical medical applications for lesion segmentation.

The clinical process from diagnosis to follow-up is made easier by segmenting medical. The study in (Gillot et al., 2022) provided a revolutionary method that efficiently completes challenging segmentation in around five minutes, which is a major improvement over the typical manual effort of trained doctors, which takes seven hours. In order to achieve a broad clinical application, the approach integrates UNet Transformers (UNETR) from the MONAI framework. Training and testing were done on 618 de-identified Cone-Beam Computed Tomography (CBCT) volumetric images of the head that were obtained with various parameters from different centers. The MONAI framework was used in (Nisar et al., 2022) to preprocess, train, and validate the neural network that served as the basis for the investigation. Using a radial, forward-looking Foresight intracardiac echocardiography (ICE) ultrasound probe, inferior vena cava (IVC) images from an animal investigation were utilized in this study and the ground truth was established by manual segmentations, which was then verified by a skilled clinician. On the dataset, a UNet architecture was trained using the MONAI platform in order to segment vessels.

2.2 Clinical Image Segmentation

Due to its non-invasive nature and capacity to produce fine-grained, three-dimensional images of live tissue, clinical magnetic resonance imaging (MRI) has established itself as a cornerstone in the diagnosis and therapy of several diseases. The capture of images with numerous contrasts is made possible by this variability, giving a

thorough view of the interior architecture of the body. Automated segmentation is highly beneficial in the clinical field for assessing illness progression, especially in the context of brain disorders and lesions. For instance, segmentation helps calculate tumor size and location, which is important for figuring out how much radiation should be given to patients during radiotherapy. Automated segmentation has also been used to categorize a number of illnesses, such as neoplasms and neurocognitive disorders. Significant number of segmentation research have each examined one of these applications, which include anatomical (Gillot et al., 2022), lesion (Valverde et al., 2019), and tumor (Hatamizadeh et al., 2021a) segmentation.

The development of clinical techniques for image segmentation for diverse medical applications has been facilitated by a number of research. (Elliott et al., 1992) concentrated on interactive image segmentation for radiation treatment planning and developed a technique that speeds up the process by combining data from edge and region detectors. A model-based technique for segmenting three-dimensional images was created by (Mitchell et al., 2002), and its effectiveness was tested by segmenting echocardiographic temporal image sequences and volumetric cardiac MR images. By merging pattern classification and level sets, (Li et al., 2004) adapted image segmentation for clinical contexts and proposed an effective architecture that includes a pattern classifier, hierarchical, and connected level sets. STAPLE, an expectation-maximization technique for concurrent truth and performance level estimation, was presented by (Warfield et al., 2004). (Li et al., 2006) presented a generic automated technique for clinical image segmentation in their study, focusing on automatic clinical image segmentation employing pathological modeling, PCA, and SVM. All in all, these studies (Li et al., 2004; Warfield et al., 2004; Li et al., 2006)help to develop and apply advanced image segmentation techniques in clinical settings, paving the way for better diagnosis, treatment planning, and decision-making processes.

Effective segmentation techniques are beneficial for the wide range of imaging modalities used in clinical settings, including computed tomography, magnetic resonance imaging, and positron emission tomography. These methods are responsible for locating and identifying areas of interest, which opens the way for more precise diagnostics and individualized treatment planning (Litjens et al., 2017). Effective medical image segmentation is hampered by a number of issues, the segmentation problem is made more difficult by the considerable variability in patient anatomy, the presence of image noise and artifacts, the low contrast between diverse tissues, and the need to adapt to varied imaging modalities and procedures (Menze et al., 2014). These problems are made considerably more obvious in the case of 3D segmentation, where the extra dimension adds more complexity (Zuluaga et al., 2015). Using a cascaded architecture of two 3D patch-wise CNNs, this study (Valverde et al., 2017) proposes a automated method for the segmentation of White

Matter (WM) lesions in Multiple Sclerosis (MS) patient images. The architecture has been successful with a small clinical dataset of manually annotated MRI images, solving the pervasive problem of label scarcity. In this cascaded architecture, the first network detects possible lesion voxels, while the second network refines this output to reduce misclassified voxels.

2.3 Pre-clinical Image Segmentation

In preclinical settings, image segmentation is just as important since it enables thorough and in-depth anatomical and pathological investigation on experimental animals, which produces informative results that influence the course of clinical research (Hernandez et al., 2003). In preclinical settings, manual segmentation was initially the preferred technique, similar to clinical applications. Manual segmentation, however, proved to be an unfeasible approach due to the complexity and sheer amount of preclinical imaging data, as well as the need for high-throughput analysis (Hulsen et al., 2019). SegCaps, a novel convolutional-deconvolutional capsule network for object segmentation, is introduced in (LaLonde et al., 2021) and shows improved performance and efficiency on a variety of clinical and pre-clinical imaging datasets.

Rodent models are very important in the area of pre-clinical research, notably in the development of new drugs and therapies (Valverde et al., 2019). For these long-term investigations, non-invasive imaging techniques like magnetic resonance imaging (MRI) have become essential for tracking the development of the illness and the effectiveness of treatment (Valverde et al., 2020). The manual segmentation of brain lesions in mouse MRI images, which takes time and is prone to inaccuracy, has been one of the persistent problems in this field. This issue emphasizes the critical necessity for trustworthy automated techniques for segmenting rodent brain lesions. The study (Valverde et al., 2020) present RatLesNetV2, the first 3D Convolutional Neural Network particularly created for automated segmentation of rodent brain lesions in MRI data. This ground-breaking approach has been trained from beginning to finish, needs no preprocessing, and has been shown effective on a large dataset of 916 rat brain MRI images from 671 animals at nine distinct lesion stages. RatLesNetV2's relevance and potential influence in this area are further supported by the fact that this dataset has been used to research focal cerebral ischemia. This innovation represents a significant step forward and provides a fix for the urgent problem of manual lesion segmentation in mouse MRI data.

In order to reduce time for researchers and doctors, the amount of MRI data has increased significantly in recent years, prompting the creation of automated segmentation techniques. However, the lack of available techniques and poor segmentation performance for preclinical MRI emphasizes the need for further

developments in this field.

2.4 Transfer Learning in Medical Image Segmentation

For medical image segmentation problems, transfer learning is a useful technology, particularly when there is a lack of training data for CNNs. The source task and domain have an impact on the effectiveness of transfer learning for certain medical image segmentation tasks. The effectiveness of various source task and domain combinations for transfer learning-based medical segmentation is critically examined in (Zoetmulder et al., 2022). In the experiment, CNNs were trained on three distinct T1 brain MRI segmentation tasks after receiving pre-training on classification, segmentation, and self-supervised tasks in the context of natural image and T1 brain MRI domains. The findings show that even with fewer datasets, pre-training with tasks and domains similar to the intended tasks and domains led to better performance. In the context of medical image analysis, the study in (Karimi et al., 2021) clarifies the impacts of transfer learning on training duration and segmentation accuracy. The study found that although transfer learning reduces training time for the target task, the degree of gain in segmentation accuracy depends on the specifics of the task at hand, as well as the properties of the data that are available. Notably, significant accuracy improvements mostly appear when the segmentation job is more difficult and when the amount of the training sample is more constrained.

2.4.1 Domain Transfer Learning

This type of transfer learning, known as domain transfer learning , keeps the task constant while changing the domain. For instance, despite the shift in the visual environment, a model that was first trained to recognize things in outdoor settings may be fine-tuned to detect objects in inside scenarios. In the field of medical imaging, the idea of transfer learning—and more particularly, inductive transfer learning—has found several uses (Cheplygina et al., 2019). According to (Cheplygina et al., 2019), early transfer learning efforts in the field of medical imaging also focused on situations when the classification goal is constant but the domains are different. This issue frequently occurs as a result of the use of data from numerous hospitals or healthcare facilities. This method takes use of the idea of learning several tasks within the same domain together as opposed to independently, such as identifying various sorts of abnormalities in lung images. The first attempts at transfer learning in medical imaging focused on situations when the classification

task remained constant but the domains differed, such as when data is collected from several hospitals. Different scanners, demographics, and imaging settings might alter data distributions, making it necessary to train classifiers differently than simply using conventional techniques. By using shared, task-independent characteristics and more data, the collaborative learning process improves the robustness of representation. This comprises self-supervised learning, which works without labeled source data, as well as multi-task learning (MTL), which makes extensive use of labeled source data. In essence, the objective itself, such as identifying diseases or classifying patients, remains the same, but the data source, or domain, varies. These variances could be attributable to, among other things, variations in patient demographics, machine settings, or hospital-specific policies. The fundamental difficulty here is preserving the same classification task while modifying the learning model to function well in these various healthcare situations. (Conjeti et al., 2016) proposed a supervised domain adaptation (DA) framework that can adjust decision trees even when there are few labeled examples available when there is a distribution change between the source and target domains. The research introduces a unique domain adaptation method that employs a hierarchical transfer relaxation strategy that corrects errors. To address the distribution shift between the domains, this technique includes domain alignment, feature normalization, and leaf posterior reweighting. Several medical imaging studies use two main task integration techniques that are in line with the MTL paradigm. The first method makes use of various label spaces and disjointed training sets for every assignment. In their work, (Bi et al., 2008) provided a demonstration of this by classifying brain MR images of the brain into AD or cognitively normal (CN) and classifying nodules in chest CT images. Despite the unique label spaces of the tasks, this method improves the effective sample size. The second technique, which involves training numerous tasks on the same dataset to create a multi-label classification scenario, is shown in the works of (Zhang et al., 2012). This strategy improves the classifier's performance via regularization even while it does not increase the sample size. This is very useful when dealing with circumstances with small sample numbers.

2.4.2 Task Transfer Learning

In this type of transfer learning, the tasks are different while the source and destination domains are the same. For instance, a model that was first trained to classify various animal species may be improved to classify various plant kinds while using the information learned from the initial training.So, although the source and target tasks are different, the data distribution across the source and target domains is the same in inductive transfer learning, also referred to as task

transfer learning. In the area of computer vision, research on this phenomena has been widely conducted, and multiple studies have looked at how various activities are related to one another (Zamir et al., 2018; Standley et al., 2020). One such study (Zamir et al., 2018) examined the connections between various source and destination tasks in an effort to create a classification system that would indicate the degree of transferability across tasks. This was accomplished by carefully adjusting each target task depending on its unique source task. The experiment also assessed the improvement in performance that might be gained when features from models that had already been trained on several source tasks were combined to learn a target task.

Instance-transfer approaches, such as instance weighting to change the source distribution or feature space transformations to align the source and destination domains, are mostly used to address these distributional differences (Cheplygina et al., 2019). The study (Chen et al., 2019) investigated if learning a single target task may be accelerated by integrating features from models that have already been trained on a range of source tasks. The subsequent phase was to do a study to create practical approaches for identifying the most important source activities that may aid decision-making, particularly when computing resources are limited. According to (Achille et al., 2019), Through the discovery of connections between various categorization tasks, complementary study has shifted their attention away from the need of network fine-tuning.

The closeness in data distribution and feature properties between the source and target tasks has a significant impact on task-specific transfer learning success. According to (Weiss et al., 2016), when tasks are more similar in terms of data distribution and feature qualities, the transfer learning process is likely to perform better. Therefore, for the efficient application of task-specific transfer learning, a thorough knowledge of the underlying data distribution and feature similarities across tasks is essential.

2.4.3 Domain-Task Transfer Learning

Both the domain and the tasks shift in this situation for domain-task transfer learning. For instance, despite the differences between the tasks and the domain (natural language processing tasks vs image classification tasks), a model that was first trained on natural language processing tasks may be converted to handle the latter. The study (Moeskops et al., 2016) used domain transfer learning to train networks that could accurately interpret images from several imaging modalities and visualize various anatomical regions using a single CNN architecture. The segmentation performance obtained by including several tasks in the training approach was comparable to that of networks trained specifically for each task.

The CNN (Moeskops et al., 2016) showed that it could adjust to different tasks, suggesting the possibility of shared representations inside the network. A pretrained CNN (AlexNet) trained on non-medical tasks is fine-tuned for differentiating between benign and malignant breast tumors in this study's (Huynh et al., 2016) evaluation of task-domain transfer learning. The findings show that classifiers built using transfer learning outperform those built using analytically derived features, and the ensemble of CNN-extracted features based classifier and analytically extracted features based classifiers, pointing to the promise of transfer learning in computer-aided diagnosis (CADx) and radiomics. (Meng et al., 2017) emphasized the difficulties in utilizing diagnostic ultrasound to provide consistent and subjective diagnosis due to different influencing factors. The study suggests a task-domain transfer learning strategy for classifying liver fibrosis that makes use of VGGNet and a fully connected network (FCNet). The pretrained VGGNet model is adjusted, and deep features are extracted, allowing the FCNet to classify liver fibrosis with more accuracy than other methods. The potential for improving classification tasks in diagnostic ultrasound is shown by the combination of transfer learning and FCNet. Early work by (Schlegl et al., 2014) shows the new method of using non-medical images as raw data for training 2D networks, which serves as an early example of task-domain transfer learning. The categorization of tissue types in chest CT slices was done using this approach. Their creative approach combined information from real images, additional chest CT scans, and head CT scans, each of which represented a different source job. It's interesting that the study showed that natural images, a totally different source, not only equaled but sometimes even beat the outcomes obtained from utilizing just lung scans. However, using brain images was less successful. The differences between CT scans of the brain and lungs may be the cause of this discrepancy. The vast homogenous patches that are often seen in brain scans contrast with the more complex texture shown in lung scans, which may have affected how well the task-domain transfer learning process worked.

We have built a strong foundation for identifying the best methods relevant to our study aims through an extensive review of the available literature. The literature highlights a variety of procedures of using pre-trained models after different fine-tuning methods. Such revelations not only deepen our comprehension but also direct our research in the direction of customizing the best strategy for the particular task at hand. The combined information from this review act as a vital guide, directing the design of our experiments and how we interpret the findings as we go into the experimental phase of our study.

Table 2.1: *Diverse transfer learning types applied to various imaging modalities for distinct tasks.*

Reference	Modalities	Function	TL type
Kouw et al. (2017)	MRI	Lesion segmentation	Domain TL
Alex et al. (2017)	MRI	Brain lesion segmentation	Task TL
Wang et al. (2013)	MRI	Brain lesion segmentation	Domain TL
Van Opbroek et al. (2014)	MRI	Brain lesion segmentation	Domain TL
van Opbroek et al. (2015)	MRI	Brain lesion segmentation	Domain TL
Moeskops et al. (2016)	Brain MRI, Breast MRI, Cardiac CTA	3 different segmentation	Domain-Task TL
Ross et al. (2018)	Endoscopy	Medical instrument segmentation	Domain TL
Tajbakhsh et al. (2016)	CT, ultrasound, endoscopy	Classification, detection, segmentation	Domain-Task TL
Bi et al. (2008)	Lung CT, heart Ultrasound	Classification, segmentation	Task TL
Ablavsky et al. (2012)	Electron microscope image stacks	Segmentation	Domain TL
Mahmood et al. (2018)	Endoscopy	Depth estimation	Domain TL
Karimi et al. (2021)	MRI, CT	Different segmentation	Domain-Task TL

3 Methodology

The goal of this work was to thoroughly investigate several techniques for segmenting stroke lesions from mouse brain MRI data, with a particular emphasis on comparing transfer learning compared to training from scratch. In order to improve the robustness and generalizability of our models, we also included alternative designs such 3D Unet, Dynamic Unet, UNetR and experimented with other data augmentation techniques, such as the use of random affine transformations are described in this chapter. This chapter provides a thorough description of our strategy for enhancing the accuracy and effectiveness of segmentation by outlining our journey through these procedures, the optimization strategies we utilized, and the particular assessment criteria we used to evaluate their performance.

3.1 Dataset

In order to investigate and understand the effectiveness of transfer learning approaches, a dataset (An et al., 2022) consisting of T2-weighted MRI scans of mice was used in this study. This significant use of cutting-edge machine learning methods, particularly transfer learning, allowed for a deeper analysis of the dataset as well as more reliable segmentation outcomes for pre-clinical images. The dataset includes a total of 382 T2 MR images, where 332 animals underwent middle cerebral artery occlusion (MCAo) and 50 had a sham operation. According to An et al. (2022), the measurements from an MRI were made 24 hours after the stroke procedure. Using a 2D RARE sequence and acquisition settings that were adapted for stroke imaging, images were obtained on a 7 T MR scanner. Rapid Acquisition with Relaxation Enhancement, or 2D RARE, sequences were used in the imaging procedure of the study. The imaging procedure involves taking 32 successive slices with a 0.5 mm thickness each. The imaging's field of view (FOV) was limited to a 25.6 mm$\hat{}$2 region. The image matrix was originally recorded at a resolution of 256 $\times$ 196, and the final resolution was zero-filled to 256 $\times$ 256. The methodology used four averages to improve signal-to-noise ratio, which led to a scan duration that was around 6 minutes and 43 seconds overall. The lesions

were annotated semi-automatically and the annotations were exported in NIfTI format. This segmentation was carried out by an experienced researcher with over 20 years of expertise, who has previously verified tracings against histology. Each segmentation took around 5 minutes.

The Rat dataset which the pretrained model was based on has 916 T2-weighted MR brain scans from 671 adult male Wistar rats, weighing between 250 and 300 grams. The scans came from 12 different studies and were provided through the Discovery Services site of Charles River Laboratories. With scans collected at various times after occlusion, the dataset was varied. To cause localized cerebral ischemia in the right hemisphere of these rats' brains, a brief, 120-minute middle cerebral artery blockage was employed. Following the occlusion, MR data was collected over a period of time. Some research also used rats that had fake operations, which included the identical surgical techniques but no real blockage. All animal experiments complied with NIH guidelines for the care and use of laboratory animals and were approved by the National Animal Experiment Board in Finland. MR data was acquired using a multi-slice multi-echo sequence, with TR =2.5 s and 12 echo times ranging from 10-120 ms in 10ms increments. The images were obtained using a horizontal 7T magnet, and T2-weighted images were computed as the sum of all echoes. A total of 18 coronal slices of 1mm thickness were collected using a 30x30 mm2 field-of-view, yielding 256x256 imaging matrices of 117x117µm resolution. The T2-weighted images were not subjected to any pre-processing steps, such as inhomogeneity correction, artifact removal, registration, or skull stripping, but were zero-centered and variance-normalized to one. The provided lesion segmentations were carried out by trained technicians from Charles River. Additional independent manual segmentation conducted in (Valverde et al., 2020) from the first acquired study to estimate inter-rater variability. The average Dice coefficient between the two manual segmentations was 0.67 ± 0.12 for 2-hour lesions and 0.79 ± 0.08 for 24-hour lesions, with an overall average of 0.73 ± 0.12. We utilized the segmentation mask created manually during the study (Valverde et al., 2020) as the ground truth for study. The RatLesNetV2 model's pre-training phase was limited to images that had a ground truth mask, in other words, that had a lesion in them. Images from sham operations that showed no lesions were not included in the training procedure. There are several restrictions on the dataset that was used to build the RatLesNetV2 model in this experiment. The rat brain's magnetic resonance imaging (MR) dataset is not available publicly due to private rights and intellectual property concerns.

3.1.1 Data preparation and splits:

All the images used in this study included certain data alterations for all three training, validation, and testing set. An essential part of the preparation workflow includes reformatting the images in order to provide the best interoperability with the deep learning models we used throughout the study. In order to prioritize the channel dimension, the data dimensions have to be rearranged. This step has significance since usually deep learning models are built to anticipate data in a particular way. Segmentation masks were also transformed into a one-hot encoding format for our binary classification problem. The model's ability to produce more accurate predictions is improved by this change. To uniformly scale the intensity range across all images, normalization—a fundamental step in the pre-processing pipeline, was applied. By giving input features a consistent scale, eliminating differences in voxel intensity across data aids in improving model performance. Therefore, including these steps in the pre-processing pipeline ensures that the data is most adequately prepared for efficient learning and performance as well as being appropriately structured for the deep learning model employed in this work.

The dataset used for our study was divided into three separate categories: training, validation and testing. The original dataset included 382 images in total, of which 51 were removed from our analysis because those mice had sham operations and there were no lesions on those images. The remaining 331 images were divided into three subgroups. The allocation was as follows: 70% (231 images) went to the training set, 10% (34 images) was used as the validation set and 20% (66 images) to the test set. We used all 231 of the training set's images to train the MONAI segmentation models. In contrast, for the transfer learning experiments, we used fewer images that were automatically selected based on the distribution of lesion area sizes. It should be emphasized that, regardless of the size of the training set, the same test and validation sets were constantly used to monitor the training and assess all models.

3.1.2 Dataset selection for Transfer learning approach

For our transfer learning experiments, we created lesion-balanced training sets derived from our original training set split. Lesion sizes were calculated using the mask files to select the training data for TL models. Later, the size of the lesion in millimeters was determined by summing the mask area and multiplying that with the voxel size of the image. Lesion sizes were divided into bins according to equal Interval binning. the entire range of data is divided into a specified number of bins such that each bin covers an equal range of values. Here, the bins are defined by the minimum and maximum lesion size in the dataset, and the interval (or

width) of each bin is the total range of the data divided by the desired number of bins. This was achieved by creating a collection of bins with equal spacing using the linspace function. Figure 3.1 shows the distribution of lesion sizes of available training samples which are based on 70% of entire mice dataset. After calculating the lesion sizes, class labels were given to each size depending on where in the bins that size belonged. As we discussed we have equal Interval binning, one image was selected from each class or bin. A random image was chosen from the available images within each class to ensure a broad representation of lesion sizes. As we have 8 bins, if less there is no image to choose from a particular bin in a given scenario, further images were picked at random until 8 images were collected. This approach was used to ensure a sufficient sample size and prevent any possible bias toward certain lesion sizes. In the first step of our process, we define uniform bins or equal intervals that cover the whole range of lesion sizes seen in our dataset. We classify all lesions into one of these bins by establishing these bins based on smallest and highest lesion sizes.

A similar strategy was used to increase the size of the training set to determine the ideal model training size. In this scenario, the goal was to expand the training set while still making sure that it was appropriate for the task at hand. First, we grouped the images into eight bins of equal size (see Figure 3.1). Second, we sampled n images from each bin, where $n \in \{1, 2, 3, 4, 5\}$ resulting in five training set of 8, 16, 24, 32, and 40 MR images, respectively. In order to study the effect of dataset size on the effectiveness of the deep learning model, the number of images per class was gradually increased. This made it possible to evaluate the impact of the training sample size on the model's accuracy and generalization abilities. The trade-off between dataset size and model performance was clearly seen throughout this repeated process of growing the training set. In order to assure the model's ability to generalize effectively to new, untested data, the study set out to determine the bare minimum number of training images necessary to reach a desired level of performance. By taking a more complete approach, this study aimed to investigate the connection between model performance and the size of the training set, providing a thorough assessment of the lesion analysis capabilities of the transfer learning model.

3.2 Model architectures

3.2.1 3D UNet

The study used the MONAI framework to create a 3D UNet model, a well-known deep learning architecture for volumetric medical image segmentation.

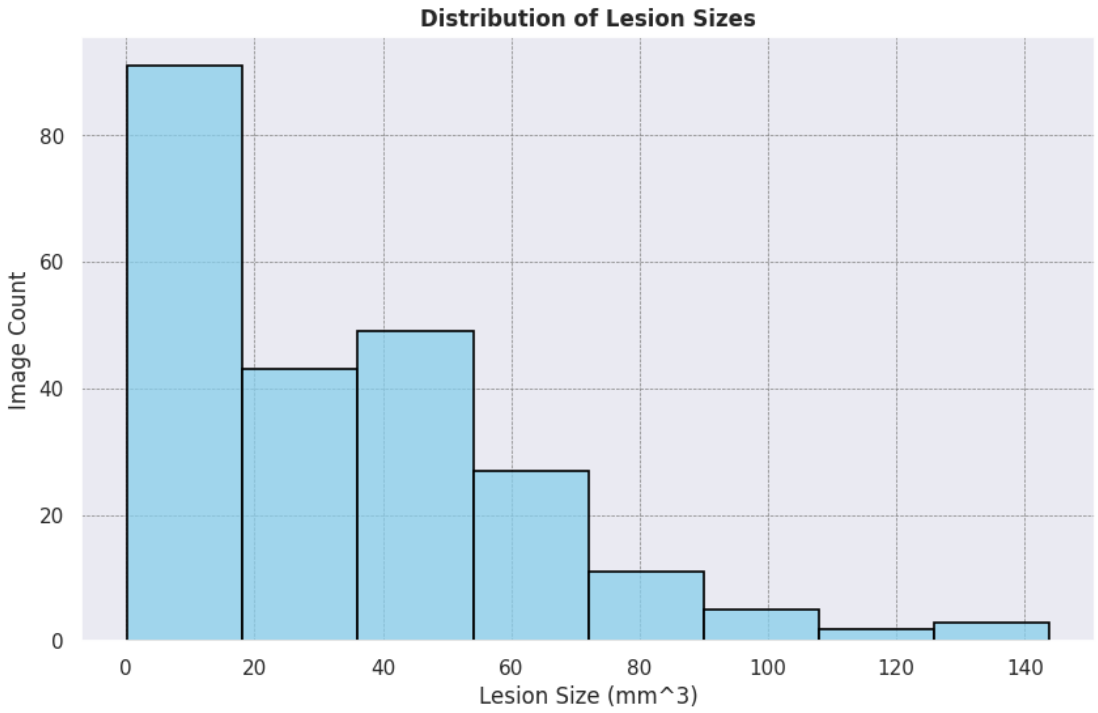

Figure 3.1: *Distribution of Lesion Sizes in the Training Sample Dataset (70% data of Mice Dataset).*

In order to deal with volumetric data, like the one used in this study, the 3D UNet model was set up with three spatial dimensions. To correspond with the two classes we have lesion and non-lesion, the model was configured to contain a single input channel and two output channels. The UNet model was constructed using an encoder and decoder, each made up of numerous layers, according to the network design. While the encoder path was made to gradually extract high-level abstract information from the input image, the decoder concentrated on utilizing these features to predict the segmentation map. This UNet architecture has five encoding layers and five decoding levels, with 16, 32, 64, 128, and 256 channels in each layer, respectively. In UNet architectures, it is usual practice to double the number of channels at each layer, enabling the model to find better feature at each depth level. The spatial dimensions of the output feature maps were halved at each level in the encoding path and then progressively recovered in the decoding path since the strides were set to 2 for all levels.

Two residual units (Xie et al., 2017) were added to each layer of the model to further improve it and speed up learning by enabling the model to build shortcut connections between levels. These connections aid in resolving the vanishing gradient issue that deep neural networks often experience, increasing the effectiveness

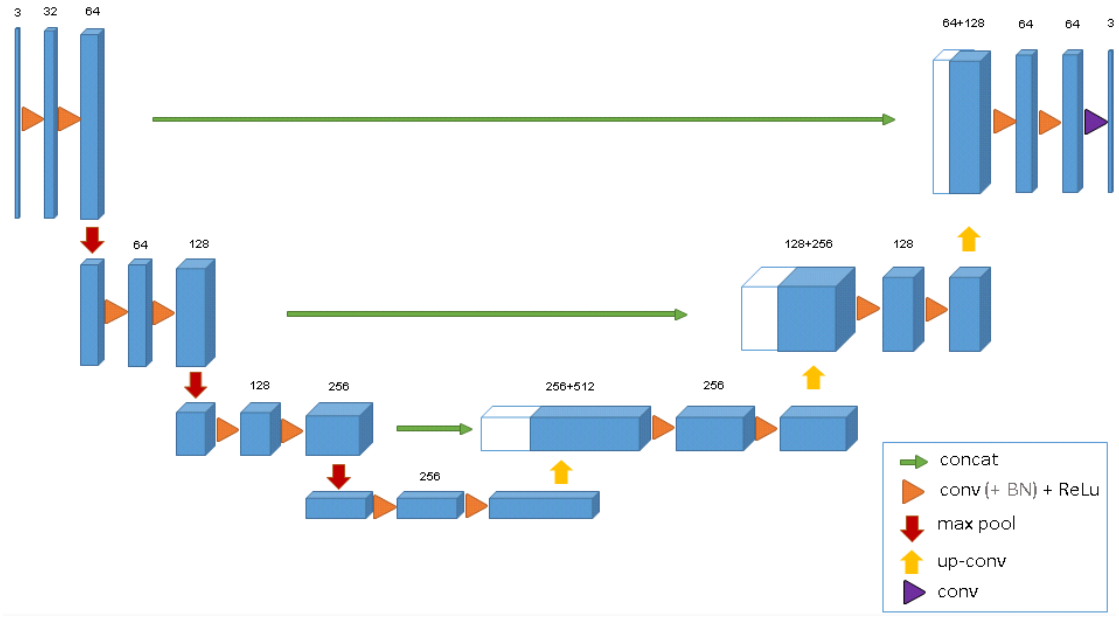

Figure 3.2: *Schematic representation of 3D UNet Model Architecture. (figure from (Çiçek et al., 2016), Creative Commons Attribution, CC BY)*

of training. To stabilize the learning process and shorten training time, batch normalization (Ioffe and Szegedy, 2015), a method that normalizes the activations of each layer, was used. This procedure lowers the likelihood of overfitting and improves generalization of the model.

Overall, the 3D UNet model's design was customized to meet the needs of this study specifically, making it well-suited for successfully completing the segmentation task. Figure 3.2 shows the model architecture of 3D-UNet from (Çiçek et al., 2016).

We defined significant differences between (Çiçek et al., 2016) 3D UNet model and the MONAI 3D UNet model used in our study while maintaining the fundamental UNet design, which consists of an encoding path, a decoding path, and lateral shortcut connections. The tuple (16, 32, 64, 128, 256) provided to the channels parameter in our study, defining the number of output channels for each layer, distinguishes our MONAI model from the (Çiçek et al., 2016) model's structure by adding five resolution stages as opposed to their model's four, although both version doubles the number of channels after each downsampling step. As with the original 3D UNet model, our MONAI technique normalizes each Rectified Linear Unit (ReLU) using batch normalization. In the MONAI 3D UNet, two residual units are used in each layer of the network. Compared to the original 3D UNet, which did not have residual units, this is an improvement. Particularly for deeper networks such this one, the introduction of residual units enhances model performance by

facilitating more efficient gradient propagation during training. This highlights the flexibility of the MONAI framework, which, despite the aforementioned differences, enables thorough modification to meet a variety of medical imaging requirements.

3.2.2 Dynamic UNet

In particular for biomedical image segmentation applications, the Dynamic UNet which was based on nnUnet (Isensee et al., 2021), and UNet are both well-known network structures for semantic segmentation from MONAI framework. However, there are several areas where they differ, particularly in terms of flexibility, deep supervision (Lee et al., 2015) and normalization. The DynUNet's inclusion of dynamic computation of the resolution levels, which enables improved flexibility, is one of the main improvements over the standard UNet. A predetermined, fixed set of resolution levels is used while operating the conventional UNet. In comparison, DynUNet offers a more adaptable method for managing a range of image sizes by dynamically adjusting the resolution levels depending on the size of the input image and the kernel. Deep supervision is used in our method, which is an additional choice offered by the DynUNet. Deep supervision is an approach in which the final layer of the network and its intermediate levels both contribute to the final prediction. This facilitates training-phase convergence more quickly and may enhance segmentation performance. The upsampling procedure used in the decoder is another significant distinction between the two models. Upsampling is accomplished in the UNet architecture using interpolation. The DynUNet model, on the other hand, employs a transposed convolution that enables the model to learn the upsampling filters during the training phase, possibly leading to more accurate feature map upsampling. In this procedure, a kernel or filter is used to create an output that has more spatial dimensions than the input by sliding over the input. The parameters of this filter, as opposed to simple interpolation, become known during training, enabling the model to adjust to the unique properties of the input. In our model architecture, batch normalization (Ioffe and Szegedy, 2015) is being employed. This is a technique that normalizes the features across all samples in a batch.

3.2.3 UNetR

UNetR (Hatamizadeh et al., 2021b) is a 3D model that includes Transformer's self-attention mechanism, which was first proposed by (Vaswani et al., 2017). Utilizing the strengths of both UNet and the Transformer, this significant combination combines the accurate localization skills of UNet with the excellent context comprehension of the Transformer. The UNetR model was created with a single input

channel and two output channels to correspond with the properties of our input data and the specifications of our binary segmentation task. UNetR's internal settings were altered, the Transformer component of UNetR was given dimensions of 768 and 3072 for the hidden layer and 3072 for the Multi-Layer Perceptron (MLP) layer. The trade-off between model complexity and computational feasibility guided the careful selection of these dimensions. Twelve distinct "heads" made up the self-attention mechanism, a crucial component of the Transformer design. This divide, as indicated by the "num_heads" option, enables a varied attention focus, allowing the model to concurrently collect several features of the data. A perception-based method was used for the positional embedding inside the Transformer. By using this technique, the model is better able to understand the spatial connections present in the volumetric data. Instance normalizing, which normalizes the characteristics of each unique instance, was the normalizing approach used. This method enhances the model's generalization skills and makes it more resilient to changes in the input distribution. The UNetR architecture, in conclusion, offers a complex and cutting-edge method for solving volumetric image segmentation issues. UNetR acts as a reliable and flexible solution in our segmentation task resources by combining the advantages of UNet and Transformer models and enabling subtle customization via numerous parameters. The decision to include residual blocks in the UNetR model for our investigation was driven by their ability to solve the vanishing gradient issue and facilitate the learning of identity functions, notably enhancing the performance of deeper layers in the architecture. Figure 3.3 shows the schematic representation of UNetR model architecture from (**?**).

3.2.4 RatLesNetV2

RatLesNetV2 is built on the idea of a UNet (Ronneberger et al., 2015), a type of convolutional neural network that has shown higher performance in biomedical image segmentation tasks. This model extends the fundamental UNet architecture by including residual connections (He et al., 2016) and making use of bottleneck layers, both of which provide special benefits in processing and comprehending 3D volumetric data. RatLesNetV2's architecture starts with a 3D convolutional layer, has a series of residual blocks and 3D max-pooling layers for the network's downsampling or encoding portion. Each residual block is made up of the following ReLU activation function, Batch Normalization 3d, Convolution 3d, ReLU, BatchNorm3d, and Conv3d, with an extra shortcut connection running from the block's input to its output, in order to prevent the model's performance from degrading with depth.

The architecture uses a succession of bottleneck layers after the encoding path, with a ResNet block and an upsampling step after each layer. The representation's

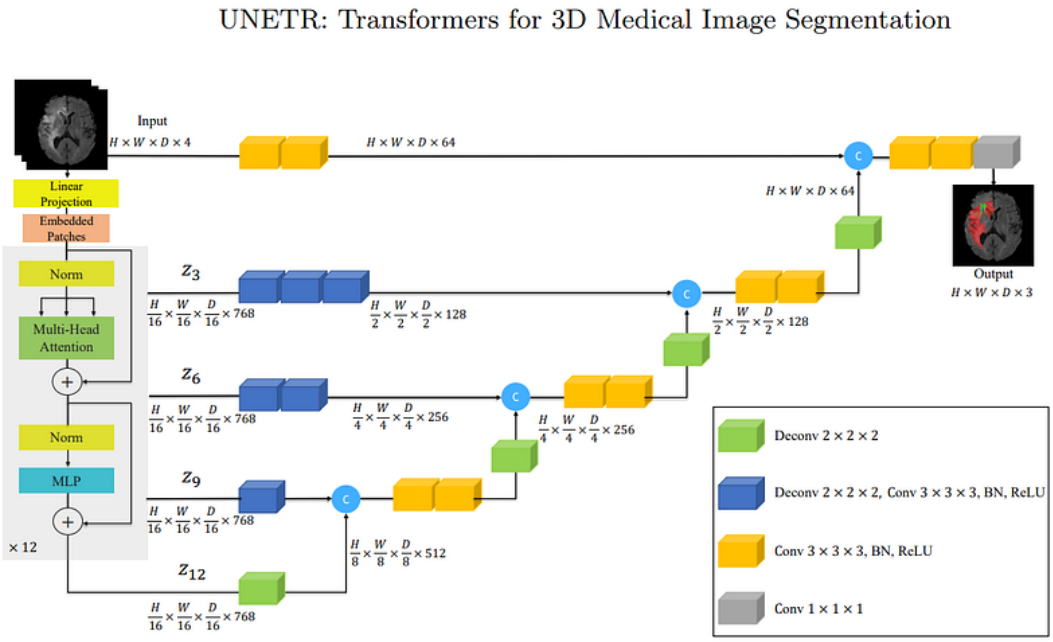

Figure 3.3: *Schematic representation of UNetR Model Architecture. Figure from (Hatamizadeh et al., 2021b), Creative Commons Attribution, CC BY*

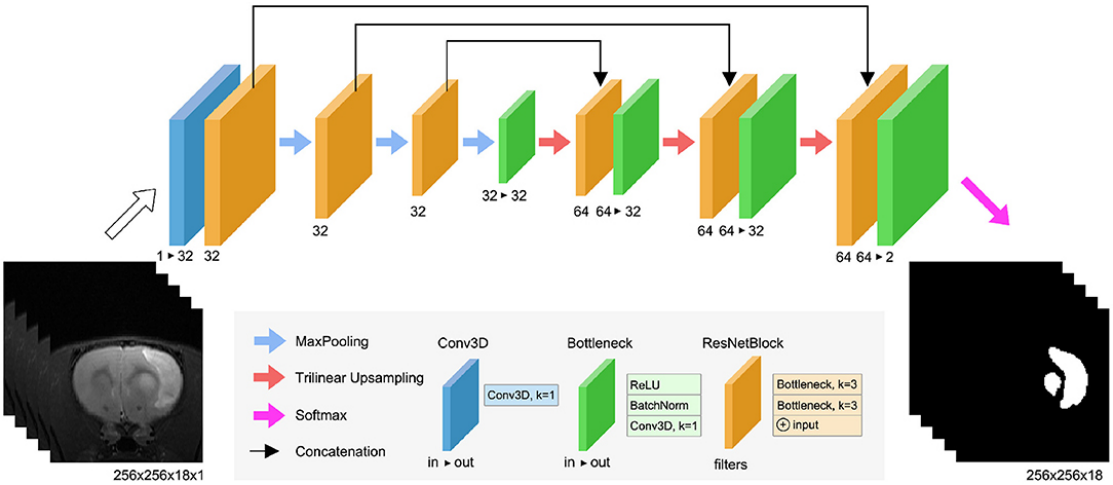

Figure 3.4: *Schematic representation of RatLesNetV2 Model Architecture.(Figure from (Valverde et al., 2020), Creative Commons Attribution License, CC BY.)*

dimensionality is decreased by the bottleneck layers, which also lowers the computational load and lowers the possibility of overfitting. The network operates at the

input data's original resolution which is a benefit of using upsampling procedure, that is essential for collecting features that are often lost during the network's encoding phase. As it takes into account the intensity values of nearby voxels to estimate the output, trilinear interpolation is a good option for 3D data and is what was utilized for upsampling in this architecture. Concatenating the output with the corresponding output from the encoding path comes after the upsampling procedure. This results in a skip connection that offers the decoder high-resolution information and aids the model in better localizing and detailing the output. This procedure is repeated along the network's upsampling or decoding chain. The output of the network is subsequently sent via a final bottleneck layer, which decreases the number of output channels to two, corresponding to the classes lesion and non-lesion. Figure 3.4 shows RatLesNetV2 architecture (Valverde et al., 2020). And in our segmentation task, the input to the model is 3D T2-weighted MRI images of size 256×256×32, where there will be 32 slices of 256×256 size.

3.2.5 Tranfer learning - Partial fine tuning

A large dataset of rat brain MR images were used to train the RatLesNetV2 model. This was a crucial step since it gave the model the ability to build advanced feature extraction skills for stroke lesion identification in the setting of neuroimaging.

For partial fine tuning based transfer learning approach, the RatLesNetV2 model's encoder was held constant, thus "freezing" the learnt weights. The generalization ability attained during the pre-trained session is preserved by this approach. The value of this comes from the generalizability of the learnt traits, which are often useful for tasks that are comparable to one another, such identifying lesions in the structural context of a different species' brain, in this example, the mouse brain. In parallel, the model's decoder part underwent fine-tuning using data from the mice brain. The goal of this step is to provide the decoder the ability to specialize for the intended task. This section of the model, which generates the output, becomes better at spotting and segmenting stroke lesions in the specific setting of mice brain imaging. As a result, the information that was learnt by exposure to rat brain images and stored into these weights is unaffected. The effectiveness of the transfer learning strategy in this study is supported by this combination of generalizable information from a related task (rat brain stroke lesion segmentation) and specialized learning for the particular task (mouse brain stroke lesion segmentation). So, we used domain transfer learning technique in this work to take into consideration the differences between the source and target domains.

We take advantage of transfer learning's benefits by using this two-step training process in this approach. On the one hand, we get advantage from the generic feature extraction capabilities of the bottom layers, which are pre-trained. The

higher layers, on the other hand, are tailored to our particular segmentation goal. As it builds upon previously learned features, this approach may result in better performance and quicker convergence than completely starting from scratch when training the model. The use of pre-trained model components minimizes the possibility of overfitting, which is another advantage of this method when we dealt with a small number of training samples or smaller datasets. By using a fine-tuning strategy, our technique takes use of the concepts of domain adaptability. We are preparing our model to adapt to a new biological domain while keeping the primary job, the 3D segmentation of lesions. The transfer learning method enables us to fine-tune the pre-trained model on a comparable but different collection of data, in this example, mice T2 MRI images. The pre-trained model was first trained on rat T2 MR images. We preserve the general interpretative capabilities such as edge detection and texture discernment by freezing the weights of the lower layers that have learnt essential characteristics from the rat MRI images. The network's higher layers are then trained using the mouse MRI data, which enables them to adapt and pick up on the nuances and particulars of this new domain. Utilizing the information gained from the prior domain (rat MRI images), the model is able to more effectively understand and interpret the details of the new domain (mice MRI images). As a consequence, our model was performing better since it basically combines extensive prior knowledge with a thorough comprehension of the particular task at hand. Thus, by concentrating the learning process on the distinctive features of the new domain, this strategy promotes the effective utilization of resources.

3.2.6 Tranfer learning - full fine tuning

In a different transfer learning technique, we used a thorough fine-tuning strategy, where the complete pre-trained RatLesNetV2 model was fine-tuned instead of limiting it to only certain parts. This meant that every layer in the network, from the input to the output, regardless of whether it was a part of the encoding or decoding portions of the network, was free to learn about and adapt to the new domain of T2 mice MRI images. While the pre-trained model already had a solid grasp of lesion segmentation in rat MRI images, it was thought that some layers might benefit from additional learning to better adapt to the nuances and particular features present in the mice MRI images. In essence, no layer was excluded, and each layer had the ability to modify its weights and biases in reaction to the new input.

This thorough approach to fine-tuning may provide a more adaptable transition to the new area. The larger number of trainable parameters does, however, mean that it often requires greater processing resources and a longer training time.

To avoid overfitting and guarantee steady convergence, regularization such as early stopping used and lower learning rate selection was also required. Overall, this approach gives the model the most flexibility to adapt to the new data, possibly producing a model with a high degree of specialization for the job of lesion segmentation in mice MRI images. As a result, the anticipated result is a model that is entirely customized to the features of the new domain and provided greater performance for the specific task at hand.

3.3 Optimization

3.3.1 Loss function

The computation of the loss function is a crucial step in the training of our model. It acts as an indicator to measure the difference between the predictions of our model and the actual ground truth. The optimizer iteratively improves the model's performance by using the estimated loss to modify the model's parameters via backpropagation (Werbos, 1990). The Dice Loss is often employed in medical imaging assignments because it is especially good at managing unbalanced data. The Dice loss was calculated using following formula:

$$\text{Dice loss}(A, B) = 1 - \frac{2 \cdot |A \cap B|}{(|A| + |B|) + \epsilon}, \tag{3.1}$$

where A is the set of voxels in predicted segmented lesion and B is the set of voxels in the ground truth. In order to avoid division by zero, the denominator of the equation has a minor positive constant called the ϵ added to it. The overlap between the expected and actual areas is measured by this function. It produces a number between 0 and 1, where 0 represents full agreement (perfect overlap) and 1 represents total disagreement (no overlap) between the projected and actual outputs. To explain it in more detail, our Dice Loss function calculates the Dice coefficient, which is basically the amount of overlap between two sets, and subtracts it from 1. To do this, it adds up the element-wise multiplications of the expected and observed outputs before scaling it by a factor of two. This total is divided by the total of the expected and actual outputs, resulting in a percentage of overlap between the two. By carefully handling the batch and channel dimensions, this configuration makes sure that the loss is calculated separately for each sample and each class. The denominator is added by a modest constant 10^{-5} to avoid division by zero. A reliable technique for evaluating prediction accuracy and directing the training of our deep learning model towards greater precision and performance is provided by Dice Loss function.

3.3.2 Optimizer

A fundamental deep learning method, backpropagation (Werbos, 1990)enables neural networks to learn and improve performance by repeatedly modifying their parameters. Calculating gradients, also known as the "slopes" of our loss function with respect to each model parameter, is the second step in the backpropagation algorithm. For each parameter, the gradients show the direction and rate of change of the loss. The gradients would indicate the slope's steepness and which direction it is sharpest if we were to visualize our loss function as a hill. We aim to identify the model parameters that give us the lowest loss conceivable, or, to put it another way, the bottom of the hill, in order to ultimately descend this metaphorical hill. This is when the optimizer comes into play, and we employed the Adam (Kingma and Ba, 2014) optimizer for our needs. In order to iteratively move the model toward a set of parameters that minimize the loss, the metaphorical hiker changes the parameters (steps) in the opposite direction of the gradients (steepest slope). When compared to other optimizers, Adam stands out because it modifies the learning rate for each parameter separately, adapting factors that need more or less aggressive updating. The Adam optimizer is started in our study with a certain learning rate and a weight decay, which adds a little penalty to big weights to assist avoid overfitting. We use a learning rate scheduler that adjusts the learning rate during training to further improve our optimization process.

Different starting learning rate and weight decay was used during training different models, we will discuss the specific values with reason in results section.

3.3.3 Regularization

Learning rate management is critical to the effectiveness of models in the context of machine learning, and especially deep learning. The model may not learn anything at all in certain circumstances if the learning rate is higher. In contrast, a low learning rate assures that the model learns more gradually, but at the risk of maybe getting stuck in a local minimum or requiring more computing power because of slow convergence. A technique used to optimize the learning process is called dynamic learning rate adjustment, sometimes referred to as learning rate scheduling. In order to achieve the advantages of beginning with a high learning rate (faster convergence at the start or avoiding local minima) and decreasing it later for fine-tuning, it modifies the learning rate throughout training by reducing it in accordance with a pre-defined schedule. For regularization in this study, we employed the Adam optimizer with a certain beginning learning rate and weight decay. We also provide a scheduler that controls learning rates, especially the StepLR scheduler from the PyTorch. Every few epochs (step size, set to 1 in

our example), this scheduler lowers the learning rate by a factor (gamma, 0.95 in our case, this implies that the learning rate decreases by 5% after each step). As training advances, the model is able to make smaller, more exact adjustments to its parameters, which often results in improved convergence to the minimum of the loss function.

By terminating the training process if the model's performance on a validation set does not increase for a certain period of consecutive epochs (referred to as "patience"), early stopping is a method used to avoid overfitting. One of the challenges in training deep neural networks is determining how many epochs are appropriate. This approach offers an automated way to choose the number of iterations. In this study, we track the training progress using a performance metric on a validation set. We terminate the training early if this metric does not improve for a certain number of subsequent epochs (the "patience"). For all subsequent predictions, the model's state at the epoch with the highest performance is preserved. The model's state at the epoch with the best performance is retained and used as the basis for all later predictions. Early stopping has two benefits that are complementary. In addition to preventing the model from overfitting to the training data by stopping before the model learns noise particular to the training data, it also conserves computing resources by eliminating unnecessary epochs. In light of this, early stopping offers a practical and efficient means of managing the deep learning training process.

3.3.4 Data Augmentation

We used the MONAI framework, as part of our technique to do data augmentation on our dataset. Without the requirement for new data acquisition, data augmentation is a strategy that enables us to increase the variety of our training data. This method is especially useful in the field of medical imaging, where logistical, economical, and privacy considerations may make it difficult to acquire data. Image and segmentation data were subjected to random affine transformations (in our instance, translations) using the RandAffined transformation. We set the range of potential translations at 10 voxels and the likelihood of applying the change at 0.5. This kind of data augmentation may improve the model's capacity to extrapolate to previously unobserved data. Preliminary experiments on other data augmentation strategies were unfavorable when we added rotation, and orientation modification as augmentation techniques.

3.4 Evaluation Metrics

3.4.1 Dice Coefficient

The MedPy package was utilized to calculate the Dice Coefficient (DC), and the segmentation method's performance is assessed mainly based on this coefficient. This statistic measures how well the segmentation method's output matches the ground truth. As a benchmark, manually segmented images represent the ground truth. DC is calculated to evaluate the segmentation's quality where the coefficient is a number between 0 and 1, with 0 signifying no overlap between the segmented result and the ground truth and 1 signifying perfect alignment.

The equation of the measure is as follows:

$$DC = \frac{2|A \cap B|}{|A| + |B|}, \tag{3.2}$$

where A represents the predicted segmented lesion's set of voxels and B represents the ground truth's set of voxels.

3.4.2 Hausdorff Distance

Even if the dice coefficient's findings are encouraging, there is yet potential to point out Hausdorff Distance to get a more robust visualization of performance based on some situation. For instance, when dealing with small structures or situations where false positives and false negatives have varying clinical significance, DC evaluation is not enough to represent the performance of the segmentation approach. Additionally, the spatial distribution of errors is not taken into consideration by this measure of DC. The Hausdorff distance (Rote, 1991) between two point sets is the maximum distance between a point in one set and the nearest point in the other set. However, since it takes into account the maximum distance, it is vulnerable to outliers. The 95th percentile Hausdorff Distance (HD95) could be used in its place to make it more resistant to outliers. Only the distance for which 95% of all distances are shorter is taken into account.

The HD95 is defined as:

$$d(A, B) = \max\left\{\max_{a \in \partial A} \min_{b \in \partial B} |b - a|, \max_{b \in \partial B} \min_{a \in \partial A} |a - b|\right\}, \tag{3.3}$$

A and B here represent the segmentation masks, and for appropriate boundary voxels ∂A and ∂B is used. The Hausdorff distance measures the segmentation error with the largest distance between the surface of the ground truth and the closest voxel in the prediction. Calculating the Hausdorff distance is crucial for studies involving the segmentation of brain lesions because higher errors in lesion borders can have a more significant effect. The HD95 calculation considers only

the distance for which 95% of all distances are smaller, making it a robust metric as it is less sensitive to the presence of outliers compared to the standard Hausdorff Distance. The calculated HD95 values are given in millimeters.

In our study, we concentrated on using the voxel spacing, which in the mice dataset is $0.1mm \times 0.1mm \times 0.5mm$, as a crucial component of calculating the HD95. The first two dimensions of the supplied voxel spacing values relate to the spatial resolution in the transverse plane, while the third value represents the spacing along the longitudinal axis. These values indicate the physical distance between voxels in the image. The measure was anchored in the exact physical dimensions of the data by including the voxel spacing in the HD95 computation, which improved its alignment with the underlying anatomical structures. The inclusion of spatial resolution in the HD95 calculation emphasizes the dedication to rigorous methodology and provides a unique insight into the similarities and differences between segmented objects and their related ground-truth labels.

3.5 Implementation

The learning and model optimization phases of our deep learning study are concentrated on the training procedure. To ensure the efficient training and assessment of the model, this procedure includes numerous steps that have a strong connection with one another. Utilizing numerous libraries that provide functionality essential for various activities in the study is the first step in the method. These consist of carrying out deep learning tasks, handling neuroimaging file formats, managing complex mathematical operations on multi-dimensional arrays and matrices, as well as managing a wide range of image processing tasks as required for the study.

Numpy adds support for large, multi-dimensional arrays and matrices as well as a substantial number of sophisticated mathematical operations that may be performed on these arrays. Numpy (Van Der Walt et al., 2011) is used to manipulate arrays when computing measurements and to save metrics throughout the training process.

os: The Python os library offers ways to use operating system-dependent features, including reading environment variables or navigating via file paths. When storing and loading parameters for the model and training metrics, the major usage of os in our script is to manage file paths.

nibabel:NIfTI and other popular medical and neuroimaging file formats could be read and write using the nibabel (Zhang et al., 2019). It enables smooth n-dimensional array opening and writing into disk files for use as medical imaging. It is used in our study to create NIfTI format files from the input data, the corresponding labels, and the anticipated output.

PyTorch: PyTorch (Paszke et al., 2019) is an effective open-source package that

offers a versatile deep learning development environment. It offers multidimensional arrays, or tensors, that are comparable to numpy's ndarray but that can be used for computation on a GPU. In our study, torch was mostly utilized in our implementation for model construction, data management, and the training procedure.

For executing the training phase of the model, the torch.optim submodule is required. It offers a selection of optimization techniques required for fine-tuning neural network parameters. In the context of our approach, an optimizer is developed by using this module, and it is then applied to the training loop to modify the model's parameters. These adjustments are guided by the computed gradients, driving the model towards improved performance.

In the meanwhile, the model's learning rate is significantly controlled by the torch.optim.lr_scheduler submodule. It gives users the option to dynamically change the learning rate in response to the advancement of epochs or the results of validation. We use the StepLR scheduler from this submodule in our case. After a certain number of epochs, the StepLR scheduler intentionally lowers the learning rate for each parameter group by a predetermined factor, or "gamma."

medpy.metric: The MedPy (Maier, 2015) library includes a number of metrics that are used in the study of medical images. The Dice coefficient, a spatial overlap index that measures the similarity between the anticipated and real segmentation, is calculated using the imported dc function. The degree of mismatch between the anticipated and actual segmentation boundaries is measured by the hd and hd95 functions, which compute the Hausdorff distance and 95th percentile Hausdorff distance, respectively.

MONAI: The Medical Open Network for AI (MONAI) (Cardoso et al., 2022)is an open-source Python library with a focus on creating, training, and analyzing deep learning models for medical imaging. Numerous classes for data transformation are available in the MONAI.transforms module, preparing data for network training. It facilitates the loading, converting, and pre-processing of input data that are essential steps in any pipeline for dealing with deep learning.

MONAI.data, this MONAI module offers features for building datasets and dataloaders that work with PyTorch. In our study, the classes DataLoader, Dataset, and CacheDataset are used. During the training phase, these classes provide an effective approach to iterate over the data.

The architecture for multiple neural networks is included in the MONAI module MONAI.networks.nets. The UNet architecture is loaded from this module into our script to create models. Several layers, norms, and activation functions are available in the MONAI.networks.layers module for use in neural network building. It is used to set up the UNet model's normalization procedure in our approach. MONAI.utils, this module offers utility methods that make it easier to manage

temporary folders, handle files, and establish determinism. In our study, we used the set_determinism function to regulate the program's ability to be replicated.

4 Results and Analysis

This chapter provides details of the experimental findings made possible by using segmentation methods experimented throughout this study on the specified dataset.

4.1 Experimental Settings

The computations were performed in an environment supported by 2.30GHz Intel Xeon CPUs. The arrangement provided enough memory for processing and executing the experimental workloads since the total memory (RAM) was roughly 26.7 GB. This was done to guarantee adequate memory availability for the computational operations. Additionally, the NVIDIA Tesla T4 GPU was used, which had a memory capacity of around 15.3 GB.

The MONAI segmentation models' training processes were constructed using a method including 100 epochs and a batch size of 1. The MONAI segmentation models were trained for 23,100 iterations in total, given the quantity of training samples (231). This number shows the total number of forward and backward passes made throughout the training phase. It is important to take this into account when analyzing training effectiveness and computing resource allocation. We chose to use only 20% or 4620 iterations, of the iterations needed for the MONAI models since our goal was to specifically shorten the training period. According to projections, this strategy would cut the amount of time needed for training into approximately 20% as the model complexity differs. In light of this, we systematically changed the number of epochs based on the size of the training set used for TL approaches. The scale used was as follows: 578 epochs for 8 training images, 289 for 16 images, 193 for 24 images, 145 for 32 images, and 116 for 40 images. We made sure the models were trained selectively altering the number of iterations. This strategic allocation provides a cutting-edge method for model training and validation within this intricate area and reflects a careful evaluation of the difficulties involved in medical image segmentation.

A learning rate of 10^{-4} and a weight decay parameter of 10^{-5} were used during the fine-tuning of the transfer learning models. The MONAI models, on the other

hand, were trained from scratch with a learning rate of 10^{-3} and a weight decay of 10^{-3}. A crucial component of achieving precise adaptation to a new task is transfer learning's lower learning rate of 10^{-4}. The pre-trained networks upon which these models are based already have their weights optimized for tasks relevant to them, thus only small modifications are required to adapt the model to the particular characteristics of the data that comes into it. As a result, the useful, previously learned features are preserved, and the risk of overfitting or catastrophic forgetting is reduced. A lower learning rate enables more regulated and subtle updates. A little regularization impact is produced by the smaller weight decay parameter, 10^{-5}. Large weights are penalized by weight decay, which helps avoid overfitting. A lesser weight decay guarantees that the pre-learned weights are not too constricted, preserving the pre-trained model's knowledge transfer. In contrast, a greater learning rate of 10^{-3} promotes faster convergence to an ideal solution for the MONAI models that were trained from scratch. Since these models lack the advantage of pre-trained weights, the lower learning rate enables more dynamic loss landscape navigation, which facilitates escape local minima. For the MONAI models, the equivalent weight decay of 10^{-4} represents a higher regularization impact. Stronger weight constraints may help avoid overfitting since these models are developing from scratch, which will help them generalize to new data.

4.2 Training from scratch

Three different MONAI segmentation models were used in our study under both data augmented and without data augmentation training conditions: 3D UNet, DynUnet, and UnetR. DC and HD95 were the two main metrics used to assess the performance of the models.

Notably, when data augmentation was used, the 3D UNet model performed better than the other models across all categories. The model's Dice Coefficient of 0.88 with standard deviation(SD) 0.07, which shows a significant overlap between predicted and manually annotated segmentation and supports the model's strong segmentation ability was achieved. Additionally, the HD95 value of this model was 0.39 with SD 0.25 mm with augmentation, indicating a good level of consistency between the predicted and real bounds. This further confirmed the model's high levels of performance. The values are given including the standard deviation. UNetR and DynUnet both produced acceptable results, but when trained with augmented data, their performance lagged below that of the 3D UNet. With augmentation, the UNetR model obtained a Dice Coefficient of 0.80 ± 0.14 and an HD95 of 1.62 ± 1.55, whereas the DynUnet recorded a Dice Coefficient of 0.77 ± 0.15 and an HD95 of 1.80 ± 1.54 under the identical circumstances. The designs of UNetR and DynUnet may not have been strong enough to tackle the difficulty of

Table 4.1: *Table depicting the results of different MONAI models (Trained using 231 images)with and without augmentation.*

Model	Without Augmentation		With Augmentation	
	Dice Coefficient	HD95	Dice Coefficient	HD95
3D UNet	0.87 ± 0.11	0.39 ± 0.55	0.88 ± 0.07	0.39 ± 0.25
DynUnet	0.76 ± 0.17	1.87 ± 1.23	0.77 ± 0.15	1.80 ± 1.54
UNetR	0.79 ± 0.16	1.73 ± 1.35	0.80 ± 0.14	1.62 ± 1.55

this particular segmentation operation.

The results of the study show that for volumetric segmentation tasks, the 3D UNet model combined with data augmentation performs much better. Table 4.1 provides a thorough side-by-side comparison of all tested models' performances, both with and without data augmentation. This layout offers an intuitive view of the relative strengths of different models and how data augmentation has influenced their outcomes. The predictions of the 3D UNet model with data augmentation are shown visually in Figure 4.1. A yellow outline shows the lesion's manual annotations, and a red contour shows the region that the model predicts will have the lesion. The closeness of the two contours highlights the advantages of using the 3D UNet and data augmentation in these challenging segmentation tasks and demonstrates to the model's strong capacity to accurately identify the boundaries of lesions.

Understanding the model's learning process requires a thorough analysis of its performance during training. With a focus on the Dice coefficient and Dice loss, Figure 4.2 provides meaningful representations of the training and validation progress over epochs. The number of training epochs is shown by the x-axis in this graph, while the Dice coefficient and Dice loss values are represented by the y-axis. The model is improving at appropriately segmenting the lesion region, as seen by the training curves' rising Dice coefficient and decreasing Dice loss over time. In addition, the validation curves provide information on how well the model generalizes to new data.

4.3 Transfer learning on RatLesNetV2

We used two methods to explore transfer learning: fine-tuning the RatLesNetV2 model after the encoder section which we can refer as partial fine-tuning, and fine-tuning the full model. We tested several training sample sizes, including 8, 16, 24, 32, and 40 images, for each of these approaches. Both data augmentation and

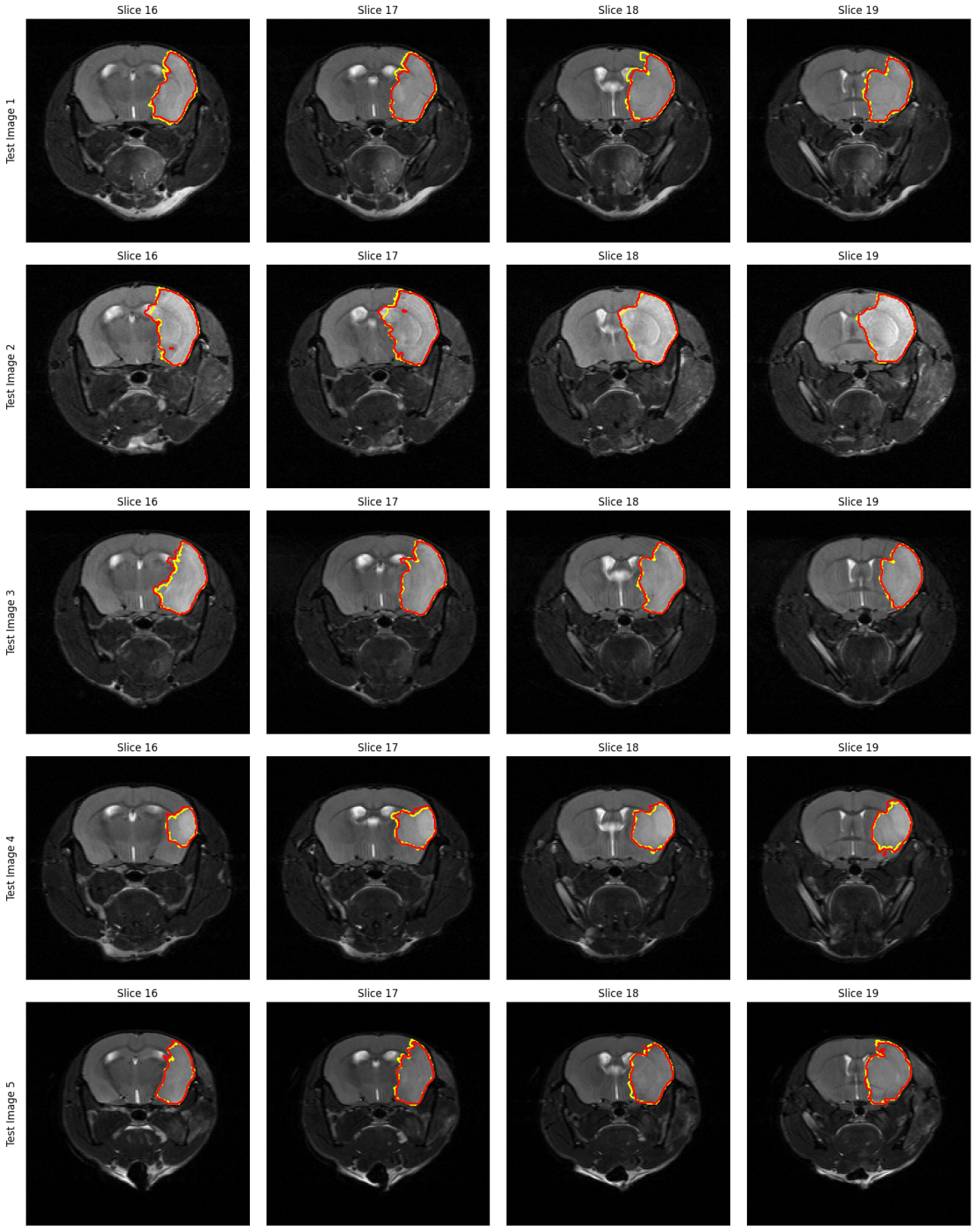

Figure 4.1: *Comparative Segmentation Outcomes of the 3D UNet with augmentation.*

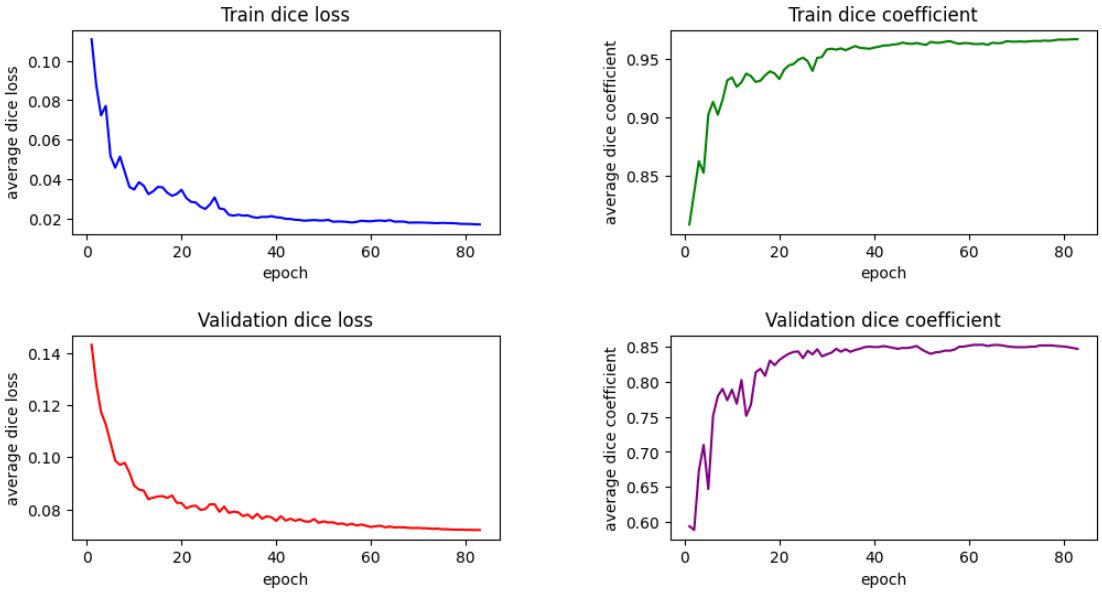

Figure 4.2: *Epoch-wise Progress Curve for MONAI UNet Model without Data Augmentation.*

no data augmentation were used to train models on each set of training images.

4.3.1 Fine-tuning the decoder

In this work, we investigated the performance of models that had been trained and partially fine tuned using various training sample sizes. The model trained on a collection of 40 images without augmentation outperformed the others, as demonstrated in Table 4.2. This model attained a high Dice coefficient of 0.85 with SD 0.10, which indicates significant overlap between predicted and real lesions, and an HD95 of 0.71, which indicates strong concordance between model-generated and actual lesion borders.

Conversely, the model trained on an augmented dataset of 16 and 32 images showcased its remarkable ability to deliver satisfactory results despite the limited training sample size if we compare the value range with other trained models. While it did not surpass the model trained on 40 non-augmented images, its performance highlights the benefits of data augmentation, especially when dealing with smaller datasets. Data augmentation broadens the variability in training data, which in turn helps to reduce overfitting and boosts the robustness of the model. Upon close examination of the SD values along with the average values, it becomes apparent that the models trained on 16, 24, and 32 training samples showcase comparable performance ranges. However, the model trained on a dataset of 40 images exhibits

Table 4.2: *Comparative Performance Metrics of Partial Model Fine-Tuning with Different Training Sample Sizes.*

Training sample size	Without Augmentation		With Augmentation	
	Dice Coefficient	HD95	Dice Coefficient	HD95
8 images	0.73 ±0.21	2.03 ±2.34	0.76 ±0.16	0.81 ±0.47
16 images	0.80 ±0.15	1.01 ±1.29	0.80 ± 0.12	0.92 ± 1.23
24 images	0.81 ±0.14	0.79 ±1.03	0.81 ± 0.15	0.88 ± 1.35
32 images	0.83 ±0.14	0.79 ±1.21	0.82 ± 0.10	0.80 ± 0.42
40 images	0.85 ±0.10	0.71 ±1.18	0.83 ±0.17	0.84 ±1.12

superior performance metrics, thus underlining its noteworthy efficacy.

Data augmentation was essential in limited situation (8 images) for increasing the variety of the training dataset, minimizing overfitting, and resulting in a more reliable model in terms of performance indexes. The segmentation results for the lesion regions in the models with the best performance after partial fine-tuning are shown in detail in Figure 4.3. These graphic illustrations highlight how effective selected fine-tuning is in improving the model's performance. In addition, Figure 4.4 provides a thorough understanding of the epoch-wise evolution of Dice loss and Dice coefficient, illuminating the complex interaction between loss reduction and accuracy maximization throughout training epochs. Together, these results highlight the possibility of a carefully planned mix of training set size, approach modification, and data augmentation in producing the best segmentation outcomes.

4.3.2 Fine-tuning all the parameters

A thorough inspection of Table 4.3 reveals a compelling narrative regarding the efficacy of full fine-tuning. Notably, the model that stood out from the rest was the one trained on a selection of 40 images without data augmentation. This model reached a Dice coefficient of 0.82 and an HD95 of 0.66, marking it as the front-runner in our examination. This accomplishment signifies the model's proficiency in leveraging the intrinsic details and spatial consistencies present in the training dataset, thereby enhancing its predictive capability without the assistance of data augmentation. When restricted to a smaller training dataset, particularly those with 8 and 16 images, the model trained with augmented data shown a respectable degree of resilience, even if it couldn't outperform its non-augmented version in the setting of 40 images. Having a close look at HD95 average and corresponding SD we can see the notable improvement when dataset is smaller. Upon scrutinizing both the

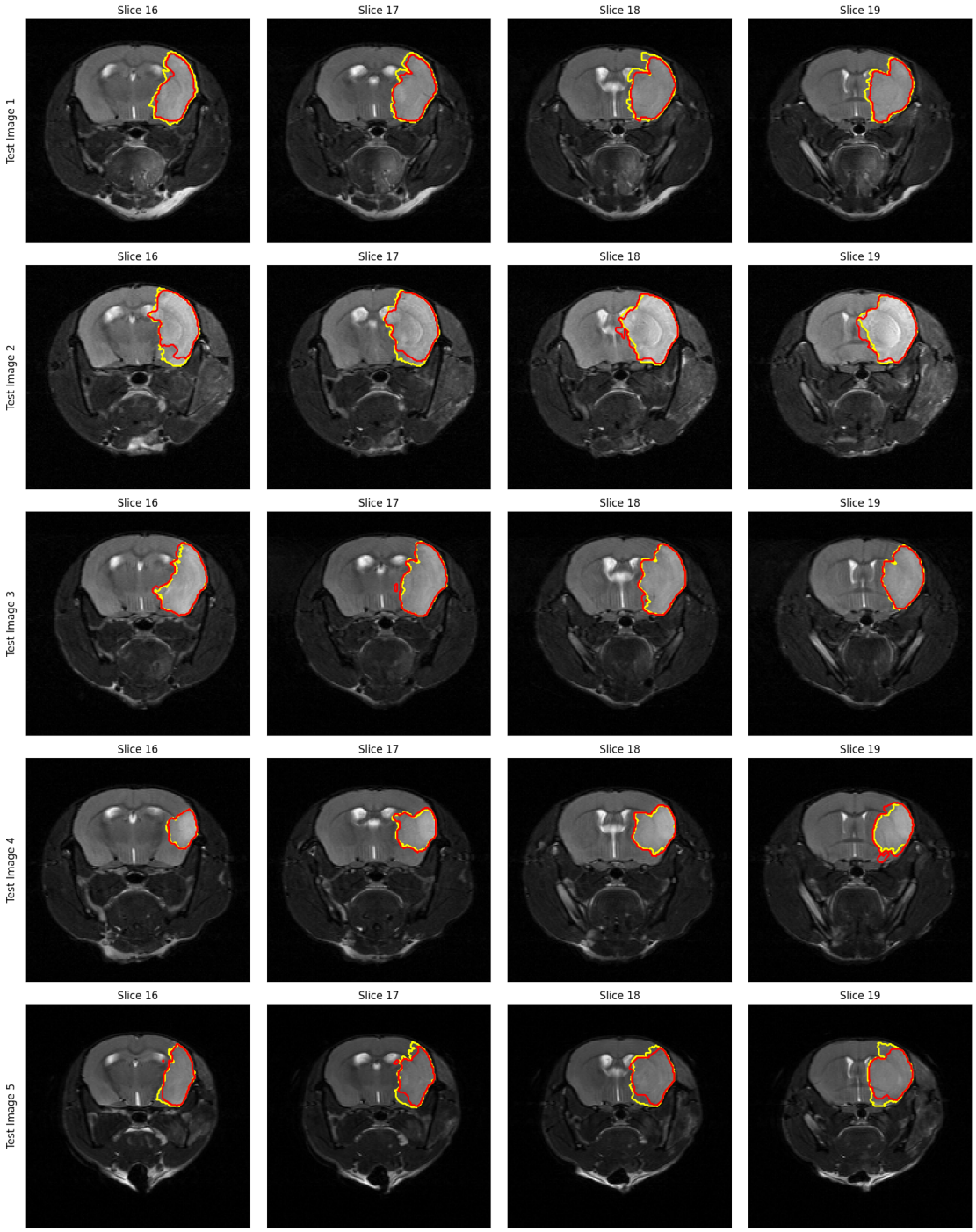

Figure 4.3: *Visual Representation of Segmentation Outcomes from Partial Fine-Tuning with 40 training samples.*

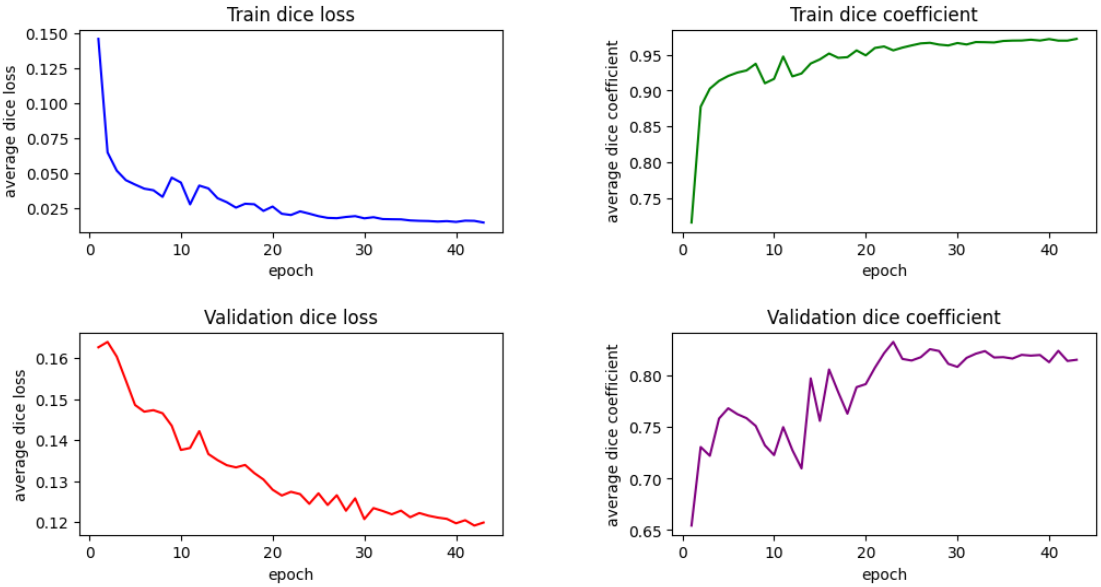

Figure 4.4: *Training and Validation Progress of the model which had 40 Training Samples without Augmentation for Partial Fine-Tuning Transfer Learning.*

average values and the associated standard deviations of HD95, it becomes evident that there's a significant performance enhancement, particularly when dealing with smaller datasets (8 and 16 images). This observation underscores the robustness and efficacy of our models when faced with limited training data, reinforcing the critical role of efficient training strategies in overcoming the challenges posed by data scarcity. The visual of lesion area segmentation for the model attaining the highest outcomes with full fine-tuning is shown in Figure 4.5, highlighting the model's skill in extracting and identifying complicated patterns. The epoch-wise trajectory of Dice loss and Dice coefficient is also shown in Figure 4.5, shedding insight on the complex interrelationship between these two measures as the model develops and improves its predictive power.

Even though a larger set with 231 training samples was available, our tests produced impressive results using just 40 training images, which represented a small portion of the whole dataset. This is an important finding that supports the efficacy of transfer learning strategies even in situations with small training data. This effectiveness needs attention because it raises the possibility that good results might be obtained without using the complete dataset, possibly lowering computing costs and training time. Using 40 training images produced an excellent performance when we concentrate on fine-tuning after the encoder stage, this clearly shows that a well-chosen (looking into the distribution of lesion) small dataset, not necessarily a huge one, improves even partial fine-tuning of the model.

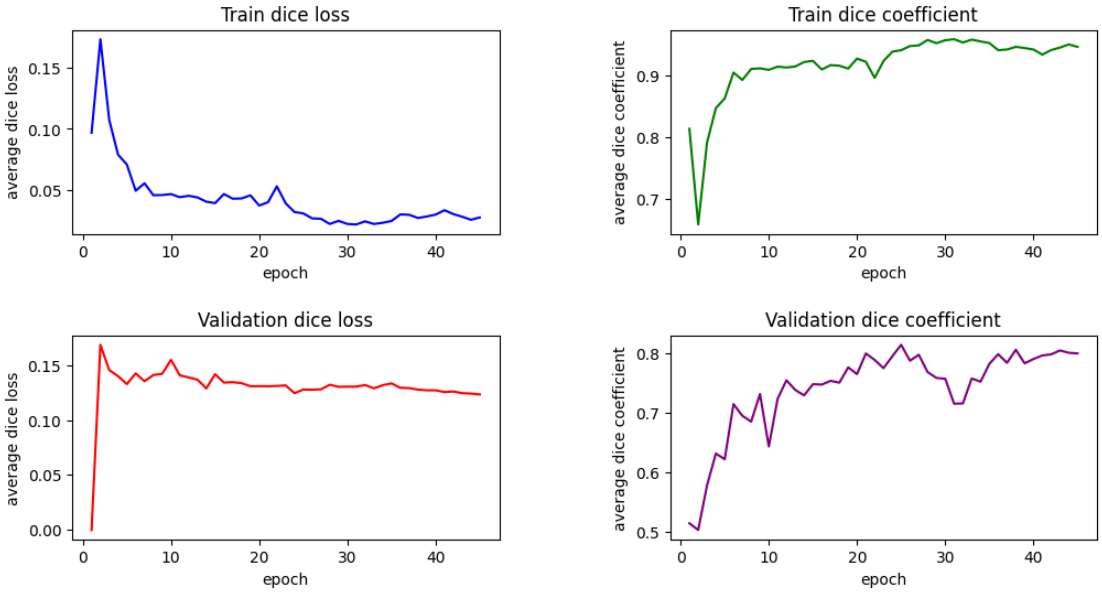

Figure 4.5: *Training and Validation Progress of the model which had 40 Training Samples without Augmentation for full Fine-Tuning Transfer Learning.*

Table 4.3: *Comparative Performance Metrics of Full Model Fine-Tuning with Different Training Sample Sizes.*

	Without Augmentation		With Augmentation	
Training sample size	Dice Coefficient	HD95	Dice Coefficient	HD95
8 images	0.73 ± 0.23	2.05 ± 2.62	0.77 ± 0.16	0.79 ± 0.52
16 images	0.76 ± 0.18	0.87 ± 1.14	0.77 ± 0.16	0.86 ± 0.59
24 images	0.81 ± 0.16	1.01 ± 1.54	0.81 ± 0.15	1.04 ± 1.51
32 images	0.81 ± 0.17	0.80 ± 1.13	0.81 ± 0.14	0.76 ± 1.11
40 images	0.82 ± 0.13	0.66 ± 1.03	0.81 ± 0.15	0.81 ± 1.14

4.4 Analysis and Findings

Performance Efficiency with Data Size: The fine-tuned model that stood out, which was trained on a selection of 40 images without data augmentation, reached a Dice coefficient of 0.85 ± 0.10 and an HD95 of 0.71 ± 1.18. This performance was achieved using only around 17.5 percent of the full training dataset size used for training models from scratch, which comprised 231 images. Despite this large discrepancy in dataset sizes, the drop in the Dice coefficient performance was only

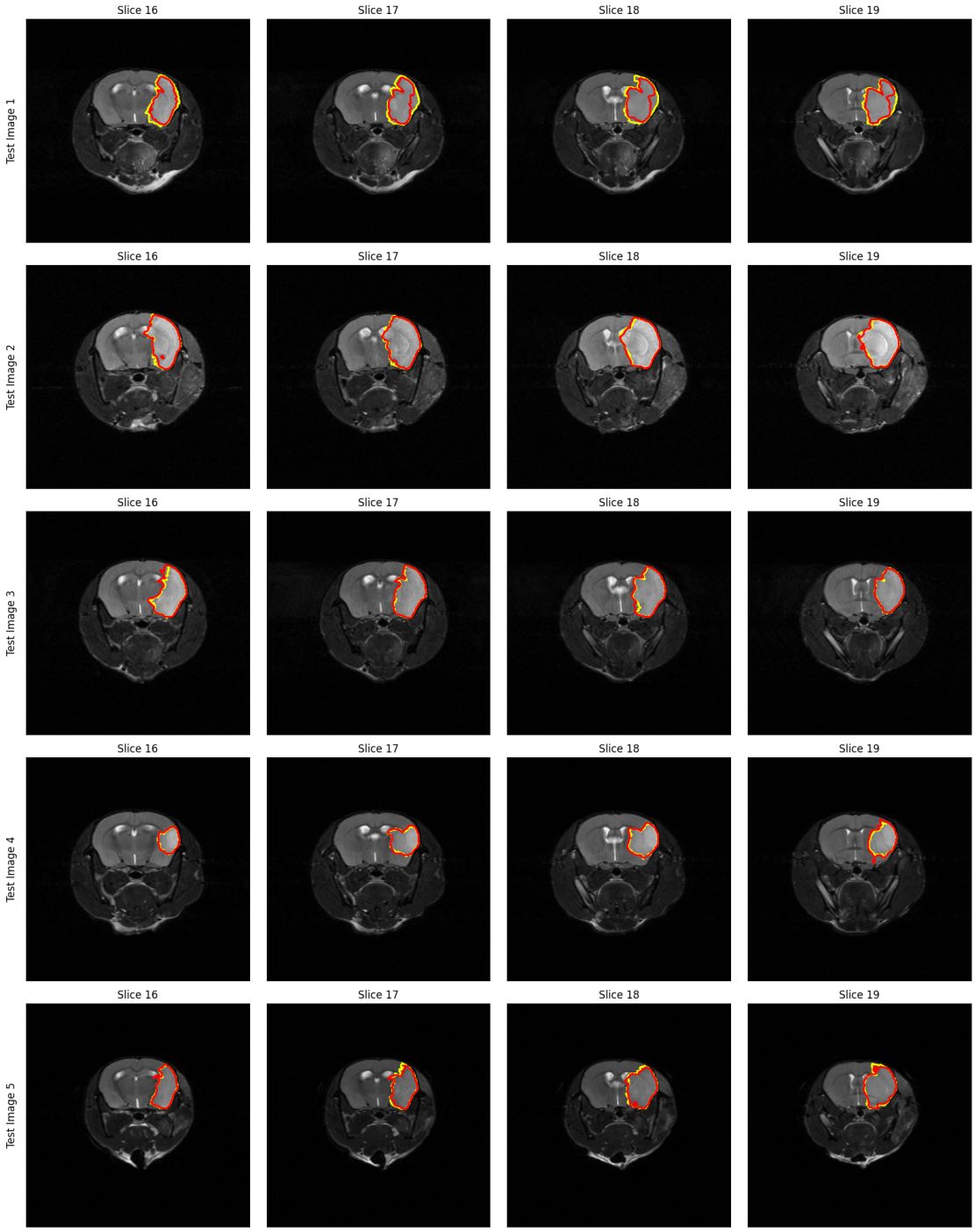

Figure 4.6: *Visual Representation of Segmentation Outcomes from Full Model Fine-Tuning with 40 training samples.*

about 0.03 (from 0.88 to 0.85), indicating that transfer learning strategies are particularly effective when dealing with smaller datasets. When we used the same 40-image training set to train the 3D UNet from scratch with data augmentation, we found DC of 0.78 ± 0.14 and HD95 of 1.00 ± 1.61.

Transfer Learning: Partial vs Full Fine-Tuning: If we compare the results of partial and full fine tuning TL technique used in this study, we can see partial fine tuning provided better segmentation result in term of performance indexes and visual result. A possible reason for partial fine-tuning's better performance can be attributed to the preservation of the pre-learned features in the initial layers. The early layers in a deep learning model usually capture more generalized features, and these preserved features can be particularly beneficial when dealing with smaller datasets, as they provide a strong base on which the model can learn task-specific features. Full fine-tuning, on the other hand, may have a higher tendency to overfit the training data, especially when the data volume is limited, hence leading to a slight decrease in its performance.

Training Efficiency: The time required for the transfer learning model with small training set(8, 16, 32, 40 images) is less than the model trained from scratch(231 images). Reducing the amount of training data directly translates to fewer computations for the model, which leads to faster training times. And we noticed the model trained with TL is getting converged in less number of epochs than the one we trained from scratch. This makes the model training process more efficient and time-effective. The findings demonstrate that transfer learning, by leveraging a smaller, well-chosen dataset, not only maintains performance but also speeds up the training process.

Impact of Data Augmentation: The study revealed a notable finding, when combined with larger datasets, data augmentation had a barely apparent effect on performance.There could be some reason , such as due to the uniformity of image acquisition, as all images were acquired using the same pipeline which have similar contrast. Although, while the precise position as well as size of the lesions varied across samples, the placement of the lesions were mostly on the same side of the brain. Therefore, it seems that the model was unable to fully benefit from the variety added by data augmentation when we used a larger dataset. On the other hand, data augmentation proved to have a noticeable impact, improving the performance of the model, when the dataset was smaller. Examining the Dice Coefficient (DC) and Hausdorff Distance (HD95) values in Tables 4.3 and **??** made this impact evident. As a result, even if data augmentation is still an effective method for improving model performance, its efficiency seems to be influenced by things such as dataset size and natural data variance.

In addition to offering a thorough comparative analysis, the comparison of these three unique techniques also reveals interesting directions for further study. This

investigation will encourage originality and might result in the optimization of methods designed especially for the difficult task at hand.

5 Conclusion

This study provides an innovative effort to investigate the possibilities of transfer learning in medical image analysis in pre-clinical settings, a field where the majority of previous studies have concentrated on clinical data. In the field of pre-clinical image analysis, this work represents an important contribution into the domain transfer learning. The study has navigated the subtle changes in image acquisition across both domains by extending the pre-trained RatLesNetV2 model to the task of segmenting mice MR images. The model's resilience is shown, and the unexplored potential of transfer learning in pre-clinical settings is highlighted by the success in generating consistent lesion segmentation across both domains. By using similarities in imaging tasks despite differences in data features, it creates opportunities for future research to explore the inter-species applicability of such models. Whereas conventional clinical approaches may fall short, this strategy may act as a basis for improving the effectiveness and efficacy of pre-clinical image analysis.

This study's investigation went into a variety of lesion segmentation methods, including transfer learning techniques utilizing the RatLesNetV2 model, the MONAI 3D UNet, UNetR, and DynUnet models. Two different applications of the transfer learning procedures were used: full model fine-tuning and partial fine-tuning after encoder. Each technique was carefully examined to determine its advantages and possible drawbacks.

The 3D UNet model with data augmentation performed well, producing a considerable Dice Coefficient and HD95. RatLesNetV2 was used as a pre-trained model in the transfer learning techniques, which likewise showed considerable results.The performance achieved through partial fine-tuning surpassed that of full fine-tuning in terms of performance and visual segmentation result, closely approaching the results of the 3D UNet. The prevalence of data augmentation, especially in constrained training contexts, emphasizes its potential use in numerous medical image segmentation challenges.

This thorough investigation has outlined areas for more study, pointed up the potential for innovation, and added to our knowledge of various methodologies in a more nuanced manner. It serves as a testimony to the unrealized

potential of pre-clinical image analysis and opens up opportunities for the use of cutting-edge machine-learning methods in this crucial area. The trained models and common training approached used in this study are publicly available here https://github.com/MJHossainS/TL-based-Mice-Brain-Leison-Segmentation.

5.1 Future work

Future study may focus on integrating novel data augmentation methods, perhaps using cutting-edge deep learning techniques like Generative Adversarial Networks (GANs). GANs have the potential to augment the training dataset because of their ability to produce artificial data. However, since artificially produced images must preserve clinical relevance and accuracy, this technique requires a strong validation process. A professional's experience in the particular pre-clinical area would likely be required for this validation stage, which would confirm the accuracy and validity of the synthetic images. Additionally, the accessibility of more specialized models, such RatLesNetV2, offers a chance to expand research into pre-clinical image analysis. The discipline may obtain new knowledge and enhance its capacity for correctly analyzing pre-clinical imaging data by using and creating these models. Therefore, the foundation of future research in the field of brain lesion segmentation on pre-cilincal images may be continuing improvements in augmentation methods and model building.

Bibliography

Ablavsky, V. H., Becker, C. J., and Fua, P. (2012). Transfer learning by sharing support vectors. Technical report. (cited on page 20)

Achille, A., Lam, M., Tewari, R., Ravichandran, A., Maji, S., Fowlkes, C. C., Soatto, S., and Perona, P. (2019). Task2vec: Task embedding for meta-learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6430–6439. (cited on page 18)

Alex, V., Vaidhya, K., Thirunavukkarasu, S., Kesavadas, C., and Krishnamurthi, G. (2017). Semisupervised learning using denoising autoencoders for brain lesion detection and segmentation. *Journal of Medical Imaging*, 4(4):041311–041311. (cited on page 20)

An, J., Wendt, L., Wiese, G., Herold, T., Rzepka, N., Mueller, S., Koch, S. P., Hoffmann, C. J., Harms, C., and Boehm-Sturm, P. (2022). Deep learning-based automated lesion segmentation on mouse stroke magnetic resonance images. *bioRxiv*, pages 2022–08. (cited on page 21)

Atlason, H. E., Love, A., Sigurdsson, S., Gudnason, V., and Ellingsen, L. M. (2019). Unsupervised brain lesion segmentation from mri using a convolutional autoencoder. In *Medical Imaging 2019: Image Processing*, volume 10949, pages 372–378. SPIE. (cited on page 4)

Balafar, M. A., Ramli, A. R., Saripan, M. I., and Mashohor, S. (2010). Review of brain mri image segmentation methods. *Artificial Intelligence Review*, 33:261–274. (cited on page 4)

Balassy, C. and Hörmann, M. (2008). Role of mri in paediatric musculoskeletal conditions. *European journal of radiology*, 68(2):245–258. (cited on page 4)

Belue, M. J., Harmon, S. A., Patel, K., Daryanani, A., Yilmaz, E. C., Pinto, P. A., Wood, B. J., Citrin, D. E., Choyke, P. L., and Turkbey, B. (2022). Development of a 3d cnn-based ai model for automated segmentation of the prostatic urethra. *Academic Radiology*, 29(9):1404–1412. (cited on page 7)

Bi, J., Xiong, T., Yu, S., Dundar, M., and Rao, R. B. (2008). An improved multi-task learning approach with applications in medical diagnosis. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part I 19*, pages 117–132. Springer. (cited on pages 17 and 20)

Byra, M., Sznajder, T., Korzinek, D., Piotrzkowska-Wroblewska, H., Dobruch-Sobczak, K., Nowicki, A., and Marasek, K. (2019). Impact of ultrasound image reconstruction method on breast lesion classification with deep learning. In *Pattern Recognition and Image Analysis: 9th Iberian Conference, IbPRIA 2019, Madrid, Spain, July 1–4, 2019, Proceedings, Part I 9*, pages 41–52. Springer. (cited on page 11)

Cardoso, M. J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al. (2022). Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701*. (cited on page 37)

Chen, C., Shen, F., Xu, J., and Yan, R. (2020). Domain adaptation-based transfer learning for gear fault diagnosis under varying working conditions. *IEEE Transactions on Instrumentation and Measurement*, 70:1–10. (cited on page 1)

Chen, H., Dou, Q., Yu, L., Qin, J., and Heng, P.-A. (2018). Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images. *NeuroImage*, 170:446–455. (cited on page 3)

Chen, M., He, Y., Ye, Q., and Zhu, J. (2019). Tuning plasmonic near-perfect absorber for selective absorption applications. *Plasmonics*, 14:1357–1364. (cited on page 18)

Cheplygina, V., de Bruijne, M., and Pluim, J. P. (2019). Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical image analysis*, 54:280–296. (cited on pages 2, 11, 16, and 18)

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer. (cited on pages 3, 12, 26, and 65)

Conjeti, S., Katouzian, A., Roy, A. G., Peter, L., Sheet, D., Carlier, S., Laine, A., and Navab, N. (2016). Supervised domain adaptation of decision forests: Transfer of models trained in vitro for in vivo intravascular ultrasound tissue characterization. *Medical image analysis*, 32:1–17. (cited on page 17)

Despotović, I., Goossens, B., and Philips, W. (2015). Mri segmentation of the human brain: challenges, methods, and applications. *Computational and mathematical methods in medicine*, 2015. (cited on page 4)

Elliott, P. J., Knapman, J. M., and Schlegel, W. (1992). Interactive image segmentation for radiation treatment planning. *IBM Systems Journal*, 31(4):620–634. (cited on page 14)

Erdi, Y. E., Mawlawi, O., Larson, S. M., Imbriaco, M., Yeung, H., Finn, R., and Humm, J. L. (1997). Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 80(S12):2505–2509. (cited on page 6)

Eskreis-Winkler, S., Zhang, Y., Zhang, J., Liu, Z., Dimov, A., Gupta, A., and Wang, Y. (2017). The clinical utility of qsm: disease diagnosis, medical management, and surgical planning. *NMR in Biomedicine*, 30(4):e3668. (cited on page 4)

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al. (2012). 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341. (cited on page 3)

Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C. R., de Leeuw, F.-E., Tempany, C. M., Van Ginneken, B., et al. (2017). Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, pages 516–524. Springer. (cited on pages 11 and 12)

Gillot, M., Baquero, B., Le, C., Deleat-Besson, R., Bianchi, J., Ruellas, A., Gurgel, M., Yatabe, M., Al Turkestani, N., Najarian, K., et al. (2022). Automatic multi-anatomical skull structure segmentation of cone-beam computed tomography scans using 3d unetr. *Plos one*, 17(10):e0275033. (cited on pages 13 and 14)

Gordillo, N., Montseny, E., and Sobrevilla, P. (2013). State of the art survey on mri brain tumor segmentation. *Magnetic resonance imaging*, 31(8):1426–1438. (cited on page 3)

Haie-Meder, C., Pötter, R., Van Limbergen, E., Briot, E., De Brabandere, M., Dimopoulos, J., Dumas, I., Hellebust, T. P., Kirisits, C., Lang, S., et al. (2005). Recommendations from gynaecological (gyn) gec-estro working group(i): concepts and terms in 3d image based 3d treatment planning in cervix cancer brachytherapy

with emphasis on mri assessment of gtv and ctv. *Radiotherapy and oncology*, 74(3):235–245. (cited on page 4)

Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., and Xu, D. (2021a). Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI Brainlesion Workshop*, pages 272–284. Springer. (cited on pages 7 and 14)

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H., and Xu, D. (2021b). Unetr: Transformers for 3d medical image segmentation. *arXiv preprint arXiv:2103.10504*. (cited on pages 27, 29, and 65)

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778. (cited on page 28)

Hernandez, M., Barrena, R., Hernandez, G., Sapiro, G., and Frangi, A. F. (2003). Pre-clinical evaluation of implicit deformable models for three-dimensional segmentation of brain aneurysms from cta images. In *Medical Imaging 2003: Image Processing*, volume 5032, pages 1264–1274. SPIE. (cited on page 15)

Hulsen, T., Jamuar, S. S., Moody, A. R., Karnes, J. H., Varga, O., Hedensted, S., Spreafico, R., Hafler, D. A., and McKinney, E. F. (2019). From big data to precision medicine. *Frontiers in medicine*, page 34. (cited on page 15)

Huynh, B. Q., Li, H., and Giger, M. L. (2016). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3):034501–034501. (cited on page 19)

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr. (cited on pages 26 and 27)

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211. (cited on pages 7 and 27)

Jyothi, P. and Singh, A. R. (2023). Deep learning models and traditional automated techniques for brain tumor segmentation in mri: a review. *Artificial intelligence review*, 56(4):2923–2969. (cited on page 6)

Karimi, D., Warfield, S. K., and Gholipour, A. (2021). Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations. *Artificial intelligence in medicine*, 116:102078. (cited on pages 11, 16, and 20)

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. (cited on page 33)

Kolarik, M., Burget, R., Travieso-Gonzalez, C. M., and Kocica, J. (2021). Planar 3d transfer learning for end to end unimodal mri unbalanced data segmentation. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6051–6058. IEEE. (cited on page 13)

Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., Meenakshi, K., Swaraja, K., Plawiak, P., and Acharya, U. R. (2022). Transfer learning techniques for medical image analysis: A review. *Biocybernetics and Biomedical Engineering*, 42(1):79–107. (cited on page 11)

Kouw, W. M., Loog, M., Bartels, L. W., and Mendrik, A. M. (2017). Mr acquisition-invariant representation learning. *arXiv preprint arXiv:1709.07944*. (cited on page 20)

Krishnapriya, S. and Karuna, Y. (2023). A survey of deep learning for mri brain tumor segmentation methods: Trends, challenges, and future directions. *Health and Technology*, 13(2):181–201. (cited on page 6)

Kumar, R. P., Sivadas, D. V., and Singh, T. (2022). Comparative study of liver segmentation using u-net and resnet50. In *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–6. IEEE. (cited on page 12)

LaLonde, R., Xu, Z., Irmakci, I., Jain, S., and Bagci, U. (2021). Capsules for biomedical image segmentation. *Medical image analysis*, 68:101889. (cited on pages 5 and 15)

Le Bihan, D. (2012). Diffusion, confusion and functional mri. *Neuroimage*, 62(2):1131–1136. (cited on page 3)

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., and Tu, Z. (2015). Deeply-supervised nets. In *Artificial intelligence and statistics*, pages 562–570. Pmlr. (cited on page 27)

Li, S., Fevens, T., and Krzyzak, A. (2004). Image segmentation adapted for clinical settings by combining pattern classification and level sets. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004: 7th International Conference, Saint-Malo, France, September 26-29, 2004. Proceedings, Part I 7*, pages 160–167. Springer. (cited on page 14)

Li, S., Fevens, T., Krzyżak, A., and Li, S. (2006). Automatic clinical image segmentation using pathological modeling, pca and svm. *Engineering Applications of Artificial Intelligence*, 19(4):403–410. (cited on page 14)

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88. (cited on page 14)

Mahmood, F., Chen, R., and Durr, N. J. (2018). Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. *IEEE transactions on medical imaging*, 37(12):2572–2581. (cited on page 20)

Maier, O. (2015). Medpy. Available at: `https://pypi.python.org/pypi/MedPy`, accessed: 2023-05-27. (cited on page 37)

McEvoy, L. K. and Brewer, J. B. (2010). Quantitative structural mri for early detection of alzheimer's disease. *Expert review of neurotherapeutics*, 10(11):1675–1688. (cited on page 4)

Meng, D., Zhang, L., Cao, G., Cao, W., Zhang, G., and Hu, B. (2017). Liver fibrosis classification based on transfer learning and fcnet for ultrasound images. *Ieee Access*, 5:5804–5810. (cited on page 19)

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al. (2014). The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024. (cited on page 14)

Milano, E. G., Capelli, C., Wray, J., Biffi, B., Layton, S., Lee, M., Caputo, M., Taylor, A. M., Schievano, S., and Biglino, G. (2019). Current and future applications of 3d printing in congenital cardiology and cardiac surgery. *The British journal of radiology*, 92(1094):20180389. (cited on page 4)

Min, J. H., Kim, J. M., Kim, Y. K., Cha, D. I., Kang, T. W., Kim, H., Choi, G. S., Choi, S.-Y., and Ahn, S. (2020). Magnetic resonance imaging with extracellular contrast detects hepatocellular carcinoma with greater accuracy than with gadoxetic acid or computed tomography. *Clinical Gastroenterology and Hepatology*, 18(9):2091–2100. (cited on page 4)

Mitchell, S. C., Bosch, J. G., Lelieveldt, B. P., Van der Geest, R. J., Reiber, J. H., and Sonka, M. (2002). 3-d active appearance models: segmentation of cardiac mr and ultrasound images. *IEEE transactions on medical imaging*, 21(9):1167–1178. (cited on page 14)

Moeskops, P., Wolterink, J. M., Van Der Velden, B. H., Gilhuijs, K. G., Leiner, T., Viergever, M. A., and Išgum, I. (2016). Deep learning for multi-task medical image segmentation in multiple modalities. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 478–486. Springer. (cited on pages 18, 19, and 20)

Nisar, H., Carnahan, P. K., Fakim, D., Akhuanzada, H., Hocking, D., Peters, T. M., and Chen, E. C. (2022). Towards ultrasound-based navigation: deep learning based ivc lumen segmentation from intracardiac echocardiography. In *Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 12034, pages 467–476. SPIE. (cited on page 13)

Ohishi, T., Takahashi, M., Abe, M., Tsuchikawa, T., Mori, M., and Nagano, A. (2005). The use of axial reconstructed images from three-dimensional mri datasets for morphological diagnosis of meniscal tears of the knee. *Archives of Orthopaedic and Trauma Surgery*, 125:622–627. (cited on page 4)

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32. (cited on page 36)

Pham, D. L., Xu, C., and Prince, J. L. (2000). Current methods in medical image segmentation. *Annual review of biomedical engineering*, 2(1):315–337. (cited on pages 4 and 5)

Ronneberger, O., Fischer, P., and Brox, T. (2015). 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. (cited on pages 3, 12, and 28)

Ross, T., Zimmerer, D., Vemuri, A., Isensee, F., Wiesenfarth, M., Bodenstedt, S., Both, F., Kessler, P., Wagner, M., Müller, B., et al. (2018). Exploiting the potential of unlabeled endoscopic video data with self-supervised learning. *International journal of computer assisted radiology and surgery*, 13:925–933. (cited on page 20)

Rote, G. (1991). Computing the minimum hausdorff distance between two point sets on a line under translation. *Information Processing Letters*, 38(3):123–127. (cited on page 35)

Schlegl, T., Ofner, J., and Langs, G. (2014). Unsupervised pre-training across image domains improves lung tissue classification. In *Medical Computer Vision: Algorithms for Big Data: International Workshop, MCV 2014, Held in Conjunction with MICCAI 2014, Cambridge, MA, USA, September 18, 2014, Revised Selected Papers*, pages 82–93. Springer. (cited on page 19)

Schwarz, A. J. (2021). The use, standardization, and interpretation of brain imaging data in clinical trials of neurodegenerative disorders. *Neurotherapeutics*, 18(2):686–708. (cited on page 6)

Singh, C. and Bala, A. (2021). An unsupervised orthogonal rotation invariant moment based fuzzy c-means approach for the segmentation of brain magnetic resonance images. *Expert Systems with Applications*, 164:113989. (cited on page 4)

Sivakumar, P. and Ganeshkumar, P. (2017). An efficient automated methodology for detecting and segmenting the ischemic stroke in brain mri images. *International Journal of Imaging Systems and Technology*, 27(3):265–272. (cited on page 12)

Standley, T., Zamir, A., Chen, D., Guibas, L., Malik, J., and Savarese, S. (2020). Which tasks should be learned together in multi-task learning? In *International Conference on Machine Learning*, pages 9120–9132. PMLR. (cited on page 18)

Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., and Lu, J. (2019). Brain tumor classification for mr images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics*, 75:34–46. (cited on page 1)

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312. (cited on page 20)

Tian, D. and Fan, L. (2007). A brain mr images segmentation method based on som neural network. In *2007 1st International Conference on Bioinformatics and Biomedical Engineering*, pages 686–689. IEEE. (cited on page 4)

Valverde, J. M., Imani, V., Abdollahzadeh, A., De Feo, R., Prakash, M., Ciszek, R., and Tohka, J. (2021). Transfer learning in magnetic resonance brain imaging: A systematic review. *Journal of imaging*, 7(4):66. (cited on page 11)

Valverde, J. M., Shatillo, A., De Feo, R., Gröhn, O., Sierra, A., and Tohka, J. (2019). Automatic rodent brain mri lesion segmentation with fully convolutional networks. In *Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10*, pages 195–202. Springer. (cited on pages 14 and 15)

Valverde, J. M., Shatillo, A., De Feo, R., Gröhn, O., Sierra, A., and Tohka, J. (2020). Ratlesnetv2: a fully convolutional network for rodent brain lesion segmentation. *Frontiers in neuroscience*, 14:610239. (cited on pages 8, 15, 22, 29, 30, and 65)

Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., Oliver, A., and Lladó, X. (2017). Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach. *NeuroImage*, 155:159–168. (cited on page 14)

Van Der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *Computing in science & engineering*, 13(2):22–30. (cited on page 36)

Van Opbroek, A., Ikram, M. A., Vernooij, M. W., and De Bruijne, M. (2014). Transfer learning improves supervised image segmentation across imaging protocols. *IEEE transactions on medical imaging*, 34(5):1018–1030. (cited on pages 11 and 20)

van Opbroek, A., Vernooij, M. W., Ikram, M. A., and de Bruijne, M. (2015). Weighting training images by maximizing distribution similarity for supervised segmentation across scanners. *Medical image analysis*, 24(1):245–254. (cited on page 20)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30. (cited on page 27)

Wang, B., Prastawa, M., Saha, A., Awate, S. P., Irimia, A., Chambers, M. C., Vespa, P. M., Van Horn, J. D., Pascucci, V., and Gerig, G. (2013). Modeling 4d changes in pathological anatomy using domain adaptation: Analysis of tbi imaging using a tumor database. In *Multimodal Brain Image Analysis: Third International Workshop, MBIA 2013, Held in Conjunction with MICCAI 2013, Nagoya, Japan, September 22, 2013, Proceedings 3*, pages 31–39. Springer. (cited on page 20)

Warfield, S. K., Zou, K. H., and Wells, W. M. (2004). Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *IEEE transactions on medical imaging*, 23(7):903–921. (cited on page 14)

Węgliński, T. and Fabijańska, A. (2011). Brain tumor segmentation from mri data sets using region growing approach. In *Perspective Technologies and Methods in MEMS Design*, pages 185–188. IEEE. (cited on page 6)

Weiss, K., Khoshgoftaar, T. M., and Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1):1–40. (cited on page 18)

Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560. (cited on pages 32 and 33)

Westbrook, C. and Talbot, J. (2018). *MRI in Practice*. John Wiley & Sons. (cited on page 3)

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500. (cited on page 25)

Zamir, A. R., Sax, A., Shen, W., Guibas, L. J., Malik, J., and Savarese, S. (2018). Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3712–3722. (cited on page 18)

Zhang, D., Shen, D., Initiative, A. D. N., et al. (2012). Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. *NeuroImage*, 59(2):895–907. (cited on page 17)

Zhang, X., Huang, J., Yang, Y., He, X., Liu, R., and Zhong, N. (2019). Applying python in brain science education. In *2019 International Joint Conference on Information, Media and Engineering (IJCIME)*, pages 396–400. IEEE. (cited on page 36)

Zhou, Y., Huang, W., Dong, P., Xia, Y., and Wang, S. (2019). D-unet: a dimension-fusion u shape network for chronic stroke lesion segmentation. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(3):940–950. (cited on page 13)

Zhou, Z., Shin, J. Y., Gurudu, S. R., Gotway, M. B., and Liang, J. (2021). Active, continual fine tuning of convolutional neural networks for reducing annotation efforts. *Medical image analysis*, 71:101997. (cited on page 11)

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76. (cited on pages 1 and 2)

Zoetmulder, R., Gavves, E., Caan, M., and Marquering, H. (2022). Domain-and task-specific transfer learning for medical segmentation tasks. *Computer Methods and Programs in Biomedicine*, 214:106539. (cited on page 16)

Zuluaga, M. A., Rodionov, R., Nowell, M., Achhala, S., Zombori, G., Mendelson, A. F., Cardoso, M. J., Miserocchi, A., McEvoy, A. W., Duncan, J. S., et al. (2015). Stability, structure and scale: improvements in multi-modal vessel extraction for seeg trajectory planning. *International journal of computer assisted radiology and surgery*, 10:1227–1237. (cited on page 14)

BIBLIOGRAPHY

List of Figures

LIST OF FIGURES

List of Tables