# 3D CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES FOR VOLUMETRIC MEDICAL IMAGE SEGMENTATION

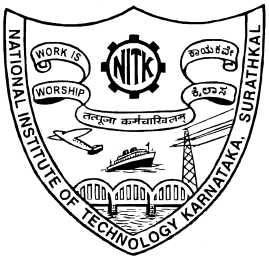
Thesis

Submitted in partial fulfillment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

by

**NIYAS S**




**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA**
**SURATHKAL, MANGALORE - 575025, INDIA**

**October 2024**

# DECLARATION

I hereby ***declare*** that the Research Thesis entitled **3D CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES FOR VOLUMETRIC MEDICAL IMAGE SEGMENTATION** which is being submitted to the ***National Institute of Technology Karnataka, Surathkal*** in partial fulfillment of the requirements for the award of the Degree of ***Doctor of Philosophy*** is a ***bona fide report of the research work carried out by me***. The material contained in this thesis has not been submitted to any University or Institution for the award of any degree.

**NIYAS S**
Registration No.: 187530
Department of Computer Science and Engineering
National Institute of Technology Karnataka
Surathkal-575025

Place: NITK - Surathkal

Date: 18-10-2024

# CERTIFICATE

This is to *certify* that the Research Thesis entitled **3D CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES FOR VOLUMETRIC MEDICAL IMAGE SEGMENTATION**, submitted by **NIYAS S** (Registration No: 187530) as the record of the research work carried out by him, is *accepted* as the *Research Thesis submission* in partial fulfillment of the requirements for the award of degree of ***Doctor of Philosophy***.

18/10/24

**Dr. Jeny Rajan**
Research Guide
Associate Professor
Department of Computer Science and Engineering
NITK, Surathkal - 575025

18/10/24

**Chairman - DRPC**
Department of Computer Science and Engineering
National Institute of Technology Karnataka
Surathkal, 575025

Chairman
DUGC / DPGC / DRPC
Dept. of Computer Engg
NITK - Surathkal
Srinivasnagar - 575 025

(Signature with Date and Seal)

***Forever Learning, Forever Curious:***
***A Research Path Begins***

# ACKNOWLEDGEMENTS

I am immensely grateful to those who have supported and guided me through the intricate journey of my Ph.D research. First and foremost, I would like to express my profound appreciation to my advisor, Dr. Jeny Rajan, Associate Professor, Department of CSE, NITK Surathkal, India, for his expert guidance, patience, and unwavering support. His insightful feedback and encouragement were pivotal in navigating the complexities of my research. He exemplified the role of a supportive advisor, consistently providing timely and constructive feedback. His mentorship not only honed my skills as a researcher but also granted me substantial autonomy in my investigative pursuits, stepping in with valuable advice whenever I faced challenges. Beyond the realm of research, I observed and learned from his kindness, patience, discipline, and exemplary time management. These qualities, which he demonstrated throughout our collaboration, are ones I aspire to integrate into my everyday life.

I feel privileged to have conducted my doctoral research at the prestigious National Institute of Technology Karnataka (NITK), Surathkal, India. My gratitude extends to the Department of Computer Science and Engineering (CSE) at NITK, which provided essential support throughout my doctoral journey. I would also like to express my deep appreciation for the Cognitive Science Research Initiative (CSRI) of the Government of India, which has provided financial support for some part of my research. I would like to extend my thanks to the members of my Doctoral Research Progress Analysis Committee, Dr. Basavaraj Talawar, Associate Professor, Department of CSE, NITK Surathkal, India, and Dr. Shyam Lal, Associate Professor, Department of ECE, NITK Surathkal, India, for their thorough evaluations and constructive suggestions that significantly enriched my doctoral research.

My sincere appreciation is extended to Dr. P. Santhi Thilagam, Dr. Alwyn Roshan Pais, Dr. Shashidhar G. Koolagudi, and Dr. Manu Basavaraju, who served as Heads of the Department of Computer Science at NITK Surathkal during my academic tenure. Their support and kindness have been invaluable to my journey. I am profoundly grateful to Prof. Karanam

Uma Maheshwar Rao, Prof. Prasad Krishna, and Prof. B Ravi the esteemed Directors of NITK Surathkal during my study period, for providing an excellent academic environment and extending their generous support.

During my tenure at NITK, I was fortunate to immerse myself in a vibrant research environment, bolstered by outstanding mentorship and collaboration. I extend my heartfelt gratitude to my colleagues at the Vision and Image Processing lab. Special thanks go to my esteemed colleagues Dr. Pawan S. J., Dr. Anoop B. N., Dr. Sankar Pariserum Perumal, Mr. Siva Bonthada, Dr. Tojo Mathew, Mr. Yamanappa, Mr. Ajith B., Ms. Akhila P., Mr. Pradyoth Hegde, Mr. Siva Krishna, Mr. Neethi A. S., Mr. Poornanand Naik, Mr. Shivam Kumar, Mr. Asjad Nabeel, and Ms. Chaitanya Lakshmi, whose support were pivotal to my journey. Their collaboration and insight greatly contributed to my professional journey and personal growth during my time at NITK. I have had the immense privilege of collaborating with an exceptional group of individuals at NITK, including Ms. Chethana Vaisali, Ms. Iwrin Show, Ms. Chandrika T. G., Mr. Edwin Thomas, Mr. Shushant Kumar, Mr. Bijay Dev, Ms. Ramya Bygari, Ms. Rachita Naik, Ms. Bhavishya Viswanath, Mr. Dhananjay Ugwekar, Ms. Shraddha Priya, and Ms. Reena Oswal. Each of them has brought unique skills and insights to our work.

I would like to extend my deepest gratitude to Dr. Chandrasekharan Kesavadas (SCTIMST, Trivandrum), Dr. S. Vinayagamani (SCTIMST, Trivandrum), Dr. Jyoti R. Kini (MAHE, KMC Mangalore), Dr. Vivek A. Saraf (DetectIQ, Kochi), and Dr. Ajith Abraham (Bennett University, Noida) for their invaluable mentorship and for co-authoring my research work. I want to express my sincere appreciation for my friend, Dr. Terry Jacob Mathew, who has been the cornerstone of my academic and professional journey. His belief in my potential and constant encouragement were the driving forces that motivated me to embark on this challenging yet rewarding journey. I'd be remiss not to mention the incredible support system I've had throughout this journey. Huge thanks to Dr. Libin P. Oommen, Dr. Fredy James, Dr. Sachin, Mr. Sudeesh, Dr. Sushan Lal, and Dr. Deepak – you folks, along with your families, have been there for me through thick and thin.

I must also pause to remember Vimal G. S., who was not only a friend but a brother in spirit. His unwavering personal support played a crucial role in my journey. Though he is no longer with us, his positive influence and cherished memories have left an indelible mark on my

life and work. I dedicate a portion of my success to his memory, grateful for the time we shared and the impact he had on my journey.

I am immensely grateful to my family, whose unconditional love, patience, and encouragement have been my constant source of strength and inspiration. To my parents, whose sacrifices and unwavering support have shaped the person I am today, I owe a debt of gratitude that words cannot express. I am profoundly grateful to my father, Shamsudeen Kutty, whose unwavering strength of character and sheer determination not only shaped my own, but also steered me towards the relentless pursuit of excellence. Expressing gratitude to my mother, Naseema, seems almost inadequate; her sacrifices and unwavering support have been the bedrock of my endeavors. To my partner, Sumayya Muhammed, whose understanding and companionship have been my solace, and to my children, Imtiaz Muhammed and Izwa Parveen who bring endless joy and motivation into my life, you all are my greatest treasures. This accomplishment would not have been possible without the countless sacrifices we made together. I am eternally grateful for your partnership and understanding. This journey would have been unimaginable without your boundless love and support. You are my sanctuary and the cherished heart of my life's endeavors.

I am deeply grateful to all those who have directly or indirectly contributed to the successful completion of my doctoral research work. I recognize that this journey wouldn't have been possible without your support, and I apologize if I have inadvertently omitted anyone's name.

NIYAS S

Place: NITK - Surathkal

Date: 18 -October-2024

# ABSTRACT

Computer-aided medical image analysis plays a critical role in supporting medical practitioners with expert clinical diagnoses and determining optimal treatment plans. Currently, convolutional neural networks (CNNs) are widely regarded as the preferred method for automated medical image analysis due to their ability to autonomously learn relevant features from training data. However, most cutting-edge semantic image segmentation techniques rely on two-dimensional (2D) CNN models, which do not fully exploit the inter-slice information available in cross-sectional imaging modalities, such as MRI volumes. This limitation underscores the need for more advanced approaches to better utilize the three-dimensional (3D) data inherent in these imaging techniques.

In this thesis, we present a comprehensive evaluation of various techniques employed in 3D deep learning for medical image segmentation. With the rapid advancements in 3D imaging systems and excellent hardware and software support to process large volumes of data, 3D deep learning methods are gaining popularity in medical image segmentation. However, traditional 3D CNN-based segmentation models require substantial computational resources, extensive memory, and typically larger datasets than 2D CNN approaches. To address these challenges, we propose a 3D CNN segmentation model that efficiently extracts information across slices and mitigates several limitations associated with traditional 3D CNN techniques. The method aims to retain the advantages of both 2D CNN and 3D CNN methods by effectively designing input data slices and the CNN architecture. In this study, we proposed a shallow sliced stacking approach to reduce the depth of input 3D data to maintain a good segmentation accuracy with minimum computation overhead and model complexity. Incorporating residual connections in the encoder path also facilitates the extraction of multi-scale features without significantly increasing the model complexity.

Accurate diagnosis of various medical conditions often requires the simultaneous analysis of multiple image characteristics. For instance, Focal Cortical Dysplasia (FCD) lesion detec-

tion can be significantly enhanced by incorporating data on cortical thickness maps along with fluid-Attenuated Inversion Recovery (FLAIR) Magnetic Resonance Imaging (MRI) scans. Additionally, employing multi-axis analysis of 3D cross-sectional imaging can substantially improve diagnostic performance. Inspired by these concepts, we propose a 3D deep learning model employing a multi-view, dual encoder-decoder architecture. The model also incorporates various architecture-wise enhancements, including an end-to-end cascaded approach for transitioning from coarse to fine segmentation, 3D Attention modules for maintaining consistency between encoder and decoder pairs, and dual-task learning. In our study, we apply this model to process FLAIR MRI volumes alongside corresponding cortical thickness maps, aiming to effectively detect FCD lesions.

Generative Adversarial Networks (GANs) have significantly impacted the field of image analysis, and they have been successfully employed for tasks such as image segmentation. Hence, this study also proposes a 3D attention-driven Vox2Vox CNN network that leverages the power of a 3D GAN to accurately segment acute stroke lesion cores in Computed Tomography Perfusion (CTP) scans. This methodology also incorporates valuable insights derived from our prior models relevant to this research. The segmentation framework incorporates two supervised GAN components: a generator and a discriminator. The generator module is designed to process 3D slices from CTP maps and learn to generate 3D binary prediction masks that closely match the ground truth for stroke lesions. Concurrently, the discriminator module is trained to distinguish between the outputs generated by the generator and the actual ground truth. Overall, this thesis demonstrates the efficacy of 3D deep learning in identifying malignancies from cross-sectional imaging modalities, including CT and MRI, thereby enhancing the capabilities of automated Computer-Aided Detection (CAD) systems.

*Keywords*: 3D Attention; 3D Deep Learning; Computer-aided Medical Image Analysis; Convolutional Neural Networks; Computed Tomography Perfusion; Dual-Task Learning; Fluid-Attenuated Inversion Recovery; Generative Adversarial Networks; Image Segmentation; Shallow Sliced Stacking; Vox2Vox

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABBREVIATIONS AND NOMENCLATURE

| | |
|---|---|
| AVD | Absolute Volume Difference |
| BCE | Binary Cross-Entropy |
| BM3D | Block-Matching and 3D Filtering |
| CADx | Computer-Aided Diagnosis |
| CBAM | Convolutional Block Attention Module |
| CNN | Convolutional Neural Network |
| CRF | Conditional Random Field |
| CT | Computed Tomography |
| DQN | Deep Q Network |
| DSC | Dice Similarity Coefficient |
| DWI | Diffusion-Weighted Imaging |
| EMMA | Ensemble of Multiple Models and Architectures |
| FCD | Focal Cortical Dysplasia |
| FCN | Fully Convolutional Network |
| FLAIR | Fluid-Attenuated Inversion Recovery |
| GANs | Generative Adversarial Networks |
| HNNs | Holistically Nested convolutional Networks |
| MRI | Magnetic Resonance Imaging |
| MTL | Multi-Task Learning |
| OBELISK | One Binary Extremely Large and Inflecting Sparse Kernel |
| OM-Net | One-pass Multi-task Network |
| PET | Positron Emission Tomography |
| RCE | Reduced Coulomb Energy |
| S3D | Separable 3D |

| | |
|---|---|
| SDM | Signed Distance Map |
| SELU | Scaled Exponential Linear Unit |
| SNR | Signal-to-Noise Ratio |
| SSL | Semi-Supervised Learning |

# CHAPTER 1

# INTRODUCTION

This chapter presents an introduction to deep learning applications in medical image segmentation. It gives an overview of Convolutional Neural Network (CNN) based segmentation approaches and discusses how 3D CNN is advantageous over 2D CNN methods in the automatic analysis of cross-sectional imaging. The chapter also delves into the inherent limitations of current 3D CNN-based segmentation techniques for medical images and discusses strategies to mitigate these challenges. Additionally, the chapter outlines the research motivation, objectives, and contributions made by this study.

## 1.1 AI-based Medical Image Analysis

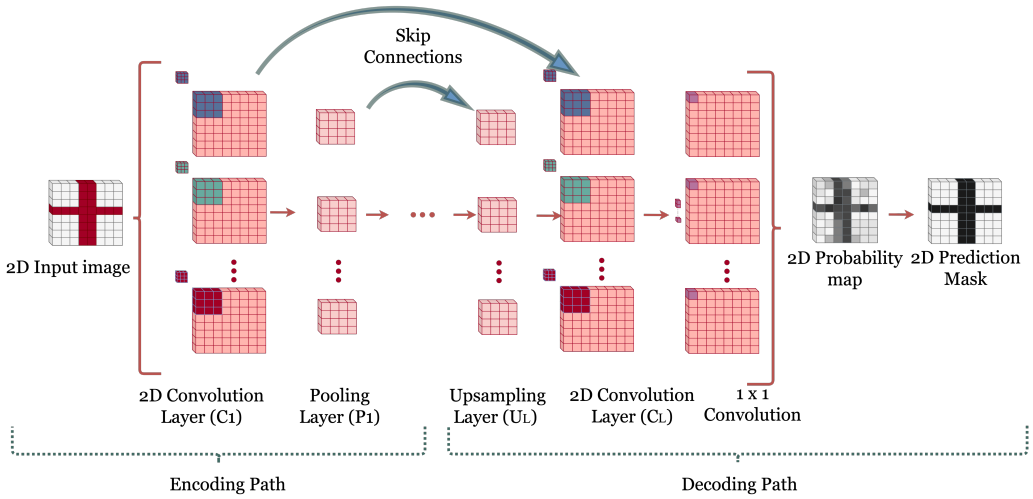
Artificial Intelligence (AI) techniques have been increasingly used in modern healthcare devices to analyze medical images. Computer-Aided Diagnosis (CADx) refers to analyzing medical data and extracting useful information using computer software to assist clinicians in making a rapid and error-free diagnosis. These diagnostic systems reduce the subjectivity in decision-making and the overall cost involved. Medical images play a significant role in clinical diagnosis, treatment planning, teaching, and research. In recent years, the quality of popular medical imaging techniques, such as X-ray, Computed Tomography (CT), Ultrasonography, Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET), has improved significantly in terms of acquisition time, image quality, and cost-effectiveness (Chakraborty *et al.*, 2018).

Medical image processing encompasses problem-specific strategies using image processing algorithms. Its common applications include image registration, denoising, enhancement, compression, classification, and segmentation. The increase in the volume and complexity of medical image data has accelerated the development of algorithms for generalized feature extraction, which has increased the need for supervised machine learning algorithms. Traditional machine learning approaches are based on application-specific feature extraction techniques for analyzing image characteristics such as contrast variation, orientation, shape, and texture patterns. Since handcrafted features are an integral part of such methods, the algorithm design requires domain experts, making human intervention inevitable.

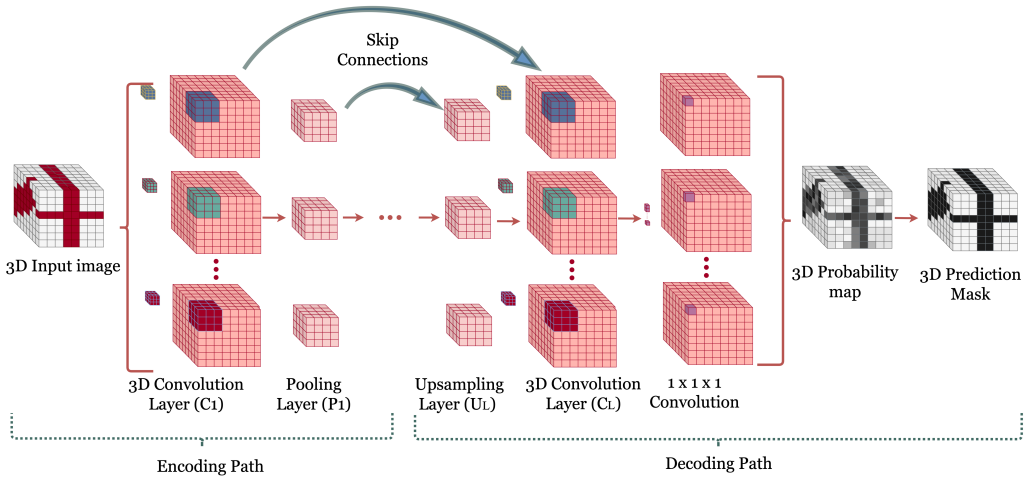
Studies on neural network-based decision-making have been conducted since the 1950s. In the last few decades, factors such as data availability, advancements in parallel distributed computing, and the surge in the semiconductor industry have accelerated research on neural networks. With the increasing prominence of image data in the digital data space and the need for adaptive feature learning, research on developing neural network models for images has increased tremendously, as evidenced by the introduction of Convolutional Neural Networks (CNNs).

### 1.1.1 An Introduction to Convolutional Neural Networks

A typical CNN model uses multiple convolution operations followed by activation and pooling layers for extracting multi-scale features; the complexity of image features advances as the model depth increases. The fully connected dense layer is designed to transform the feature map from the previous layer to one-dimensional feature vectors. The final fully connected layer operates as a classification layer and provides the probabilities of the target classification task. Furthermore, a CNN model uses several additional modules, such as batch normalization (Ioffe and Szegedy, 2015), regularization, and dropout (Srivastava *et al.*, 2014), to improve the learning process (Singh *et al.*, 2020).




(a)




(b)

Figure 1.1: (a) A 2D CNN architecture for segmentation and (b) A 3D CNN architecture for segmentation.

In image segmentation, every pixel needs to be classified simultaneously. CNN models for segmentation use convolution and pooling layers similar to those in a classification network. However, the fully connected layers are usually excluded in the segmentation model to retain the spatial relationship of pixels throughout the network. In addition,

adequate upsampling layers are added to the final layers of the network to compensate for the pooling operation. Hence, the probability of each pixel is estimated, and finally, the segmentation mask is created. A general workflow of an encoding–decoding architecture for segmentation is depicted in Figure 1.1 (a).

### 1.1.2 An Overview on 3D Medical Data

Medical images use different modalities based on imaging principles. The characteristics of these images differ in terms of spatial resolution, image intensity range, size of the image, and noise. In addition, they vary in terms of dimensionality and its way of representation. 2D data mostly use relatively simple Euclidean representation, which describes each data point using pixel intensity values. In medical images, these pixel values represent the state of the anatomical structure under consideration. For instance, the pixel intensity of X-ray images varies with the radiation absorption, whereas it depends on the acoustic pressure in ultrasound images or the radio frequency (RF) signal amplitude in MRI. The representation of 3D data is more complex. In 3D medical image representation, volumetric elements or *voxels* (analogous to *pixels* in a 2D image) are used to model 3D data by describing how the 3D object is distributed through the three axes: axial (front to back), coronal (top to bottom), or sagittal (side to side).

The 3D image visualization has provided a great opportunity for clinicians to evaluate the cross-section of anatomic structures. This has increased the understanding of complex patterns and structural morphology, mostly in radiology. Currently, 3D imaging is commonly used in several modalities, such as CT, MRI, USG, and PET. Although 3D imaging techniques have numerous advantages over 2D images, they have certain limitations as well. For example, compared to 2D imaging methods, they require significantly large storage space and are often expensive. However, with decreasing computational expenses, higher networking speeds, and the availability of powerful graphics processing units (GPU), visualization and analysis of 3D medical images have become easy (Zhou *et al.*, 2021*a*).

### 1.1.3 An Introduction to 3D CNN-based Segmentation

A 3D CNN-based segmentation model uses 3D images as input, and a similar-sized 3D prediction mask is expected as the output. The basic network structure is similar to that of a standard 2D CNN model, but the convolution and pooling layers use 3D extensions to process the volumetric data. Here, the convolution layers perform filtering with 3D kernels, and the 3D pooling layers subsample the data in all three dimensions to compress the size of the feature space. Hence the volumetric data is analyzed as cubic patches from layer to layer and can learn spatial features in all three dimensions. A 3D CNN architecture for segmentation is shown in Figure 1.1 (b).

The segmentation architecture usually consists of an encoding and decoding path like the U-Net architecture proposed by Ronneberger et al. (Ronneberger *et al.*, 2015). In the encoding path, 3D convolution layers and pooling layers create 4D feature space. Since the pooling reduces the feature space dimension in the encoding path, the decoding path uses a sufficient number of upsampling layers to gradually increase the feature space dimension similar to the input data. Normally, transpose convolution is used for upsampling the data in a learnable fashion. Skip connections are also used to preserve the fine features by merging features between the corresponding encoding and decoding layers. A $1 \times 1 \times 1$ filter is used in the final layer to project the stacked features into a feature space with the same dimension as the input image. The probability map is then created using a non-linear activation function, followed by a thresholding that creates the final prediction mask.

## 1.2 Motivation

Medical image segmentation plays a crucial role across various phases of clinical practice. Consequently, the development of reliable automated methods for this process carries broad implications. Most of the state-of-the-art medical image segmentation approaches

use 2-dimensional (2D) deep learning approaches to limit the memory consumption and computation cost while training the model. However, 2D CNN frameworks couldn't learn the inter-slice relationship between the frames, which is important while analyzing 3D images. 3D CNN helps to retrieve features from multiple adjacent slices and will be useful for extracting inter-slice relationships along with the 2D spatial information. In medical images, lesions or abnormalities are often spread over multiple frames, and hence, 3D CNN can significantly improve the segmentation performance over such data. However, it is very challenging to design a 3D CNN architecture for medical image segmentation. The main challenges are listed below.

1. **Lack of large 3D datasets:**

   Building a sufficiently large 3D dataset is quite challenging due to the intrusive nature of some medical imaging modalities, the prolonged imaging duration, and the laborious annotation required while making a large 3D cross-sectional image dataset.

2. **Computational cost and hardware limitations:**

   3D CNN uses convolution kernels with three dimensions and makes the computations significantly higher while processing large 3D image datasets. This attracts the need for higher computation requirements and costlier GPUs.

## 1.3 Problem Statement

This research proposal aims to develop efficient 3D CNN-based segmentation algorithms to detect and segment abnormal regions from 3D medical images without a substantial increase in computational complexity while maintaining competitive performance.

### 1.3.1 Research Objectives

1. **Objective 1**: Design and Develop a Simplified 3D CNN-Based Segmentation Model with Optimal Performance.

In deep 3D CNN, the number of trainable parameters is significantly higher than in corresponding 2D networks. This demands huge memory and undesirably long training time, even with high-performance GPUs. Moreover, 3D networks are often forced to work with low batch sizes, which leads to non-optimized weight updation and eventually degrades the segmentation performance.

2. **Objective 2**: Develop a High-Performance 3D Cascaded CNN Model for Multi-modal Medical Image Segmentation with Accurate Lesion Segmentation.

In conventional CNN networks, when the model tries to segment the lesion regions with high recall, the number of false detections also increases and will impact the precision performance. Therefore, by integrating a localization phase followed by a focused segmentation phase, cascaded networks can be employed. This approach facilitates fine-level segmentation, resulting in a more precise and thorough segmentation.

3. **Objective 3**: Develop an optimized 3D CNN segmentation technique that can effectively use Generative AI techniques.

Several medical images can learn from multiple image modalities together, and advancements in generative AI techniques can boost their performance. Hence, developing such a 3D CNN model is a need of the hour.

## 1.4 Major Contributions

This thesis significantly adds contributions in the field of medical image segmentation by employing 3D deep learning techniques. The main contributions in this thesis are summarized as follows:

1. **A Deep 3D CNN Segmentation Model with Minimal Computation and Complexity.**

We propose a deep 3D CNN segmentation model that efficiently extracts information across slices, overcoming the limitations of traditional CNN techniques. This model retains the advantages of both 2D and 3D CNN methods through effectively designed input data slices and customized encoder-decoder structures. A shallow sliced stacking approach is introduced to reduce the size of input 3D data, maintaining high segmentation accuracy with minimal computation overhead and model complexity. Additionally, incorporating residual connections in the encoder path facilitates the extraction of multi-scale features without significantly increasing the model complexity.

2. **Multi-View Dual Encoder-Decoder Architecture for 3D Deep Learning.**

Inspired by multi-axis analysis (sagittal, axial, and coronal axis) of 3D cross-sectional imaging, which improves the detection performance of neuro-radiologists, we propose a novel 3D deep learning model employing a multi-view dual encoder-decoder architecture. This model includes architecture-wise enhancements such as an end-to-end cascaded approach for transitioning from coarse to fine segmentation, 3D Attention modules to ensure consistency between encoder and decoder pairs, and dual-task learning. We apply this model to process Fluid-Attenuated Inversion Recovery (FLAIR) MRI volumes of the brain and the corresponding cortical thickness maps to detect Focal Cortical Dysplasia (FCD) lesions.

3. **3D Attention-Driven Vox2Vox Network using GANs.**

Leveraging the transformative impact of Generative Adversarial Networks (GANs) in image analysis, our research presents a 3D attention-driven Vox2Vox network that utilizes the capabilities of a supervised 3D GAN. This network is designed to accurately segment acute stroke lesion cores in Computed Tomography Perfusion (CTP) scans. The approach incorporates valuable insights derived from our prior models, enhancing its effectiveness and relevance to current medical imaging challenges.

## 1.5 Organization of this Thesis

Rest of the thesis is organized as follows:

**Chapter 2** presents an in-depth review on state-of-the-art 3D deep learning-based medical image segmentation.

**Chapter 3** presents a 3D CNN model for medical image segmentation that leverages the advantages of both 2D and 3D CNN approaches.

**Chapter 4** presents a segmentation model with a dual encoder-decoder 3D CNN architecture with multi-view training, 3D attention, and an end-to-end cascaded network design.

**Chapter 5** presents a novel 3D attention-driven Vox2Vox framework that investigates the transformation of a 3D image translational CNN network for segmenting lesions from 3D cross-sectional imaging.

**Chapter 6** concludes this thesis by summarizing the research conducted and exploring potential future directions, emphasizing the importance of 3D deep learning-based segmentation techniques for biomedical image analysis.

# CHAPTER 2

# LITERATURE SURVEY

This chapter presents an extensive overview of the popular 3D deep learning methods used for medical image segmentation. Additionally, it explores the current challenges and prospective developments in the field of medical image segmentation utilizing 3D deep learning technologies. This study scrutinizes 3D CNN approaches for segmenting cross-sectional medical images, concentrating on recently published scholarly articles. Based on our review, the 3D CNN segmentation approaches for medical image segmentation are organized into various categories, as outlined below.

1. Fully supervised 3D CNN models

   (a) Direct 3D CNN models

   (b) 3D Patch-wise segmentation models

   (c) Multi-task learning models

2. 3D CNN with Semi-supervised learning

3. 3D CNN with Weakly-supervised learning

4. Cost-effective approximations of 3D CNN

[1]The work described in this chapter has been published in: **Niyas S.** Pawan S. J., Anand Kumar M., and Jeny Rajan (2022). Medical image segmentation with 3D convolutional neural networks: A survey. Neurocomputing. 493, 397-413.

## 2.1 Fully Supervised 3D CNN Models

### 2.1.1 Direct 3D CNN Models

A straightforward implementation of a 3D CNN model is possible by replacing the 2D convolution and pooling operations in conventional CNN models with 3D counterparts. Such CNN models are discussed under this category.

Milletari et al. (Milletari *et al.*, 2016) presented an advanced 3D U-Net architecture (V-Net) with residual blocks (instead of cascaded convolution blocks) and strided convolution (instead of max-pooling). The methodology uses a Dice score-based loss function to reduce the class imbalance between the voxel classes. Bui et al. (Bui *et al.*, 2017) proposed another 3D CNN architecture inspired by the Fully Convolutional Network (FCN) model proposed by Long et al. (Long *et al.*, 2015) for brain segmentation in infant brain MRI volumes. The encoder path uses multiple coarse and fine 3D convolution layers to extract multi-scale features. Downsampling uses strided convolution to reduce the feature size and increase the receptive field. Multiple convolution layers are used to extract four different scales of feature space, then upsampled and concatenated to generate the final probability map.

Kayalibay et al. (Kayalibay *et al.*, 2017) proposed another 3D encoder-decoder architecture with deep supervision. In this model, the feature space created at different decoder levels in the network is merged using an element-wise summation of the features. In the approach, the learning process directly depends on the coarse features from the different decoder levels and helps to speed up the convergence. However, the element-wise summation may limit the learning process when the semantic gap between the different decoder levels is significant. Dou et al. (Dou *et al.*, 2017) proposed a similar approach that can work on relatively small 3D datasets.

Li et al. (Li *et al.*, 2017) presented a 3D CNN model using dilated convolution instead of max-pooling and residual connections. For utilizing multi-scale features, the

dilation factor of the 3D convolution kernel is steadily increased in the subsequent layers. Hence, the spatial resolution of the feature space is kept constant throughout the network. The residual blocks merge the features from different layers to reduce the vanishing gradients problem and feature degradation. Chen et al. (Chen *et al.*, 2018*a*) proposed another residual deep 3D CNN architecture for segmenting the brain region from 3D MRI volumes. The proposed architecture employs a Voxel-wise Residual Network (VoxResNet). The architecture uses a 3D extension of 2D deep residual networks that extracts features from multiple scales using a series of convolution operations, and the features at multiple scales are fused to generate the segmentation mask.

Schlemper et al. (Schlemper *et al.*, 2019) proposed an attention gate based 3D U-Net architecture for medical image segmentation which automatically learns to concentrate on various target structures. Wang et al. (Wang *et al.*, 2019) proposed another 3D FCN method that integrates recursive residual blocks and pyramid pooling to extract more complex features. The recursive residual blocks contain multiple residual connections that can minimize feature degradation problems. Pyramid pooling (He *et al.*, 2015*b*) generates fused feature maps at different decoding levels for obtaining both local and global information. It helps to eliminate the fixed size constraints of CNN without losing spatial information.

A 3D version of multi-scale U-Net segmentation was presented by Peng et al. (Peng *et al.*, 2020). The model uses multiple U-Net modules to extract long-distance spatial information at different scales. The U-Net blocks use Xception (Chollet, 2017) modules instead of normal convolution to extract more complex features. Zhou et al. (Zhou *et al.*, 2020) proposed a lightweight deep 3D CNN architecture: One-pass Multi-task Network (OM-Net) that can handle the native flaws in the MC approach. OM-Net combines discrete segmentation tasks into a one-pass deep model to learn joint features and solve the class imbalance problem better than MC. The prediction results between tasks correlate using a cross-task guided attention block that can adaptively re-calibrate channel-wise feature responses.

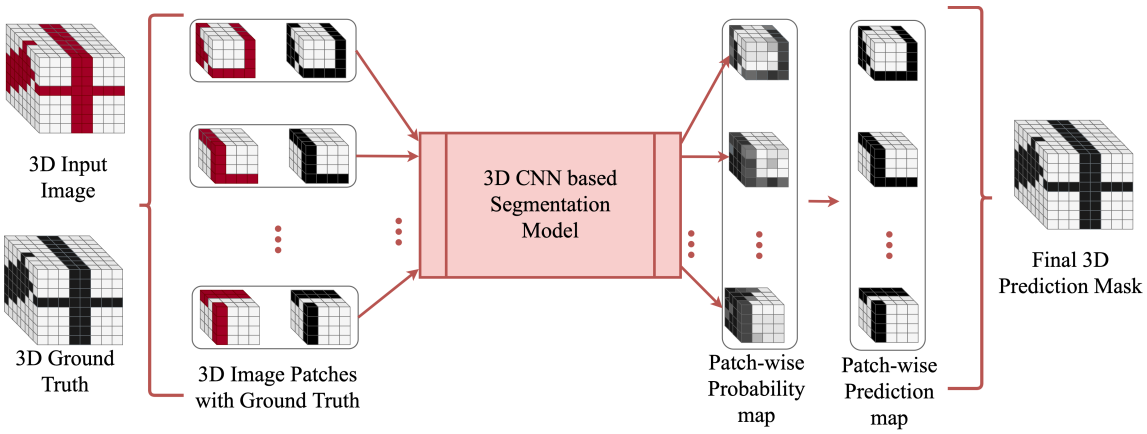



Figure 2.1: A high-level block diagram of 3D Patch-wise segmentation model.

In the above-mentioned medical image segmentation approaches, the 3D CNN models mainly use a straightforward extension of various 2D CNN building blocks. However, the volumetric data analysis brings a significant increase in the memory requirement and computation costs while training deep 3D CNN models compared to corresponding 2D versions. Several studies were also conducted in developing 3D segmentation models that work with limited training data and hardware resources. Such approaches are discussed in the following sections.

### 2.1.2 3D Patch-wise Segmentation Models

The resolution of input 3D data is the primary reason behind the need for large memory and higher computation complexity in training a 3D CNN segmentation model. However, this could be circumvented with a patch-based approach. In the patch-wise approach, the input 3D images are converted to small 3D sub-samples and analyzed individually. Finally, the patch-wise prediction outcomes will be merged to create the final 3D prediction map. A high-level block diagram of 3D Patch-wise segmentation is shown in Figure 2.1. The patch-wise analysis is an interesting area of research in 3D medical image analysis, and such patch-wise 3D deep learning based techniques are discussed in this section.

Yu et al.(Yu *et al.*, 2016) proposed a deep supervised 3D fractal network for whole

heart and great vessel segmentation in MRI volumes. This method expands the FractalNet (Larsson *et al.*, 2016) and uses deep supervision using multiple auxiliary classifiers deployed at expanding layers to reduce the vanishing gradient problem. The methodology uses cropped 3D patches of size $64 \times 64 \times 64$ and reported good results in the HVSMR 2016 challenge dataset (Pace *et al.*, 2015).

Kamnitsas et al. (Kamnitsas *et al.*, 2017*a*) presented an Ensemble of Multiple Models and Architectures (EMMA) for robust performance by aggregating predictions from multiple models. This method uses an ensemble of a DeepMedic model (Kamnitsas *et al.*, 2016), three 3D FCN models (Long *et al.*, 2015), and two 3D U-Net models (Ronneberger *et al.*, 2015). An ensembler computes the confidence maps and finds the average class confidence of the individual models. Nevertheless, the ensemble with seven 3D architectures requires higher computation complexity and training time.

A deep dual pathway 3D CNN architecture was also proposed by Kamnitsas et al. (Kamnitsas *et al.*, 2017*b*) for brain lesion segmentation. They employed a two-way network that simultaneously learns from multiple image scales to analyze features from different receptive fields. The approach uses 3D patches at two different scales and classifies the center voxels as any target classes using fully connected dense layers. A 3D fully connected Conditional Random Field (CRF) is also utilized in the post-processing stage to reduce false alarms in the final segmentation mask.

Chen et al. (Chen *et al.*, 2018*b*) proposed a hierarchical 3D CNN architecture to segment Glioma from multi-modal brain MRI volumes. The model uses two distinct scales of 3D image patches to examine multi-scale features. The 3D CNN architecture uses a series of hierarchical dense convolutions without pooling layers. The model uses a patch-wise analysis that reduces the class imbalance and the computational cost of training large 3D datasets. The patch-based segmentation helps to reduce the hardware resources for training and generates a large number of data samples that favor adequate learning. However, the patch-wise analysis often fails to extract global features from the actual image volumes. This may limit the learning performance when the abnormality is

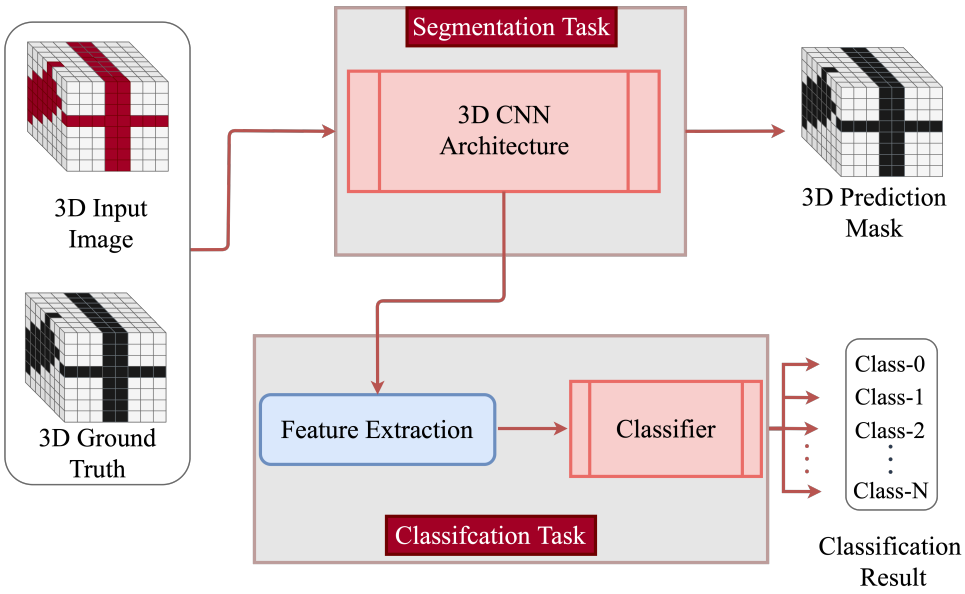



Figure 2.2: A high-level block diagram of 3D CNN based Multi-Task learning model.

region-specific.

### 2.1.3 Multi-Task Learning Models

Multi-Task Learning (MTL) is a machine learning approach that can assess multiple related tasks with a single network (Vafaeikia *et al.*, 2020). MTL is advantageous in medical images because many tasks, such as classification and segmentation, may be required concurrently throughout the diagnostic process (Wu *et al.*, 2021). A graphical representation of a 3D CNN-based MTL is shown in Figure 2.2. This section discusses a few such approaches that use 3D CNN for multi-task learning.

Zhou et al. (Zhou *et al.*, 2021*b*) proposed multi-task learning of classification and segmentation using a 3D CNN model for classifying tumors in breast ultrasound images. The authors used a modified version of 3D V-Net (Milletari *et al.*, 2016) as the backbone of this model. To conduct a multi-scale analysis, feature maps from three higher-level encoder and decoder layers of the segmentation network are fused together to perform the classification task.

Another multi-tasking approach has been reported by Gordaliza et al. (Gordaliza *et al.*, 2019) to infer tuberculosis from CT images. The model uses higher-order encoder features and processes two individual feed-forward neural networks to understand the tuberculosis condition. The approach also used several optimizations such as self-normalization, the use of the Scaled Exponential Linear Unit (SELU) activation function, and uncertainty-weighted multi-task loss to improve the performance both for detection and counting the number of nodules.

Ge et al. (Ge *et al.*, 2019) proposed a multi-task learning approach for segmenting and quantifying the left ventricle from paired apical views (A4C and A2C) of echo sequences. The method offers an overall cardiac analysis using a multi-task network: K-Net, an end-to-end network that can simultaneously segment the left ventricle and quantify its extent over the 3D plane. The methodology uses 2D convolution in different stages to segment and quantify the 3D structure of the left ventricle using complex echo data. The reported results over a sufficiently large echo dataset also prove the proposed model's superiority in the heterogeneous learning approach.

## 2.2 Semi-Supervised Learning

The 3D CNN models discussed in Section 2.1 mostly use fully supervised deep learning algorithms that require thousands of annotated 3D volumes. However, accurate marking of ground-truth images is a labor-intensive and tedious process. Hence, the supervised learning algorithms are more expensive in terms of time and cost. Consequently, research also commenced on alternatives to process sparsely annotated training data. Semi-Supervised Learning (SSL) methods (Zhu and Goldberg, 2009; Rasmus *et al.*, 2015; Snell *et al.*, 2017) are one of those types that require a few labeled image samples with a large number of unlabeled samples, and such SSL-based 3D medical image segmentation works are discussed in this section. A basic 3D SSL framework is shown in Figure 2.3.

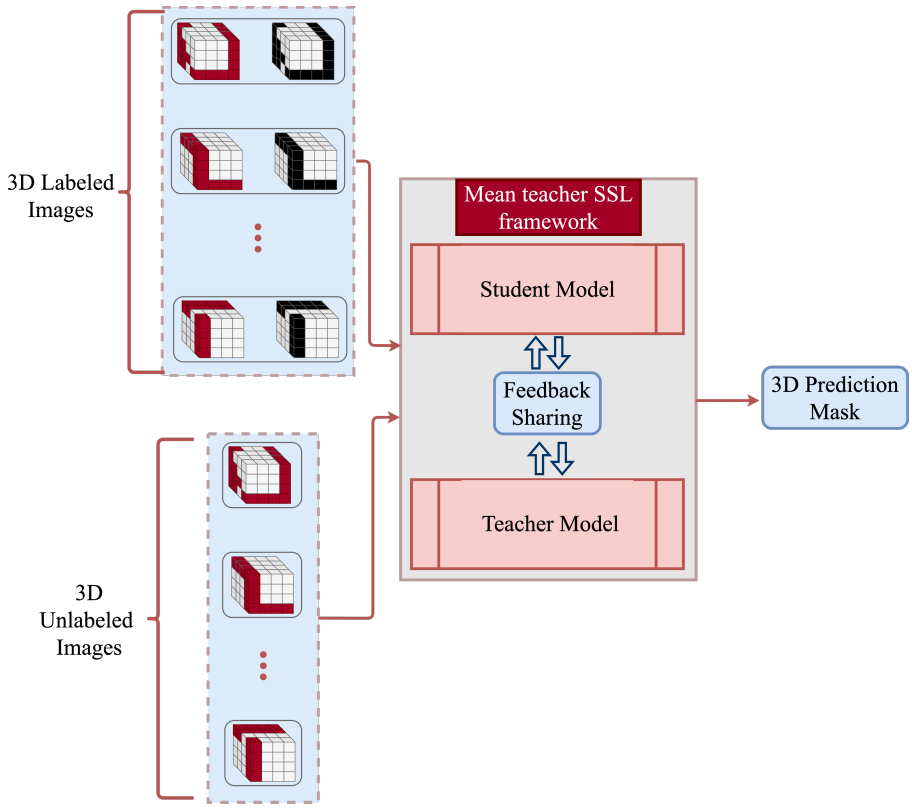



Figure 2.3: A high-level representation of a 3D CNN based Mean teacher Semi-supervised learning model.

Çiçek et al. (Çiçek *et al.*, 2016) presented one of the first promising 3D CNN network for volumetric segmentation that learns from sparsely labeled 3D images. This study outlined two cases: a semi-automated model and a fully automated model. In both cases, the network learns from sparse annotated data and this helps to reduce the human effort in ground truth labeling without considerable degradation in the segmentation performance.

Mondal et al. (Mondal *et al.*, 2018) proposed a 3D multi-modal medical image segmentation method based on Generative Adversarial Networks (GANs) (Goodfellow *et al.*, 2014). The method uses a semi-supervised training with a mix of labeled and unlabeled images. The proposed architecture uses several design considerations to modify the standard adversarial learning approaches to generate 3D volumes of multiple modalities. The 3D GANs generate fake samples and are used along with labeled and unlabeled 3D vol-

umes, and a discriminator defines separate loss functions for these labeled, unlabeled and synthetic (fake) training samples. However, generating useful synthetic samples may not represent the actual data distribution and thus becomes challenging working with 3D medical image data.

Zhou et al. (Zhou *et al.*, 2019*a*) proposed a straight-forward semi-supervised segmentation approach named deep multi-planar co-training (DMPCT), which uses parallel training to extract information from multiple planes. The multiplanar fusion generates reliable pseudo-labeling to train deep segmentation networks. The DMPCT framework consists of a teacher network that uses fully labeled images for training. The trained model then creates the pseudo labels from the unlabelled training data with a multi-planar fusion module. Subsequently, the student model uses both labeled and pseudo-labeled data for the final training process.

Yang et al. (Yang *et al.*, 2020*b*) proposed a similar semi-supervised segmentation technique to detect catheters from volumetric ultrasound images. The segmentation model uses a Deep Q Network (DQN) for localizing the target region. After the catheter localization, the method uses a twin-UNet model for the semantic segmentation of the catheter volume around the localized region by a patch-based strategy. This model uses a typical teacher network followed by a student network to train labeled and unlabeled 3D patches based on a set of hybrid constraints.

Li et al. (Li *et al.*, 2020) presented a shape-aware 3D segmentation for medical images to use extensive unlabeled data to enforce a geometric shape analysis on the segmentation output. The model uses a deep CNN architecture that predicts semantic segmentation and Signed Distance Map (SDM) of object surfaces. The network uses an adversarial loss between the predicted SDMs of labeled and unlabeled data during training to leverage shape-aware features. The integration of adversarial loss, which uses a generative discrimination function, helps supervise learning with unlabelled data and extract generalized features.

Wang et al. (Wang *et al.*, 2020*a*) proposed a tailored modern SSL method named

as FocalMix for the detection of lesions from 3D medical images. The model is built on MixMatch (Berthelot *et al.*, 2019) SSL framework, which uses a prediction for unlabeled images and MixUp augmentation. The proposed 3D CNN model uses a Soft-target Focal Loss along with an anchor-level target prediction to improve lesion detection. The study also uses two adaptive augmentation methods: image-level MixUp and object-level MixUp, to generate the final training data.

Zhang et al. (Zhang and Zhang, 2021) proposed a 3D medical image segmentation model using semi-supervised 3D CNN. This dual-task mutual learning model uses two side-by-side frameworks. One network works on the region-based shape constraint, while the learning in the other network focuses on boundary-based surface mismatch. The main contribution of this model is the use of a signed distance map (SDM) and the conventional ground truth maps to get a better intuition of region features and shape features together. During the training with labeled image volumes, the loss function concentrates more towards the difference in segmentation map and actual ground truth. On the other hand, while training with unlabeled images, the model uses a consistency loss function based on distance maps to ensure prediction consistency while working with similar images. Hence, the dual network tries to reduce both losses and aids a better segmentation accuracy.

Another semi-supervised method is presented by Li et al. (Li *et al.*, 2021), which uses a 3D CNN-based mean teacher framework with hierarchical consistency regularization for 3D left atrium MR images. The model felicitates the prediction consistency between the teacher and student network at multiple scales. During training, the student network uses multi-scale deep supervision while hierarchically regularizing the network's prediction consistency. Hence, the model learns from both the labeled and unlabeled volumes by minimizing the supervised loss and consistency loss concurrently.

In semi-supervised approaches, the training demands a subset of data with accurately marked ground truths. Usually, multiple networks are used in semi-supervised models to process the labeled and unlabeled data using a shared feedback mechanism. The

learning is often synchronized by evaluating the segmentation losses and consistency in predicting unlabeled data. The performance of a semi-supervised model highly depends on this feedback and the accuracy of the annotated images. Since SSL requires precise annotation (for a subset of data), it is challenging while dealing with large 3D image datasets. Hence, several algorithms that can learn from weakly labeled data have also been published in the medical image segmentation domain. Those methods are discussed in the below section.

## 2.3 Weakly Supervised Learning

Weakly supervised is a learning paradigm that uses noisy or low-quality annotations in the learning process. In weakly supervised learning, the data labeling need not be as highly accurate as in the case of fully supervised learning. In medical image segmentation, weakly-supervised learning is highly significant as the abstract level annotation is relatively easier and may be accomplished by non-experts with minimal support from radiologists. The labels for weakly-supervised learning can also be made from batch clustering or from noisy predictions from comparable pre-trained models. Hence, the data preparation becomes relatively cheaper and practical, but at the cost of more noise or false labeling in the training data. A block-level representation of a weakly-supervised 3D CNN is shown in Figure 2.4. Several 3D CNN models with weak supervision have been reported recently, and those works are discussed here.

Yang et al. (Yang *et al.*, 2020*a*) proposed a weakly-supervised method for segmenting catheters from 3D frustum ultrasound images. The methodology uses data annotated with 3D bounding boxes over the catheter regions. A pseudo-label generator module is introduced here to reduce the impact of inaccurate ground truth marking. This model uses a sequential network with a localization module to detect frustum volumes, followed by the segmentation stage to extract catheter voxels. The localization network used 3D ResNet-10 encoder architecture, and the feature maps were then converted to the final

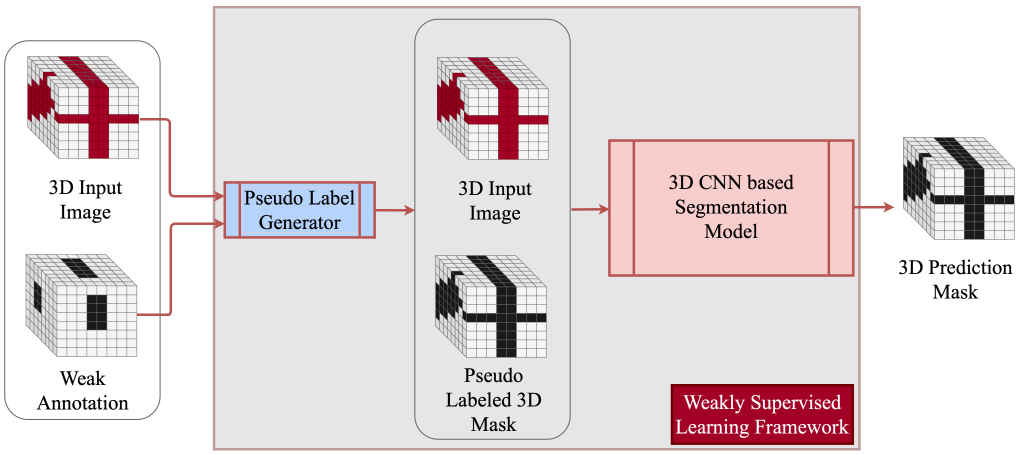



Figure 2.4: Block diagram of a Weakly-supervised 3D CNN model.

segmentation map using a decoder with multiple transpose convolutions.

Wu et al. (Wu *et al.*, 2019) proposed a weakly-supervised brain lesion segmentation using attentional representation learning from 3D image volumes with image-level annotation. The approach used an attention mechanism that is dimensionally independent on the class activation map and can estimate the lesion labels with minimum demands of the trainable parameters and learn the representation model from the dimensional independent class activation map to extract the foreground voxels.

Another similar weakly-supervised 3D CNN-based segmentation is proposed by Zhu et al. (Zhu *et al.*, 2021) that requires only image-level class labels. The proposed model: CIVA-Net, uses weakly annotated labels for volumetric image segmentation on 3D cryo-ET datasets. The input to the network is image-level class labels, and a pre-processing seed generation stage is used initially for generating pseudo labels for each voxel. Using the *cross-image consensus* stage, the similar voxel groups are merged using co-occurrence learning to generate the pseudo localization map. An inter-voxel affinity learning is also proposed to analyze the inter-pixel relationship from the pseudo localization map to forge the affinity voxel pairs. The final segmentation stage uses VoxResNet (Chen *et al.*, 2016) to predict the segmentation map using the pseudo localization map and the affinity pairs generated from previous steps.

Weakly supervised learning is highly recommended for 3D medical images where the voxel shares a similar pattern in the region of interest. However, the pseudo-labeling of the region of the region of interest just from the image class label is highly susceptible to errors and can reduce the segmentation performance significantly.

## 2.4 Cost-effective Approximations of 3D CNN

Since the straight-forward 3D CNN models are highly susceptible to large computation costs and require large datasets, several alternative techniques have been proposed that can use simulated 3D architectures with cost-effective approximations. For instance, 2D frames from a 3D image can be processed in three orthogonal directions and then merged to create the final 3D prediction mask, and such a model is shown in Figure 2.5. Numerous prominent approximations for segmenting 3D medical images have been reported and discussed in this subsection.

Extraction of inter-slice features from a 3D volume is possible by analyzing three orthogonal planes (sagittal, coronal, and axial planes) and by classifying the voxel at the intersection of three planes. This method was successfully applied by Ciompi et al. (Ciompi *et al.*, 2017) for classifying lung nodules. This model learns from 2D patches in three perpendicular planes centered at a given voxel at three different scales. Fully connected layers combine the streams and perform the voxel classification. However, CNN-based models with fully connected layers are computationally less efficient for segmentation than FCN architectures. A similar 3D segmentation approach by combining information from three orthogonal planes was discussed by Kitrungrotsakul et al. (Kitrungrotsakul *et al.*, 2019). The proposed method (VesselNet) uses a 2D DenseNet (Huang *et al.*, 2017) network for classifying voxels in three different 2D planes. The features from the individual networks are then fused to get the probability of predicting the segmentation map. However, this approach is not an end-to-end segmentation model and hence misses most contextual and location-based information.

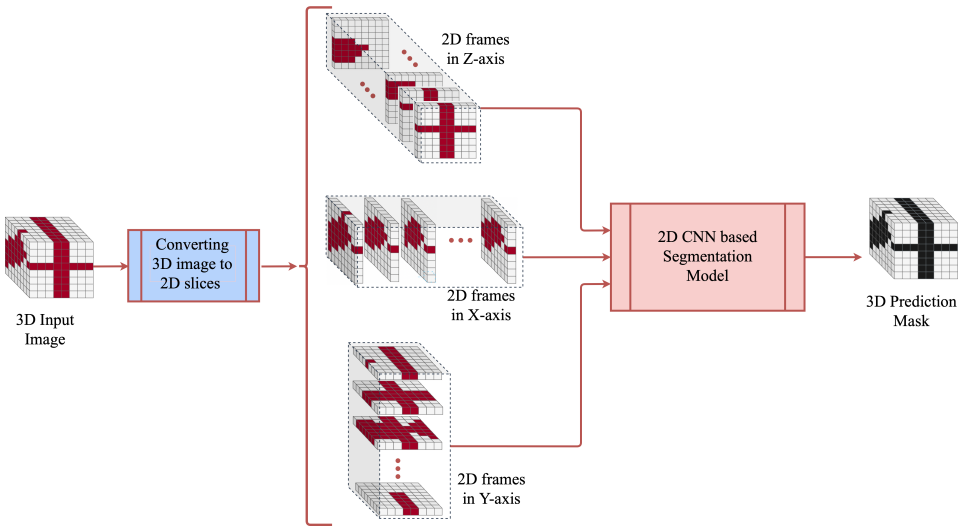



Figure 2.5: A cost-effective approximation of a 3D CNN model.

Mlynarski et al. (Mlynarski *et al.*, 2019) proposed a similar CNN architecture that combines the benefits of the short-range 3D context and the long-range 2D context. This architecture uses modality-specific sub-networks to focus on missing MR sequences. During training, three individual 2D U-Net models were used to create features from axial, coronal, and sagittal slices. The final 3D network was then trained using these 2D features along with 3D patches from the input volumes. The study also suggests considering the outputs from multiple 3D models to minimize the constraints of specific choices of neural network models. However, the use of multiple 2D and 3D models demands more hardware resources, and patch-wise analysis fails to learn region-specific features.

Another 3D CNN model was introduced by Heinrich et al. (Heinrich *et al.*, 2019), which uses trainable 3D convolution kernels that learn both filter coefficients and spatial filter offsets in a continuous space based on the principle of differentiable image interpolation. The proposed One Binary Extremely Large and Inflecting Sparse Kernel (OBELISK) works with fewer parameters and reduced memory consumption. The deformable convolutions in the OBELISK filter replace the continuously sampled spatial filter with a sparse sampling approach. This helps to extract information from wider spatial contexts and replace multiple small filter kernels at different scales. However,

the computation complexity may be higher in OBELISK due to the unoptimized filter sampling.

Roth et al. (Roth *et al.*, 2018*a*) proposed an automated segmentation system for 3D CT pancreas volumes based on a dual-stage cascaded method. It included a localization method followed by a segmentation. 2D Holistically Nested convolutional Networks (HNNs) (Xie and Tu, 2015) on the three orthogonal directions were used in the localization stage. The 2D HNN pixel probability maps are then merged to get a 3D bounding box of the foreground regions. In the second stage, the model focuses on semantic segmentation over the voxels in the bounding box. Two different HNN realizations are integrated into the segmentation stage that extracts mid-level cues of deeply-learned boundary maps of the target organ. The authors also presented an advanced multi-class segmentation model (Roth *et al.*, 2018*b*) for segmenting the liver, spleen, and pancreas from CT volumes. The cascaded network helps to provide boundary-preserving segmentation and reduces false detection in the 3D volumes.

Chen et al. (Chen *et al.*, 2018*c*) presented a separable 3D U-Net architecture that targets to extract spatial information from the image volumes with limited memory and computation cost. The model uses a U-Net architecture for the end-to-end training with separable 3D (S3D) convolution blocks. The S3D block is an arrangement of parallel 2D convolutions and exploits the advantages of the residual inception architecture. The model is evaluated on BRATS 2018 (Bakas *et al.*, 2018) data, and the results justify the improvement in the segmentation performance compared to a standard 2D or 3D U-Net architecture.

Another 3D CNN architecture proposed by Rickmann et al. (Rickmann *et al.*, 2020) incorporates a compress-process-recalibrate pipeline using 3D recalibration methods. The method uses a project & excite module that compress the intermediate high dimensional 4D feature maps into three 2D projection vectors. The convolution layers in the *processor* module learns from this 2D projections and the final recalibration stage generate the 4D tensors for the subsequent layers. This approach reduces the computation cost, without

degrading the segmentation performance. The paper reported improved overall accuracy over different multi-class segmentation datasets.

The volumetric medical segmentation using the above-discussed approaches is highly useful in reducing the computation complexity and the need for large datasets. The results in the above discussed works prove its supremacy over other conventional 2D and 3D CNN approaches. However, the 3D approximations are highly dependent on the characteristics of the input data and the type of segmentation task. This may limit the generic use of models across different medical image modalities, and the designing of such a universal model remains an open research challenge.

Some pre-trained 3D medical segmentation models have also been discussed in the literature. Chen et al. (Chen *et al.*, 2019) proposed a heterogeneous 3D network called Med3D by co-training multi-domain 3D datasets to develop multiple pre-trained 3D medical image segmentation models. The authors use datasets from various medical challenges to create a 3D segmentation data set (3DSeg-8) with different modalities, target organs and pathologies. The pre-trained models were experimented on several 3D medical datasets (Armato *et al.*, 2011) using transfer learning to achieve performance gain and faster convergence. The model can work well on similar image modalities with less training time and can be considered as a suitable option for small 3D image datasets. Zhou et al. (Zhou *et al.*, 2019*b*) proposed a similar set of pre-trained 3D CNN networks for classification and segmentation tasks in CT and MRI. The models are collectively known as Generic Autodidactic Models (Models Genesis), which uses learning by self-supervision. However, transfer learning may not be an appropriate solution in various scenarios due to the difference in image features across different imaging modalities and targeted abnormalities.

## 2.5 Summary

In this research, we conducted a comprehensive evaluation of the latest advancements in medical image segmentation through the application of 3D deep learning technologies. We contend that deep learning plays a crucial role in enhancing segmentation tasks involving 3D medical images. Despite the dominance of 2D methods, there has been a notable surge in research papers exploring 3D deep learning for medical image analysis over recent years. However, numerous segmentation approaches that employ 3D CNNs encounter various optimization challenges, which restricts the analysis of 3D data to its full potential.

This study also provides insights into future directions in the domain of 3D CNN-based cross-sectional image segmentation. This analysis aims to provide the research community with deeper understanding of current trends, future opportunities, and prevailing challenges, thereby facilitating the generation of innovative solutions that bridge existing research gaps in 3D medical image segmentation.

# CHAPTER 3

# SEGMENTATION OF FCD LESIONS FROM MRI USING 3D CNNs

## 3.1 Introduction

This chapter presents a 3D CNN model that efficiently extracts information across slices, overcoming the limitations of traditional 2D and 3D CNN techniques. This model automates the segmentation of FCD lesions from 3D fluid attenuation inversion recovery (FLAIR) Magnetic Resonance Imaging (MRI) images of brain by effectively designing input data slices and a customized encoder-decoder structure.

FCD is a type of neuronal malformation in the brain cortex and is the leading cause of intractable epilepsy, regardless of gender or age differences. Since neuron-related abnormalities are usually resistant to drug therapy, surgical resection has been the main treatment approach for patients with intractable epilepsy (Dingledine and Hassel, 2016). Automating the identification and segmentation of FCD is useful for neuroradiologists in pre-surgical evaluations. However, only 60-70% of the patients enjoy a seizure-free life after the surgery (Alexandre Jr *et al.*, 2006; Hauptman and Mathern, 2012; Simpson and Prayson, 2014). The post-surgical seizures are due to the presence of residual dysplastic lesions caused by incomplete resection of the lesion region. Inaccurate lesion boundary estimation is the main reason behind such incomplete resection. Hence the treatment will be more effective if it is possible to identify the true extent of the lesion and perform a

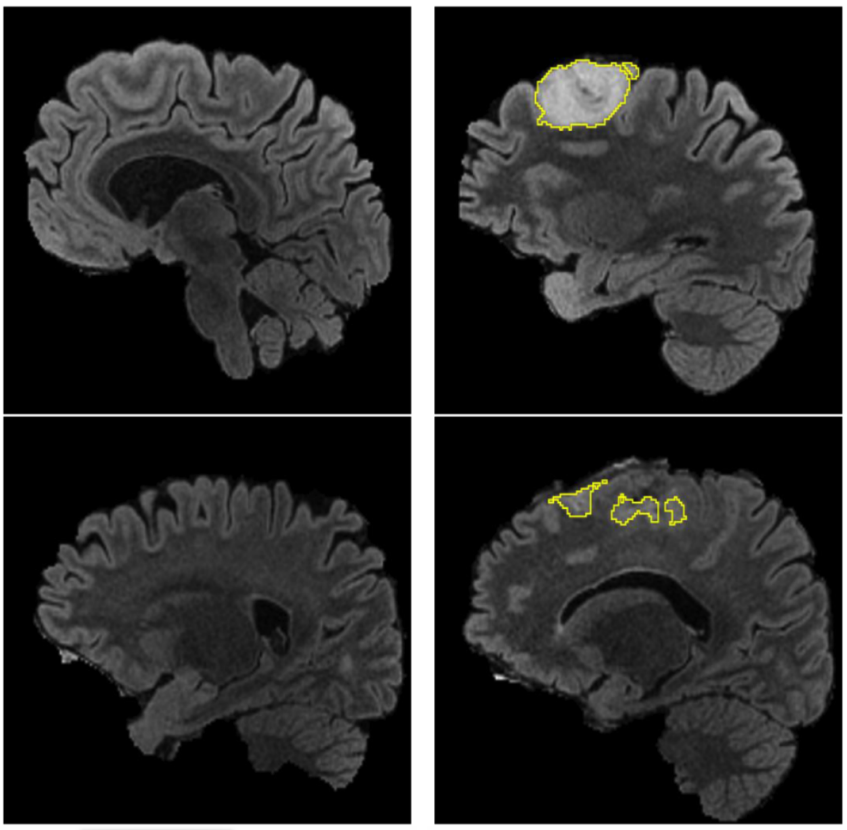
[3]The work described in this chapter has been published in: **S. Niyas**, S. C. Vaisali, Iwrin Show, T. G. Chandrika, S. Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan (2021). Segmentation of focal cortical dysplasia lesions from magnetic resonance images using 3D convolutional neural networks. Biomedical Signal Processing and Control. 70, 102951.

complete resection of the lesion (Krsek *et al.*, 2009). However, accurate estimation of the lesion region/boundary manually through visual analysis is laborious and challenging.

One of the common imaging techniques for assessing brain pathology is MRI (Kabat and Król, 2012). Automated segmentation of FCD lesions from MRI is a possible solution for accurately estimating the FCD regions. Figure 3.1 depicts FLAIR MRI slices of skull-stripped brain images with and without FCD lesions. The characteristics of FCD lesions in FLAIR MR images include cortical thickening, blurring of white matter-grey matter junction, altered signal intensity from white matter with or without the penetration through cortex (transmantle sign), altered signal from gray matter, abnormal sulcal or gyral pattern and segmental and/or lobar hypoplasia/atrophy (FCD, 2020; Kabat and Król, 2012; Antel *et al.*, 2002). Since the manual tracing of these features is laborious and challenging, automated segmentation methods can be of great assistance to neuroradiologists in accurately examining the lesion regions with reduced time.

Brain disorders are usually associated with structural changes in the brain, and thus the brain symmetry analysis helps in detecting these diseases (Crow *et al.*, 1989). However, structural asymmetry can also be detected in a healthy brain, and hence it cannot be considered a reliable measure for identifying FCD regions. Volume-based features such as skewness and kurtosis of the cortical thickness, blurring in white matter-grey matter junction, intensity transitions of the voxels, orientation of the gradient vectors, etc., are also proven to be very useful in detecting FCD lesions, as they enable the analysis of distributional characteristics, formed by a set of neighboring voxels (Yang *et al.*, 2011; Feng *et al.*, 2020*a*; Jin *et al.*, 2018). However, assessment of these features using mathematical models and handcrafted feature extraction methods is quite challenging due to several external factors such as contrast variations and noise.

A complex diffusion-based method was proposed by Rajan et al. (Rajan *et al.*, 2009) for improving the visibility of the FCD-affected regions in T1-weighted brain MR images. While this approach can highlight most potential FCD lesions, it also detects several non-FCD regions, resulting in low precision rates. Bergo et al. (Bergo *et al.*, 2008) proposed



(a) (b)

Figure 3.1: Sample brain MRI scans with and without FCD (FCD regions are highlighted). (a) MRI slices without FCD lesion and (b) MRI slices with FCD lesion.

another approach for detecting FCD lesions by comparing the features of voxels from opposing brain hemispheres using a Reduced Coulomb Energy (RCE) classifier. Adler et al.(Adler *et al.*, 2017) utilized surface-based features to identify focal abnormalities in the pediatric cohort of cortical development. The need for symmetric brain templates and the inability to extract complex lesion features are the main drawbacks of this method. Colliot et al. (Colliot *et al.*, 2005, 2006) proposed an FCD segmentation algorithm using a 3D deformable model and a level set framework, mainly driven by the known MR characteristics of FCD. One main limitation of this approach is that only MRI-visible FCD lesions with high-intensity values can be segmented, rather than the entire spectrum of FCD.

Ahmed et al. (Ahmed *et al.*, 2015) proposed a morphometry-based approach to detect MRI-negative FCD lesions. This method requires accurate co-registration of the pathological sample with the MRI, allowing for the matching of MRI and pathological slice. Wang et al. (Wang *et al.*, 2018) proposed an FCD detection approach, which used voxel-based multi-modal features from diffusion tensor MRI and T2 weighted MRI. Tan et al. (Tan *et al.*, 2018) proposed an FCD detection and segmentation approach which used multi-modal features. Their model used a linear SVM classifier to assess the lesion with handcrafted features, and hence, the detection performance is highly influenced by the nature and quality of the input images.

Azami et al. (El Azami *et al.*, 2013) proposed a voxel-wise analysis using T1-weighted MR images, considering features such as cortical thickness and gray matter-white matter characteristics, using One-Class SVM. Focke et al. (Focke *et al.*, 2008) also reported a voxel-based FCD detection with 3D FLAIR images. In this process, various mathematical models were employed to restrict the region of interest to specific locations such as white matter-grey matter junctions. These aforementioned models use handcrafted features that are highly sensitive to image type, nature of noise, and contrast variations. Nevertheless, the main disadvantage of these approaches is the lack of semantic analysis.

Several neural network-based methods were also proposed for detecting and localizing the FCD lesions. Mo et al. (Mo *et al.*, 2019) proposed a model that requires inputs from multiple MRI scans, such as T1, T2-FLAIR, and PET images, which may not be practical in all scenarios. With the recently emerged Fully Convolutional Neural Networks (FCN) and deep learning concepts, image segmentation capability has dramatically improved and contributes much to the process of automated medical image segmentation. Feng et al. proposed a deep four-layer CNN model for FCD localization in (Feng *et al.*, 2020*c*). This model focused on screening MR images for detecting dysplasia regions. Although this method detects the frames with dysplasia lesions, it does not provide any information regarding the location or extent of the FCD regions. Gill et al. (Gill *et al.*, 2018, 2019) introduced patch-wise CNN models designed for FCD segmentation. Due to its reliance

on patch-wise analysis, these approaches are limited in their ability to capture global contextual information from the entire brain area. Dev et al. (Dev *et al.*, 2019) proposed a deep learning CNN-based FCD segmentation on FLAIR MR images, which combines both 3T and 1.5T data for detecting and quantifying FCD lesions. The methodology uses 2D MRI slices and does not consider the inter-slice information among the MRI volumes. Very recently, Thomas et al. (Thomas *et al.*, 2020) proposed a hybrid skip connection-based FCD segmentation model with a modified U-Net architecture, reinforced with Multi-residual convolution blocks and attention modules, using 3T FLAIR MR images. Though this method shows the performance of FCD segmentation, it does not utilize the inter-slice information from 3D MRI volumes.

In this study, we propose a 3D CNN segmentation model integrated with residual connections for FCD lesion detection and segmentation from 3D MRI volumes. The main contributions of this work can be summarised as follows.

1. shallow sliced stacking approach is proposed to reduce the size of input 3D objects so as to maintain a good segmentation accuracy with minimum computation overhead and model complexity.

2. A customized 3D U-Net architecture is proposed by integrating residual connections in the encoder path, which helps in extracting multi-scale features with a relatively smaller number of parameters.

## 3.2 Methods

The proposed methodology consists of three main stages: pre-processing, segmentation, and post-processing. Figure 3.2 shows the block-level representation of the training and testing stages in the proposed methodology.

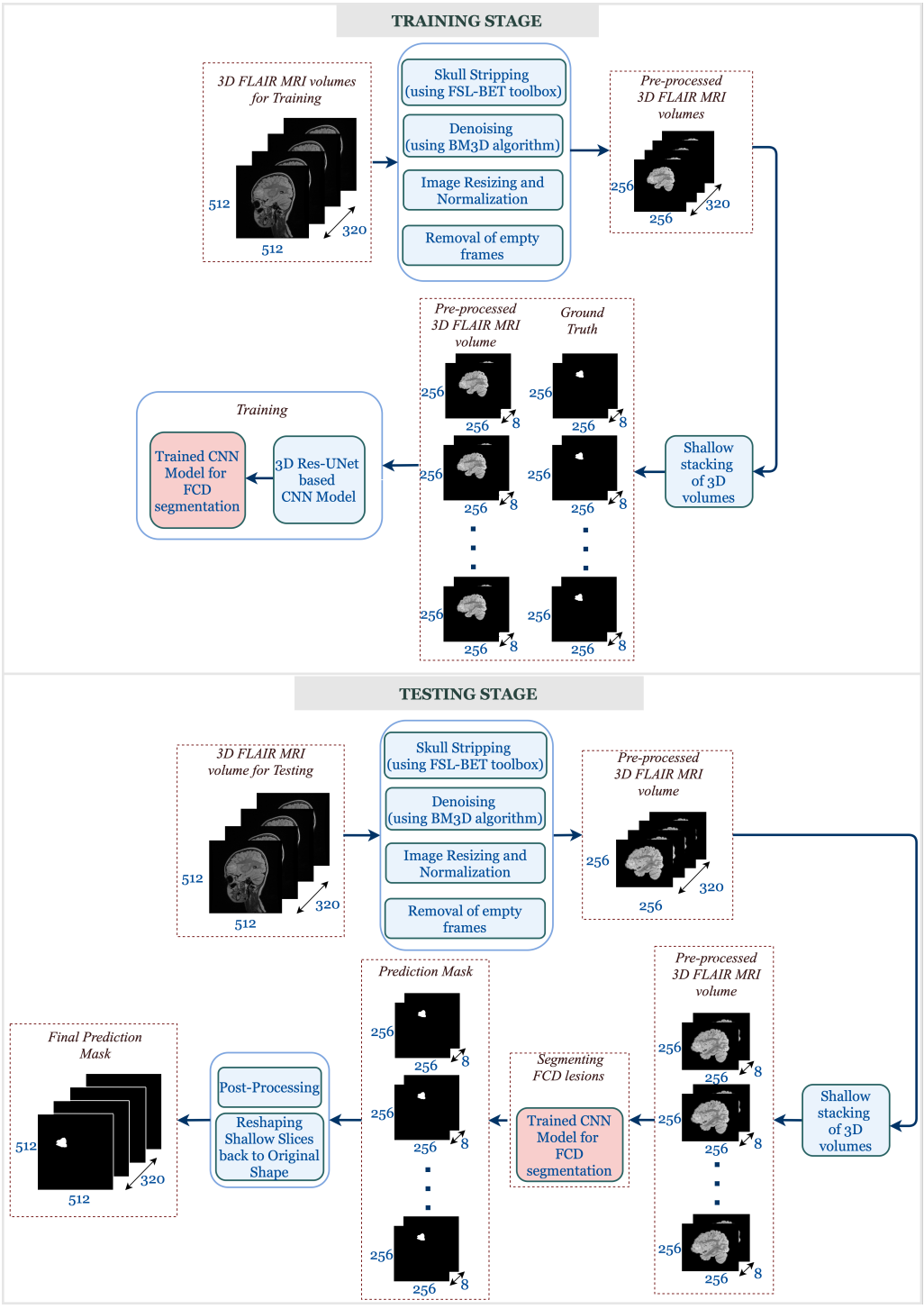



Figure 3.2: Schematic diagram showing the training and testing stages in the proposed methodology.

### 3.2.1 Preprocessing

The acquired MR images are corrupted by noise, and the Block-Matching and 3D Filtering (BM3D) algorithm is employed to remove the noise from the MRI volume (Maggioni *et al.*, 2012). The general features of FCD include cortical thickening, blurring of white matter-grey matter junction, etc. FSL-BET toolbox (Smith, 2002) is used for skull-stripping operation so as to restrict the scope to the brain area. The proposed segmentation architecture is designed for input volumes of size $256 \times 256 \times k$, where $k$ is the number of slices in the volume. In this study, $k$ is set to 8. So, the input MRI slices need to be resized to the order of $256 \times 256$. All slices are also standardized using Z-score normalization (Karpathy *et al.*, 2016) to generate slices with zero mean and unit variance. A few frames at the start and end position of each MRI volume are also removed as the frames are devoid of any useful information. The removal of such frames is done automatically based on the number of brain region pixels in each slice.

### 3.2.2 FCD Segmentation using the Proposed 3D CNN

The proposed model uses a depth-3 3D encoder-decoder architecture, where the residual blocks are used instead of normal convolution blocks in the encoder part. The proposed model addresses several drawbacks of the conventional 3D CNN by customizing the network architecture. The specific customizations integrated into the proposed architecture are detailed in the following subsections.

### 3.2.3 Network Architecture

The proposed architecture is depicted in Figure 3.3. Higher computational cost, extra memory, and the need for a large number of 3D training samples are the main drawbacks of the traditional 3D CNN models. This motivated us to design a 3D CNN-based segmentation method that can extract the inter-slice features by minimizing the aforementioned

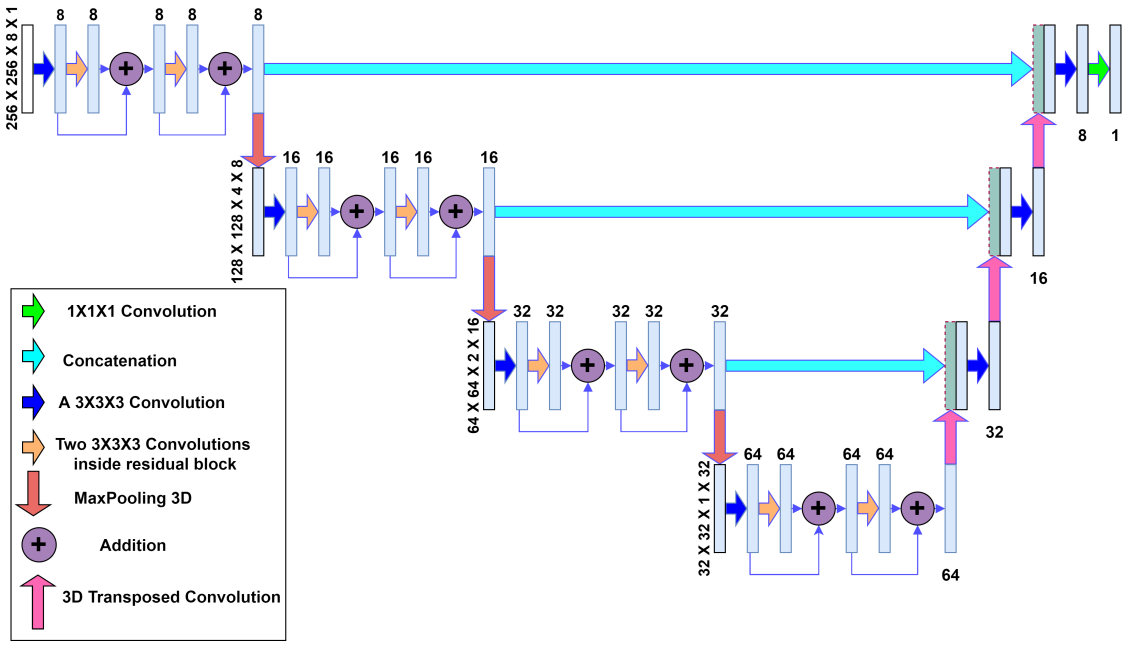



Figure 3.3: Proposed 3D CNN model for FCD segmentation.

drawbacks. The main design considerations of the proposed architecture are listed below.

1. *Shallow Slice Stacking for reducing computation cost and data augmentation*

In medical imaging, often several image acquisition techniques such as MRI, CT, etc., are used to scan the ROI for generating closely stacked 2D images with low inter-slice spacing and hence considered as 3D volumes. These images help to retain the continuity in scanning by preserving the inter-slice information and hence provide the volumetric analysis of the abnormalities that spread over multiple slices. Consider a 3D CNN model with an input 3D object of size $m \times n \times k$, and $N1$ number of 3D filters, having a kernel of size $h \times h \times h$ in the first convolution layer $L1$. Considering the same convolution, the total number of parameters in the first convolution layer is $h^3 \times N1$, the number of multiplications is $h^3 \times m \times n \times k \times N1$, and the output feature space is of size $m \times n \times k \times N1$. Assuming $N2$ number of filters in the second convolution layer $L2$, the number of parameters, multiplications, and feature space dimensions will be set as $h^3 \times N1 \times N2$, $h^3 \times N1 \times m \times n \times k \times N2$ and $m \times n \times k \times N2$, respectively. Hence, the computation overhead, as well as the memory usage, can be controlled by adjusting the

above factors. Since $m$ and $n$ are indicative of the number of rows and columns in the input data, reducing them may result in the loss of information. The kernel size $h$ and the number of filters used in each layer $N1$, $N2$, etc. are the next most contributing factors. However, these factors (convolution kernel size and number of filters in the convolution layers) are usually designed to extract the features with an optimal bias-variance trade-off. Hence, in order to reduce the computation cost and model complexity, we focused on the parameter $k$, which represents the number of slices in the input object.

In the proposed model, we used shallow stacking of MRI slices to limit the depth of input data samples by dividing a single MRI volume into multiple overlapped shallow 3D sub-volumes. For example, if a single MRI volume is of the size $m \times n \times K$, then it can be split into multiple overlapping $m \times n \times k$ sub-volumes (where $k < K$). A window of size $m \times n \times k$ with stride $S$ along the depth (z-direction) is used to generate the samples. The number of 3D sub volumes per MRI volume is given as:

$$l = floor\left(\left(\frac{K-k}{S}\right) + 1\right) \tag{3.1}$$

In Equation (3.1), $l$ is the number of sub-volumes, $k$ is the number of slices per sub-volume, $K$ is the number of slices in the actual MRI volume, and $S$ is the stride used for overlapping. Thus, the number of data samples generated can be controlled by adjusting the stride used for sampling the slices. This results in performance improvements such as:

(a) Reduction in the computation complexity and memory overhead due to the reduced depth of input volumes.

(b) Creation of a sufficient number of 3D data samples for training without extensive data augmentation.

(c) Retaining the pixel continuity in two dimensions by 3D shallow slices and hence preserving the global features.

For the final selection of hyper-parameters like the number of slices per sub-volume, stride for overlapping etc, we conducted several experiments with different depths and

stride values. The memory requirement and execution time required for training the model are highly correlated to the depth of the shallow slices. For instance, the time and GPU memory required for training one epoch with an input object size of $256 \times 256 \times 8$ is 16 seconds and 0.55 GB, respectively. In the case of input objects with size $256 \times 256 \times 32$, it is 81 seconds and 2.19 GB. Similarly, while considering input with size $256 \times 256 \times 128$, it takes nearly 809 seconds to complete one epoch and consumes 8.75 GB of GPU memory (refer Table 3.4). On the other hand, the segmentation performance of input objects with depth 8 is nearly the same as with depth 16, 32, or 64. Experiments were also conducted with a lower depth of 4, which shows inferior results than those with higher depths. The stride value for overlapping is another relevant parameter that controls the number of generated 3D sub-volumes. We conducted experiments with different stride values, such as 1, 3, and 5, and better segmentation performance was observed with lesser stride values due to the generation of more training data. Hence, the depth and stride of the input shallow slices are selected as 8 and 1, respectively.

2. *Integration of ResNet blocks in the encoder layers for multi-scale feature extraction*

In the original 3D U-Net architecture, a sequence of two convolution layers is used in each encoder depth, and both layers use the same number of filters. The main drawbacks of this sequential convolution are the quadratic effect in computation cost and the lack of multi-scale feature extraction. This can be demonstrated with an example. Consider the case with a data sample of size $m \times n \times k$ and a 3D kernel of size $h \times h \times h$. Suppose $N$ number of filters are used in each convolution layer. Then, the number of parameters and multiplications required in the first and second layers are $h^3 \times N, h^3 \times m \times n \times k \times N$, and $h^3 \times N^2, h^3 \times m \times n \times k \times N^2$, respectively. In general, the number of parameters and multiplications required up to the $l^{th}$ layer of 3D convolution are given as $NP$ and $NM$, respectively, in Equation (3.2) and Equation (3.3).

$$NP = h^3 \times \left(N + (l-1)N^2\right) \tag{3.2}$$

$$NM = h^3 \times m \times n \times k \times \left(N + (l-1)N^2\right) \tag{3.3}$$

From these equations, it is clear that the requirement of memory usage and number of computations are in the order of $N^2$. The residual connections are capable of easing the training process using identity mappings. The identity mappings in the residual blocks effectively simplify the network, using fewer layers in the initial stages, thus resulting in faster convergence. Therefore, in each encoder depth, we used a combination of a convolution layer followed by two residual blocks, having two convolutions with $\frac{N}{2}$ filters each, thus reducing the memory usage and complexity to the order of $(\frac{N}{2})^2$. The incorporation of ResNet blocks into our architecture has two advantages in addition to the inherent advantages of skip connections. They are listed below.

(a) *Reduction in the computation cost and memory overhead*

In the case of a 2-layer convolution block with $N$ filters in each layer, the number of parameters and computation cost will be affected by a factor of $N + N^2$. Alternatively, if we use one convolution block followed by two ResNet blocks with $\frac{N}{2}$ filters in each layer, the factor affecting the number of parameters and computation cost is $\frac{N}{2} + 4 \times (\frac{N}{2})^2$. This shows the reduction in computation cost, number of parameters, and memory usage while replacing UNet convolution blocks with Residual-UNet blocks.

(b) *Extraction of multi-scale features*

Since all the kernels in the ResNet blocks are of size $3 \times 3 \times 3$, the use of the proposed architecture results in four convolution layers in each encoder level. Hence, the output of the second, third, fourth, and fifth convolutions cover an ROI equivalent to $5 \times 5 \times 5$, $7 \times 7 \times 7$, $9 \times 9 \times 9$ and $11 \times 11 \times 11$, respectively. It eventually extracts features in four different scales without any additional computations or memory requirements. This helps the model to achieve better performance with reduced depths of encoder-decoder-based CNN segmentation models.

In the encoding path, each convolutional block consists of a 3D convolution layer followed by *batch normalization* (Ioffe and Szegedy, 2015) and *ReLU* activation (Nair and

Hinton, 2010). Similarly, residual blocks are also followed by *batch normalization* and the $ReLU$ activation function. Since the main advantage of the residual blocks is to extract multi-scale features, we integrated the Res-blocks only in the encoding path of the CNN architecture. The sequence of residual blocks extracts multi-scale characteristics from the input feature space, following which, a $2 \times 2 \times 2$ max-pooling operation is performed. The 3D max-pooling will reduce the number of rows, columns, and slices of the output into half of the input feature space. The number of filters in the subsequent layers is doubled to account for the loss of spatial information by max-pooling. The depth of the encoder stage can be extended up to $\log_2 k$, where k is the number of slices in the input object.

The proposed architecture is of depth-3 and is designed for input samples of size $256 \times 256 \times 8$. In depth-0, all convolutions (including the convolutions in the ResNet blocks) use eight filters each, and the number of filters is doubled in subsequent depths. Hence, the encoder stages in depth-1, depth-2, and depth-3 take 16, 32, and 64 filters in each convolution block. In the expanding path, the spatial dimension of the feature map is increased by a transposed convolution operation. The high-resolution feature maps from the contracting path are concatenated to the upsampled resultant image. Dropout and regularization are also integrated into the architecture to avoid overfitting. As the output of the model is binary with only one output neuron, the sigmoid activation function is used in the last layer.

The dataset used in this work is imbalanced, and the number of lesion voxels is less than 5% of the negative voxels in the 3D samples with FCD lesions. Therefore, training on this dataset with popular loss functions such as Dice loss or a combination of Binary Cross-Entropy (BCE) and Dice loss leads to predictions that are severely biased towards high Precision and low Recall. Hence, we used the Tversky loss function (Salehi *et al.*, 2017) to eliminate this class imbalance problem to an extent. Tversky loss function appears to perform better with class-imbalanced datasets. The equations for Dice Similarity Coefficient (DSC) and Dice loss (Dice, 1945) are given in Equations 3.4 and 3.5. Tversky similarity index is a generalized representation of the DSC and is defined in Equations

3.6 and 3.7.

$$\text{DSC} = \frac{2 \times TP}{(2 \times TP) + FP + FN} \tag{3.4}$$

$$\text{Dice Loss} = 1 - \text{DSC} \tag{3.5}$$

$$\text{Tversky similarity index} = \frac{TP}{TP + (\alpha \times FP) + (\beta \times FN)} \tag{3.6}$$

$$\text{Tversky Loss} = 1 - \text{Tversky similarity index} \tag{3.7}$$

where $TP$ is *true positives*, $FP$ is *false positives*, $FN$ is *false negatives*, and $\alpha$ and $\beta$ decide the magnitude of penalties for $FP$ and $FN$, respectively.

When $\alpha = \beta = 0.5$, the penalty for $FP$ and $FN$ will be equal, and in this case, the Tversky similarity index becomes identical to that of DSC. In our experiments with loss functions such as BCE loss and the Dice loss, it was observed that the Precision values are significantly higher than that of the Recall values since the ratio of voxels in the FCD region to the normal brain region is very high. It will also contribute to the overall class imbalance and limit the model from detecting many of the FCD voxels. While using Tversky loss, $FN$ can be weighted with higher penalties and improves the Recall rate to a sufficient range. In the proposed architecture, we conducted experiments with different combinations of Tversky loss parameters and empirically selected the values of $\alpha$ and $\beta$ as 0.25 and 0.75, respectively. All other hyperparameters are fine-tuned based on the trial-and-error approach.

### 3.2.4 Post Processing

The segmentation masks obtained using CNN models may have a few false detections due to the noise elements in the input images. Hence, a post-processing stage is also applied to the prediction mask obtained from the 3D CNN output. A morphological opening operation with a disk-shaped structural element of pixel radius 3 followed by

a morphological hole-filling operation is used in the post-processing stage. The post-processing reduces the noisy pixels in the prediction mask, smoothens the lesion boundary regions, and eliminates small holes in the prediction mask. Finally, the binary predicted masks corresponding to the shallow stacked slices of each test volume are combined to generate the final 3D prediction mask with respect to the original MRI volume.

## 3.3 Results and Analysis

This section presents the hardware details, evaluation metrics, ablation study, and discussion involving qualitative and quantitative analysis.

### 3.3.1 Hardware Details

All experiments were conducted on NVIDIA® DGX-1® machine loaded with Canonical Ubuntu OS, Dual 20-Core Intel® Xeon E5-2698 v4 CPU @2.2 GHz, 512 GB of RAM, and 8X NVIDIA® Tesla® V100 GPU with 32GB dedicated memory. Keras and TensorFlow libraries of Python are used to implement our model.

### 3.3.2 Evaluation Metrics

The Precision, Recall, and DSC are used as metrics for quantitative analysis. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. Recall is the ratio of correctly predicted positive observations to all observations in the actual class, and the DSC is the weighted average of precision and recall. DSC takes both *FP* and *FN* into account and is a popular evaluation benchmark in analyzing the overlapping between the predicted result and the ground truth. The mathematical representation of Precision, Recall, and DSC are given in Equations 3.8, 3.9, and 3.10, respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.9}$$

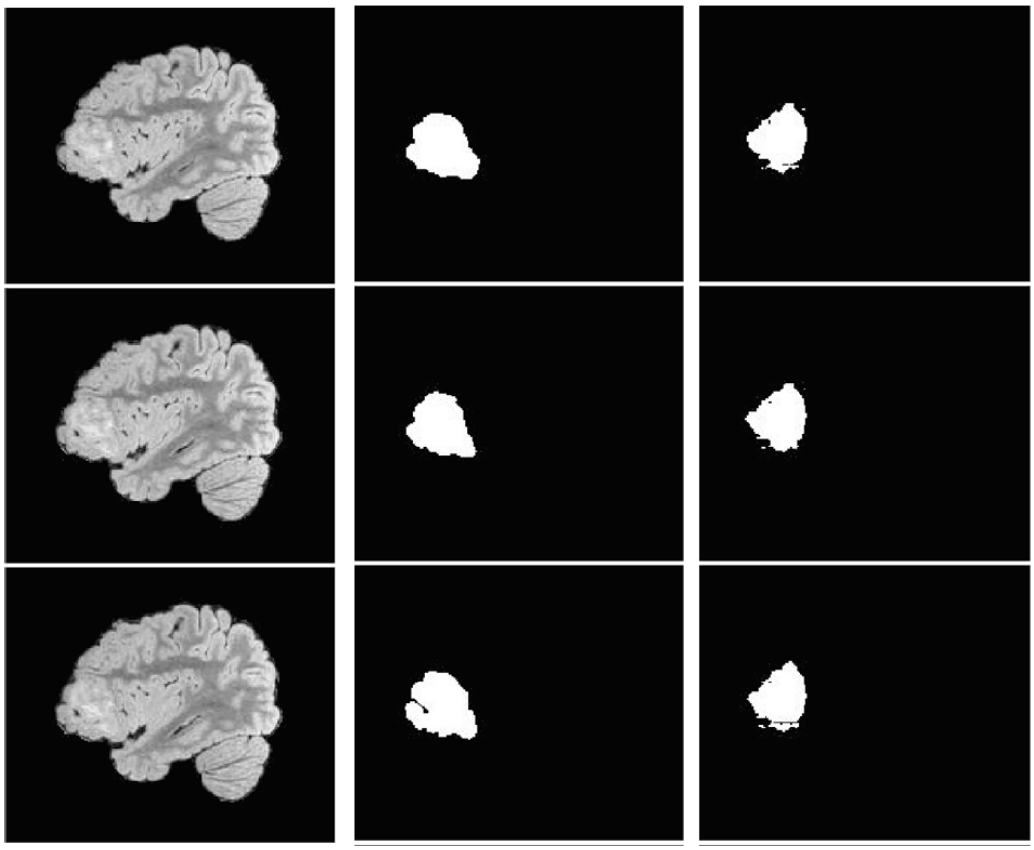
$$\text{DSC} = \frac{2 \times TP}{(2 \times TP) + FP + FN} \tag{3.10}$$

### 3.3.3 Datasets

The dataset used in this work is collected from Sree Chitra Tirunal Institute for Medical Sciences and Technology (SCTIMST), Trivandrum, India. This study was conducted with the approval of the institutional ethics committee of SCTIMST (No. IEC/1073). The dataset consists of 3D FLAIR brain MR images collected from 26 patients. Images were acquired in the sagittal plane on a 3T scanner (GE Healthcare, UK) with a slice thickness of 1 mm and pixel spacing of 0.5 mm per patient. The TR/TE/TI/flip angle used was $7200ms/117.241ms/1936ms/90°$, respectively. Each volume is of size $512 \times 512 \times 320$. For the experiments, these volumes were pre-processed and resized to $256 \times 256 \times 320$ resolution to avoid redundant details and to reduce the training time.

### 3.3.4 Training methodology

Since each MRI volume consists of a few blank frames in the initial and final locations, such frames were initially filtered out using an automatic thresholding process over the number of brain pixels in the image. We only considered the frames, which had at least 10% of brain pixels. For experiments, we opted for 5-fold cross-validation to avoid statistical uncertainty and bias. To create the train, test, and validation sets, we used random indexing of patients by randomly selecting 18, 5, and 3 patients, respectively. While creating the folds, it was ensured that no common patient exists among train, test, and validation splits. Since we need uniformly sized 3D sub-volumes to feed the model input layer of size $256 \times 256 \times 8$, the training data and validation data are created



(a) (b) (c)

Figure 3.4: Qualitative analysis of the proposed model, (a) Brain MRI scans with FCD, (b) Ground truth, (c) Predicted output of the proposed 3D Res-UNet.

as overlapping 3D objects of shape $256 \times 256 \times 8$, as explained in Section 3.2.3, with a stride of one voxel. Since there is no need for augmenting the test data, the test set is created with non-overlapping $256 \times 256 \times 8$ sub-volumes from the MRI test volumes. During training, a limited offline augmentation is used to increase the number of training samples three times that of the actual training data.

Hyper-parameters such as the number of Res-blocks in each level, the number of filters in convolution layers, etc., were decided empirically based on the results of several experiments. Adam optimizer (Kingma and Ba, 2014) with Adagrad algorithm (Reddi *et al.*, 2018) was used to update the weights. The learning rate used was 0.001, with a batch size of 16. Weights of the network are initialized using *Xavier* initialization (Glorot

and Bengio, 2010). The weight $W$ is initialized as in Equation 3.11,

$$W \in U\left[-\frac{\sqrt{6}}{\sqrt{c_{input}+c_{output}}}, \frac{\sqrt{6}}{\sqrt{c_{input}+c_{output}}}\right] \tag{3.11}$$

where $U[-a, a]$ is the uniform distribution in the interval $(-a, a)$ and $c_{input}$ and $c_{output}$ are the input and output sizes of the corresponding convolutional layer. L2 regularization is used in all convolutional and transposed convolutional layers with $l2 = 0.01$. The added regularization penalty is computed as $loss = l2 * reduce\_sum(square(x))$, where $x$ is the weight tensor. The proposed model has 808,329 parameters in total, out of which 807,257 are trainable parameters. The training is limited to a maximum of 500 epochs with an early stopping after a patience of 100 epochs. During the training process, the best weights are saved based on the best validation DSC obtained after each epoch.

### 3.3.5 Results and Discussion

The performance of the proposed method is analyzed qualitatively and quantitatively and is compared with two state-of-the-art FCD segmentation models given in (Dev *et al.*, 2019) and (Thomas *et al.*, 2020). We used a predicted segmentation mask for qualitative analysis, and this is shown in Figure 3.4. It is evident that the predictions are very close to the ground truth, thus correctly locating and segmenting the FCD lesions. Based on the visual analysis conducted over many test results, it can be concluded that the proposed model segments the FCD lesion regions with sufficient accuracy.

Five-fold cross-validation is used in the quantitative evaluation, which includes pixel-wise, region-wise, and patient-wise analysis. In pixel-wise analysis, we evaluated the performance of the model to discriminate FCD pixels from non-FCD pixels. In region-wise analysis, the evaluation is based on how well the model is detecting the presence of an FCD lesion in an MRI slice. Similarly, in patient-wise analysis, the evaluation is based on how well the model is detecting whether an MRI volume has an FCD lesion or not. The reported results are the average of three 5-fold cross-validations. Table 3.1 represents

Table 3.1: Pixel-wise performance comparison of the benchmark models and the proposed model. The results of the 5-fold evaluation are presented in terms of Precision, Recall, and DSC.

| **Folds** | **UNet (Dev *et al.*, 2019)** | | | **MultiRes Attention-UNet (Thomas *et al.*, 2020)** | | | **Proposed 3D Model** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **DSC** | **Precision** | **Recall** | **DSC** | **Precision** | **Recall** | **DSC** | **Precision** | **Recall** |
| **Fold 1** | 58.14 | 59.87 | 56.96 | 62.55 | 59.91 | 65.63 | 69.53 | 69.91 | 69.17 |
| **Fold 2** | 42.44 | 72.06 | 30.67 | 45.67 | 73.67 | 33.12 | 46.75 | 66.13 | 36.16 |
| **Fold 3** | 69.34 | 74.12 | 66.46 | 72.54 | 68.65 | 77.00 | 73.35 | 67.80 | 79.89 |
| **Fold 4** | 70.77 | 71.42 | 70.67 | 71.49 | 66.72 | 77.33 | 72.09 | 70.81 | 73.42 |
| **Fold 5** | 57.32 | 74.85 | 47.42 | 56.81 | 71.57 | 48.05 | 59.89 | 73.25 | 50.66 |
| **Average** | **59.64** | **70.46** | **54.44** | **61.81** | **68.10** | **60.23** | **64.32** | **69.58** | **61.86** |

Table 3.2: Region-wise results of the benchmark models and the proposed model.

| **Folds** | **UNet (Dev *et al.*, 2019)** | | | **MultiRes Attention-UNet (Thomas *et al.*, 2020)** | | | **Proposed 3D Model** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **DSC** | **Precision** | **Recall** | **DSC** | **Precision** | **Recall** | **DSC** | **Precision** | **Recall** |
| **Fold 1** | 63.57 | 91.44 | 48.86 | 77.91 | 94.00 | 66.58 | 78.65 | 78.00 | 79.32 |
| **Fold 2** | 44.25 | 78.50 | 31.61 | 57.04 | 74.21 | 46.79 | 59.44 | 55.73 | 63.69 |
| **Fold 3** | 85.54 | 95.54 | 77.50 | 84.81 | 90.47 | 79.84 | 81.11 | 76.39 | 86.46 |
| **Fold 4** | 88.68 | 96.86 | 82.10 | 87.27 | 93.25 | 82.10 | 86.01 | 81.61 | 90.91 |
| **Fold 5** | 75.60 | 97.26 | 62.35 | 71.05 | 87.92 | 60.12 | 77.28 | 82.02 | 73.05 |
| **Average** | **71.53** | **91.92** | **60.48** | **76.62** | **87.97** | **67.09** | **76.67** | **74.75** | **78.69** |

the pixel-wise results for the segmentation of FCD lesions and shows a comparison with the state-of-the-art FCD segmentation approaches: 2D UNet (Dev *et al.*, 2019) and MultiRes-Attention UNet (Thomas *et al.*, 2020). The proposed method detects more boundary pixels in the FCD lesion and leads to an improved Recall rate in comparison with 2D UNet and 2D MultiRes-Attention UNet models.

Table 3.2 presents the region-wise analysis, where the proposed model shows superior

Table 3.3: Patient-wise results of the benchmark models and the proposed model.

| **Folds** | **Recall** | | |
|---|---|---|---|
| | **UNet (Dev *et al.*, 2019)** | **MultiRes-Attention UNet (Thomas *et al.*, 2020)** | **Proposed 3D Model** |
| **Fold 1** | 60.0 | 90.0 | 90.0 |
| **Fold 2** | 40.0 | 75.0 | 80.0 |
| **Fold 3** | 100.0 | 100.0 | 100.0 |
| **Fold 4** | 100.0 | 100.0 | 100.0 |
| **Fold 5** | 95.0 | 95.0 | 95.0 |
| **Average** | **79.0** | **92.0** | **93.0** |

FCD detection performance (in terms of Recall rate) than the 2D UNet models. However, the region-wise precision value of the proposed model is lesser compared to 2D U-Net models due to comparatively higher false detection in the proposed model. Patient-wise detection performance is also evaluated and compared with the 2D UNet approaches, and the results are shown in Table 3.3. The dataset used in the evaluation process consists of MRI volumes from 26 patients confirmed with FCD. The FCD lesions are present only over a small region in a few frames out of the 320 frames per patient. So, there are both positive and negative pixels per frame and both positive and negative frames per patient. Hence, we include the performance metrics: Precision, Recall, and DSC in both pixel-wise analysis and region-wise analysis. Since there are no negative patients (without any FCD lesion in the entire MRI volume), no false alarms will be there, and the chance of getting a *FP* is always zero in the patient-wise analysis. When *FP* is zero, the Precision will always give 100% (refer Equation 3.8), irrespective of the segmentation performance. Hence, Precision analysis is irrelevant in the patient-wise evaluation. Since DSC is the harmonic mean of the Precision and Recall values, it will also be misleading in the patient-wise analysis. Hence, the Precision and DSC evaluation is excluded from the patient-wise analysis.

Performance of the proposed model is also evaluated in terms of the number of train-

Table 3.4: Performance comparison of the proposed model with 3D UNet with respect to different shall slice depths, in terms of trainable parameters and computation complexity ( $T_N$, $T_T$, and $T_M$ represents the number of trainable parameters(in million), training time per epoch (in seconds), and the memory required for training (in GB), respectively).

| **Models** | **Input shape (256x256x128)** | | | **Input shape (256x256x32)** | | | **Input shape (256x256x8)** | | |
|---|---|---|---|---|---|---|---|---|---|
| | $T_N$ | $T_T$ | $T_M$ | $T_N$ | $T_T$ | $T_M$ | $T_N$ | $T_T$ | $T_M$ |
| **3D UNet** | 1.658 | 824 | 8.9 | 1.658 | 84 | 2.22 | 1.658 | 17 | 0.56 |
| **Proposed 3D Model** | 0.807 | 809 | 8.75 | 0.807 | 81 | 2.19 | 0.807 | 16 | 0.55 |

Table 3.5: Performance comparison of the proposed method with state-of-the-art approaches, in terms of number of trainable parameters ($T_N$), training time per epoch $T_T$ and GPU memory required $T_M$.

| **CNN Models** | **Benchmark Metrics** | | |
|---|---|---|---|
| | $T_N$ **(in Million)** | $T_T$ **(in Seconds)** | $T_M$ **(in GB)** |
| **UNet (Dev *et al.*, 2019)** | 1.926 | 36 | 0.168 |
| **MultiRes-Attention UNet (Thomas *et al.*, 2020)** | 1.036 | 58 | 0.154 |
| **3D UNet** | 1.658 | 17 | 0.56 |
| **Proposed 3D model** | 0.807 | 16 | 0.55 |

able parameters ($T_N$), Training time $T_T$ and GPU memory required $T_M$ and are shown in Table 3.4 and Table 3.5. Table 3.4 shows the computational cost when 3D models are used with different 3D input shapes. Table 3.5 shows a similar comparison of the proposed method with respect to the state-of-the-art approaches.

Table 3.6: Pixel-wise performance Comparison between 3D U-Net (with normal convolution layers) and the proposed 3D model. The results of 5-fold evaluation is presented in terms of Precision, Recall and DSC.

| **Folds** | **3D UNet** | | | **Proposed 3D Model** | | |
|---|---|---|---|---|---|---|
| | **DSC** | **Precision** | **Recall** | **DSC** | **Precision** | **Recall** |
| **Fold 1** | 42.35 | 47.34 | 39.17 | 69.53 | 69.91 | 69.17 |
| **Fold 2** | 36.17 | 58.16 | 27.53 | 46.75 | 66.13 | 36.16 |
| **Fold 3** | 58.59 | 62.14 | 55.57 | 73.35 | 67.80 | 79.89 |
| **Fold 4** | 58.15 | 67.846 | 51.825 | 72.09 | 70.81 | 73.42 |
| **Fold 5** | 37.10 | 65.29 | 25.99 | 59.89 | 73.25 | 50.66 |
| **Average** | **46.47** | **60.28** | **40.02** | **64.32** | **69.58** | **61.86** |

The pixel-wise analysis in Table 3.1 shows a significantly higher Recall rate for the proposed model in comparison with the 2D UNet model presented in (Dev *et al.*, 2019), while the Precision is slightly lesser for the proposed methodology. Since the study focuses on screening the brain MRI for finding FCD lesions (to assist radiologists), the primary objective is to reduce the 'Miss Rate.' Hence, more emphasis is given to improving the Recall rate rather than the Precision performance, and several contributing factors in the proposed methodology, such as loss function and the augmented training data, are fine-tuned to provide such a Precision-Recall trade-off. While comparing with the MultiRes-Attention UNet (Thomas *et al.*, 2020), the pixel-wise performance is higher for the proposed method in terms of all benchmark metrics. The inter-slice feature extraction using 3D convolution layers is the foremost reason behind the performance improvement of the proposed method in comparison with state-of-the-art 2D FCD segmentation models.

Patient-wise FCD segmentation performance (in Table 3.3) also shows the advantage of the proposed model, which detects several true FCD frames from the MRI volumes in comparison with the 2D approaches. The generation of large 3D sub-volumes using the shallow sliced stacking approach and the adaptive augmentation of the training data (selection of positive and negative 3D sub-volumes in the ratio 1:3) and Tversky loss function contribute to reducing the class imbalance between the FCD and non-FCD voxels. While analyzing the segmentation performance in various perspectives, the variance

observed among different folds is also reported in the tables (Table 3.1, 3.2, 3.3, and 3.6). The proposed method also shows lesser variance in comparison with other approaches due to more generalized learning from the training data.

Since the proposed method uses less number of filters in comparison with state-of-the-art 2D CNN methods, the number of trainable parameters and the training time are relatively less in the proposed model (in Table 3.4 and Table 3.5). It can be clearly seen that the GPU memory usage and training time are increasing with the increase in the depth of the 3D input data. The optimal depth (here 8) is selected based on these observations as well as the segmentation performance obtained in those different cases. However, the proposed model requires more memory in comparison with 2D models to save the intermediate 4D feature spaces while training the model. The proposed method also shows a significant reduction in the number of trainable parameters and marginal improvement in the training time and GPU memory usage when compared with the 3D UNet model.

While designing the proposed 3D architecture, the basic 3D UNet and its several variants were considered for conducting the preliminary experiments. Several models/combinations were evaluated by integrating deep learning convolution structures, such as ResNet, DenseNet, etc., and the final model was decided based on the observations

Table 3.7: Pixel-wise performance of the proposed method on BRATS 2015 Dataset (Menze *et al.*, 2014) (for brain tumor segmentation). The results of the 5-fold evaluation are presented in terms of Precision, Recall, and DSC.

| **Folds** | **Proposed 3D Model** | | |
|---|---|---|---|
| | **DSC** | **Precision** | **Recall** |
| **Fold 1** | 82.55 | 75.22 | 91.47 |
| **Fold 2** | 83.75 | 81.80 | 85.81 |
| **Fold 3** | 84.21 | 81.81 | 86.76 |
| **Fold 4** | 83.54 | 79.13 | 88.48 |
| **Fold 5** | 82.22 | 81.71 | 82.74 |
| **Average** | **83.25** | **79.93** | **87.05** |

in the ablation study. Hyperparameters such as the number of Res-blocks in each level, the number of filters in Res-blocks, etc., are also decided based on the results of different experiments. For instance, the results of an ablation study with 3D U-Net (without residual blocks) and 3D Res U-Net (with residual blocks) are shown in Table 3.6. It shows the significance of residual modules in the proposed architecture. The multi-scale feature extraction and the formation of multiple learnable paths favor improved results while using the residual blocks.

Even though the proposed architecture is designed and fine-tuned for FCD segmentation, we tried it on the publicly available BRATS 2015 dataset (Menze *et al.*, 2014) (for Brain tumor segmentation), and the results are shown in Table 3.7. This experiment additionally demonstrates that the proposed model can also be used for other similar segmentation problems. The performance of the model on this dataset can be further improved by fine-tuning the hyperparameters. Though several design elements in the proposed model help reduce class imbalance up to an extent, the huge class imbalance in the dataset (the voxel ratio between the 'FCD' class and the 'background' class is almost 1:100) can still influence the segmentation performance. Future works are planned to integrate design considerations such as the usage of an adaptive data augmentation (to reduce the impact of class imbalance), and modifying the 3D CNN architecture by analyzing the data from all three perpendicular directions (Sagittal, Coronal, and Axial ) of the 3D MRI, cascading post-segmentation algorithms that can provide more attention on the lesion regions, etc.

## 3.4 Summary

This chapter explores the development and implementation of an automated 3D CNN-based model for the segmentation of FCD lesions utilizing 3D FLAIR MRI scans. The proposed 3D CNN model is crafted to overcome inherent limitations associated with both 2D and 3D CNNs while harnessing the benefits of each approach. Design features

of the model include the use of shallow slicing to generate numerous shallow 3D volumes and the incorporation of residual blocks within the CNN framework. These features enable the model to learn effectively with fewer parameters, reduced memory consumption and decreased training time compared to traditional 2D methods. The architecture demonstrates superior performance in FCD segmentation relative to leading 2D methods. Experimental results indicate that the proposed 3D deep learning model achieves improvements of 2.7%, 14.7%, and 1.1% in pixel-wise, region-wise, and patient-wise FCD detection rates (measured in terms of recall), respectively, compared to the state-of-the-art approaches in FCD segmentation. Therefore, this model offers a valuable tool for neuro-radiologists in identifying FCD regions in patients with intractable epilepsy. Furthermore, this research lays the groundwork for future studies focused on the development of efficient 3D CNN architectures optimized for constrained hardware environments and smaller datasets.

# CHAPTER 4

# A DUAL ENCODER-DECODER MULTI-TASK 3D DEEP LEARNING FRAMEWORK FOR THE SEGMENTATION OF FCD LESIONS

## 4.1 Introduction

In the context of medical imaging, 3D CNNs are especially valuable for interpreting 3D cross-sectional data, such as MRI and CT scans. The key advantage of using 3D CNNs for segmentation lies in their ability to capture the inter-slice features in 3D data, which allows the network to learn features with a full understanding of the volumetric context, leading to more accurate and consistent detections of abnormalities.

A dual encoder CNN model represents a powerful approach to handling multi-input data, particularly in the context of medical cross-sectional scans, to efficiently utilize the correlations over different input characteristics simultaneously. In this study, we propose a dual input 3D CNN model that processes FLAIR MRI volumes and their corresponding cortical thickness maps for segmenting FCD lesions.

Recent advancements in deep learning techniques have significantly elevated the performance of image segmentation, playing a pivotal role in the automation of medical image analysis (Moeskops *et al.*, 2016; Guo *et al.*, 2015). In the study by Feng et al. (Feng *et al.*, 2020*c*), a deep CNN was developed to localize FCD in MR images. While

[3]The work described in this chapter has been submitted for publication to: **S. Niyas**, Chandrasekharan Kesavadas, and Jeny Rajan (2024). A Dual Encoder-Decoder Multi-task 3D Deep Learning Framework for the Segmentation of Focal Cortical Dysplasia Lesions. Biomedical Signal Processing and Control.

effective at identifying frames containing dysplasia lesions, the model fails to provide specific details about FCD lesions. Gill et al. (Gill *et al.*, 2018, 2019) introduced patch-wise FCD segmentation models, highlighting their efficiency in analyzing individual patches of medical images with minimal computational resources. However, these patch-wise approaches encounter limitations in capturing global contextual information from entire MR frames and are susceptible to inaccurate predictions. Dev et al. (Dev *et al.*, 2019) introduced a customized U-NET model, a sophisticated deep learning-based semantic segmentation approach that can localize and segment FCD lesions using FLAIR MRI. To improve FCD segmentation, Thomas et al. (Thomas *et al.*, 2020) introduced a model that combines an improved U-Net network with multi-residual convolutional and attention blocks. Although these 2D CNN models deliver good segmentation performance, they cannot account for the inter-slice correlations in 3D MRI data sets.

In their study, Niyas et al. (Niyas *et al.*, 2021) proposed an efficient 3D Residual U-Net framework for the segmentation of FCD in 3D brain MRI. This approach benefits the advantages of 2D and 3D CNNs through a strategic arrangement of input data slices and network architecture. The model employed a novel technique of shallow sliced stacking to augment the number of 3D samples generated from limited datasets. The aforementioned segmentation model also incorporated convolutional residual blocks in the contracting path, facilitating the extraction of multi-scale features while minimizing the training complexity.

This chapter introduces an advaced 3D deep learning model that employs a multi-view dual encoder-decoder architecture for precise segmentation of FCD lesions within MRI volumes. This model leverages a 3D CNN framework enhanced with residual connections, forming the core of our segmentation network. Several architectural enhancements have been incorporated into our approach. Firstly, we adopt multi-view training, inspired by the methodologies used by neuro-radiologists in examining MRI volumes. Our model processes input data along the sagittal, axial, and coronal axes, updating the model weights accordingly. To further enhance the segmentation performance of FCD, the model simul-

taneously processes FLAIR MRI volumes and their associated cortical thickness maps through a dual-encoder network. Additionally, our model includes a dual-decoder stage, which facilitates dual-task learning by leveraging distance maps derived from ground truth data.

The main contributions presented in this research work can be summarized as follows:

1. ***Multi-view training:*** The model evaluates data from three distinct orientations: sagittal, axial, and coronal, and subsequently adjusts the model weights, drawing insights from this tri-axial analysis.

2. ***A Dual encoder-decoder model optimized for learning from cortical thickness:*** The model simultaneously processes FLAIR MRI and their associated cortical thickness maps using a dual-encoder network. The individual encoders communicate within this network through a 3D attention mechanism, enriching the feature extraction process. Additionally, the architecture incorporates a dual-decoder stage designed for dual-task learning. This stage leverages distance maps generated from the ground truth labels, enhancing segmentation accuracy.

3. ***A cascading strategy for transitioning from coarse to fine segmentation:*** The entire segmentation process operates in a cascaded fashion, encompassing two distinct stages. The first stage works on the full-sized image to localize FCD lesions, minimizing the rate of false detections. In the second stage, the model narrows its focus to a localized neighborhood surrounding the lesions identified in the first stage. This enables fine-level segmentation, allowing for a more detailed and comprehensive delineation of the complete FCD regions.

4. ***3D Attention network for maintaining consistency between encoder and decoder pairs:*** The proposed architecture incorporates a Dual Input 3D Convolutional Block Attention Module (3D CBAM) attention mechanism to reinforce the feature maps in the encoding stage.

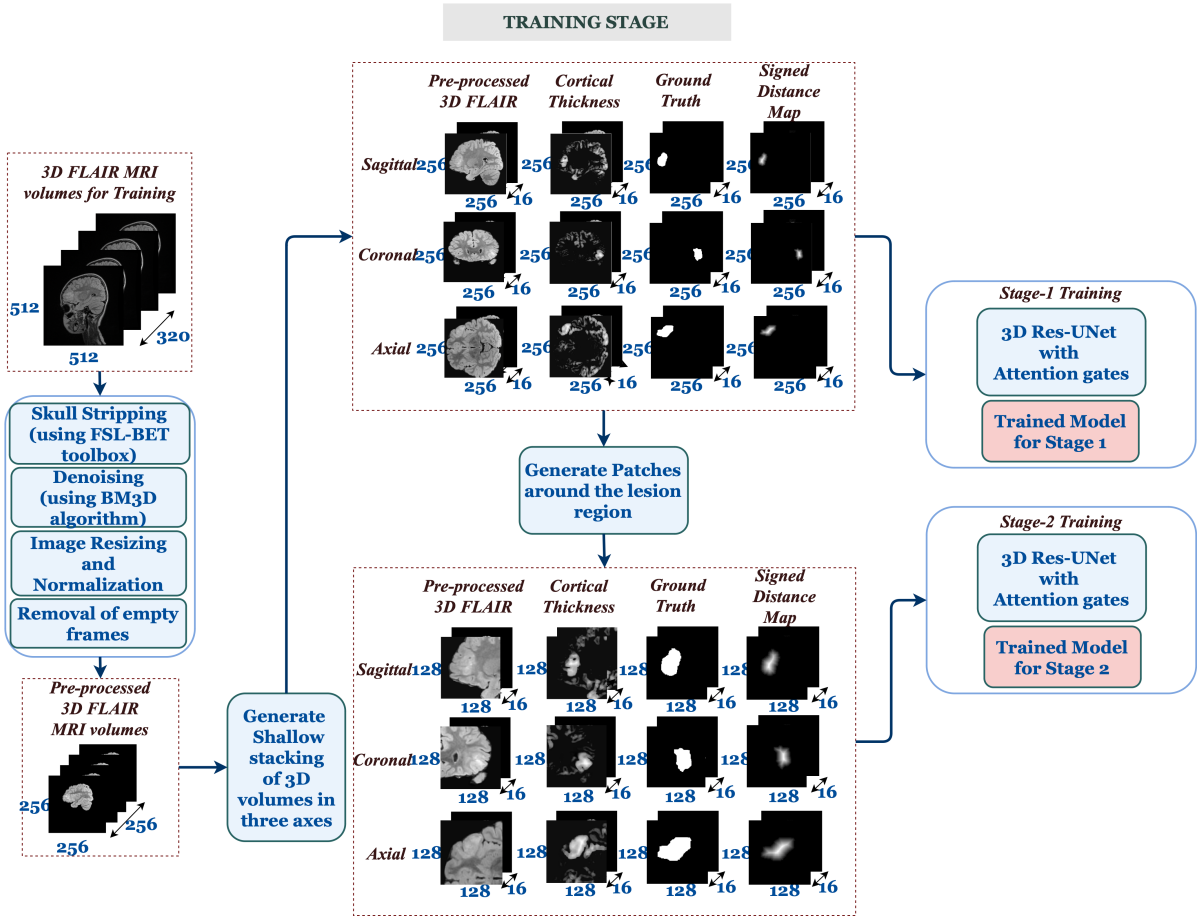



Figure 4.1: Training stage of the proposed FCD segmentation model.

## 4.2 Methods

Figure 4.1 and 4.2 depict the training and testing phases of the proposed FCD segmentation method.

### 4.2.1 Preprocessing

During MRI scanning, signal corruption may arise from various sources, such as the thermal motion of the patient and radio frequency emissions from MRI coils and associated electronics. Consequently, MRI scans are frequently affected by noise, necessitating denoising as an essential preprocessing step. In this study, we utilize the Block-Matching

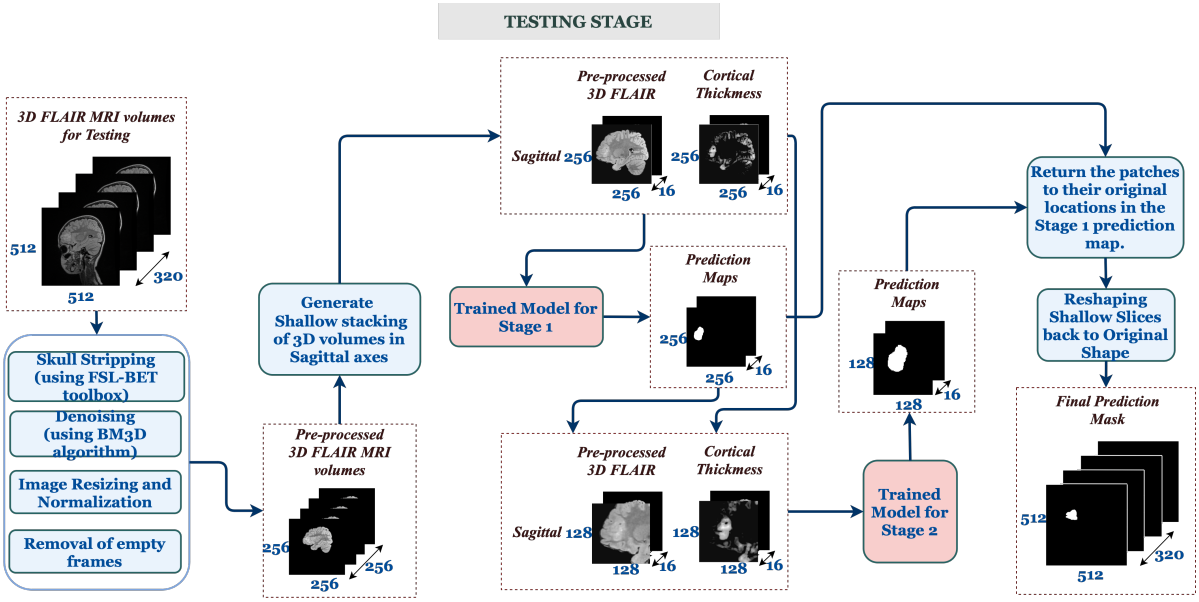
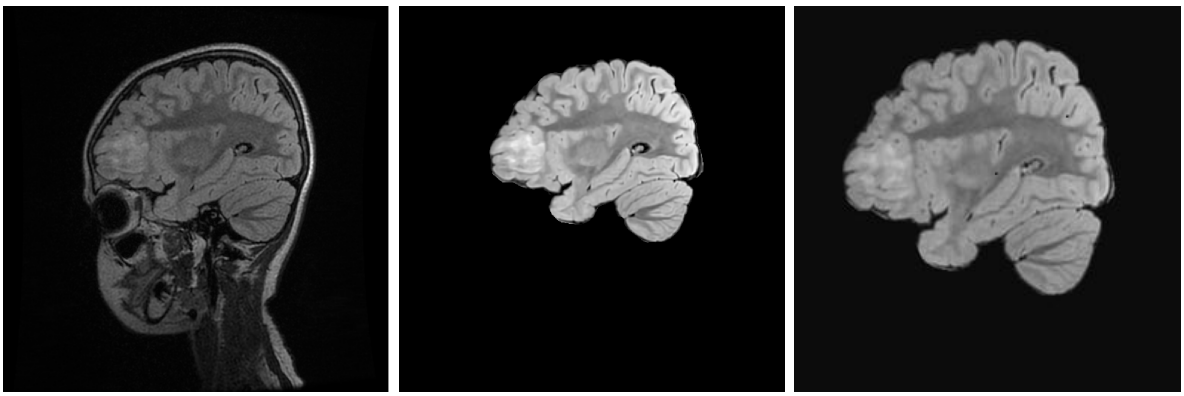



Figure 4.2: Testing stage of the proposed FCD segmentation model.

and 3D Filtering (BM3D) algorithm (Maggioni and Foi, 2012; Maggioni *et al.*, 2012) to perform the denoising, which significantly enhances the Signal-to-Noise Ratio (SNR) of the input data. Characteristically, for detecting FCD, focusing on the brain area minimizes false alarms in later processing. We employ the FSL-BET toolbox (Smith, 2002) for skull-stripping to focus solely on the brain region.

Each input volume in the current FCD dataset has dimensions of $512 \times 512 \times 320$.



(a) (b) (c)

Figure 4.3: (a) Raw MRI slice, (b) Result after skull-stripping, and (c) Result after cropping & denoising

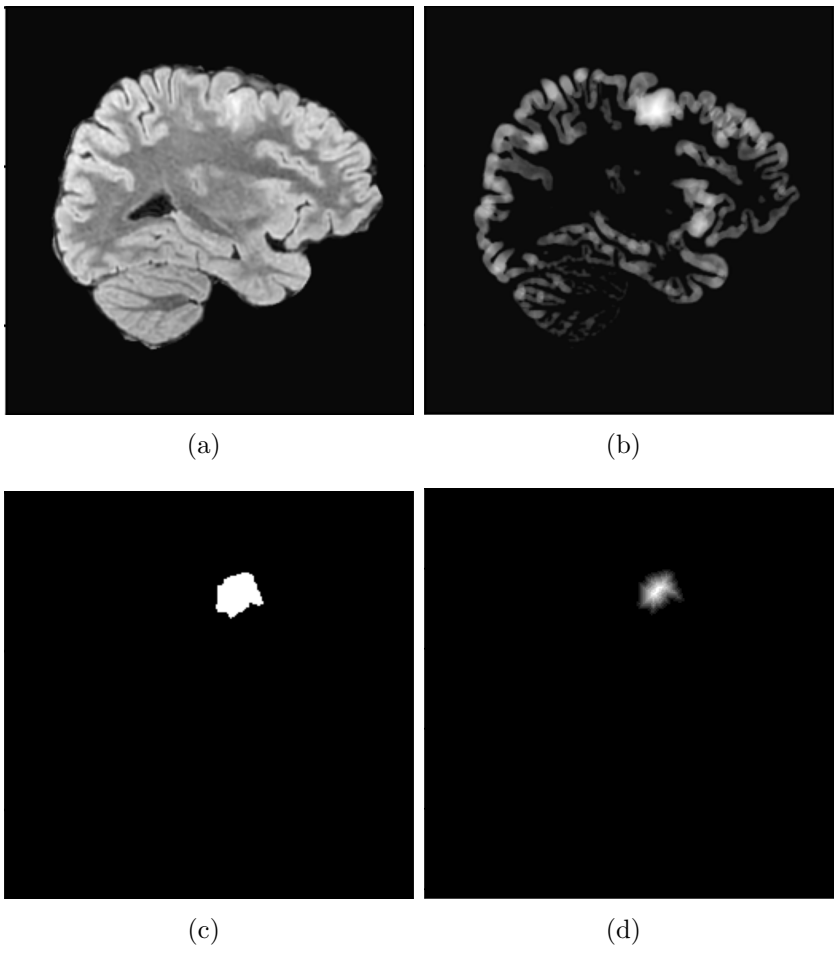
However, the proposed segmentation model is targeted for MRI data having a resolution of $256 \times 256 \times 16$, aiming to balance segmentation accuracy and computational load effectively. In the preprocessing stage, we aim to resize the data without compromising the intricate details of the brain region. To achieve this objective, MRI volume trimming is conducted in two sequential stages. Initially, slices deemed irrelevant to the study are eliminated. Initial and final slices of each MRI volume typically contain minimal or no relevant information concerning brain anomalies. Consequently, these frames are excluded, resulting in a volume shape of $512 \times 512 \times 256$ pixels. Subsequently, in the second stage, the slices are cropped and resized to focus solely on the brain region, thereby minimizing the presence of extraneous empty pixels. The cropping uses an adaptive bounding box approach across the MRI volumes to eliminate sparse regions without losing brain information.

The MRI volumes, after cropping, are subsequently adjusted to dimensions of $256 \times 256 \times 256$, ensuring uniformity in slice numbers across all three axes. Identical cropping handles are used to crop the MRI volume, Cortical thickness map, and ground truth, ensuring their spatial coherence is preserved. Additionally, each slice is subjected to Z-score normalization to achieve a zero mean and unit variance. This normalization process helps eliminate redundant details and facilitates faster learning when training with CNN models. Figure 4.3 illustrates the images before and after undergoing various preprocessing steps. It shows that preprocessing effectively removes the skull region, while adaptive cropping preserves the region of interest and optimizes the spatial resolution of the brain slices.

### 4.2.2 A Dual Encoder-Decoder Segmentation Framework .

Figure 4.4 shows a sample MRI scan, cortical thickness map, binary mask for highlighting FCD lesions, and the distance map derived from the binary mask. It is worth highlighting that cortical thickness, which is calculated as the distance between the innermost and outermost layers of the gray matter, is crucial for assessing the structural integrity of the

cerebral cortex (Feng *et al.*, 2020*b*). Hence, the proposed model is designed to analyze FLAIR MRI scans and their corresponding cortical thickness maps concurrently. The FLAIR MRI scans and the corresponding cortical thickness maps are given into a specially designed dual-encoder framework to facilitate parallel learning from two diverse input data. In the proposed architecture, the individual encoders are deployed for specific tasks: the left encoder is designated for the processing of MRI data, while the right encoder is intended for learning features associated with cortical thickness. These encoders interact through a 3D attention mechanism that substantially enhances the feature extraction process, leading to more consistent and comprehensive analysis.



(a) (b)

(c) (d)

Figure 4.4: (a) Sample brain MRI slice, (b) Cortical thickness map, (c) Ground truth and (d) Distance map

### 4.2.3 Multi-view Learning

The segmentation framework employs a multi-view training approach that incorporates sub-volumes from axial, coronal, and sagittal axes for comprehensive learning. The concept of multi-view training is motivated by the neuroradiologists' approach while examining an MRI volume to isolate FCD or similar anomaly regions.

The utilization of 3D scans ensures data consistency across slices, facilitating a thorough examination of anomalies across multiple slices. Nevertheless, the adoption of a 3D CNN framework leads to a significant escalation in trainable parameters, computational complexity, and memory requisites compared to 2D CNN methodologies. This escalation correlates directly with the depth of frames in the 3D samples. In this study, shallow stacked MRI slices are utilized to modulate the depth of input data samples, minimizing computational and memory demands effectively. This shallow slice stacking is accomplished by partitioning MRI volumes into multiple overlapping shallow slices, influenced by the 3D CNN approach explained in the previous chapter (Chapter 3).

As illustrated in Figure 4.1, the segmentation model processes the entire MRI volume of dimension $256 \times 256 \times 256$, with the associated cortical thickness maps through individual encoder paths. The model dynamically generates shallow stacked slices of size $256 \times 256 \times 16$ across three planes: sagittal, axial, and coronal, which are subsequently passed into the 3D CNN framework. Training is designed first to undergo ten epochs using shallow stacked slices from the sagittal plane, followed by five epochs each for the shallow stacked slices from the axial and coronal planes. Cumulatively, these 20 epochs constitute a super-epoch, and the overall training process is carried out by up to 30 super-epochs. We conducted experiments using various weight combinations along different image axes. The ablation study revealed superior results when the distribution of epochs placed greater emphasis on the sagittal plane.

### 4.2.4 A Dual-task Learning for Preserving Lesion Boundaries

The proposed model introduces a dual-decoder phase specifically tailored for dual-task learning. This phase utilizes distance maps derived from binary ground truth labels to reinforce learning. The feature space from the bottleneck layer is processed in parallel by dual-decoder networks. One decoder functions similarly to a traditional decoder stage, receiving long skip connections from the encoder layers that process MRI data. In contrast, the second decoder refines the feature space by accepting skip connections from the encoder layers that handle the cortical thickness map.

The dual-task learning process involves evaluating the difference between the probability maps of both classification decoder layers, relative to the binary ground truth and the ground truth's distance transform. The output of the first decoder is compared to the binary ground truth, ensuring equal weightage on loss calculation across all lesion regions. Conversely, the output of the second decoder is assessed against the distance map, which places greater emphasis on the central lesion region compared to the lesion edges.

To generate these distance maps, we employ a transformation function. This function determines the distance transform of the binary input image, converting each foreground (non-zero) element to its nearest distance from background pixels, as given in Equation 4.1.

$$Y_{DM(i,j)} = \sqrt{\left(Y_{(i,j)} - b_{(i,j)}\right)^2} \tag{4.1}$$

where $Y_{(i,j)}$ is the binary ground truth image with pixel coordinates $(i, j)$, $b_{(i,j)}$ is the background point (value 0) with the smallest Euclidean distance to $Y_{i,j}$, and $Y_{DM}$ is the distance map image.

Figure 4.4(c) and Figure 4.4(d) depict the sample binary ground truth and distance map, respectively. Similar to the encoder stage, the decoders also engage through a 3D

attention mechanism, significantly enhancing the feature extraction process. Distance map-based loss computation provides a spatially aware mechanism for segmentation, which enhances the model's ability to learn complex shapes such as FCD lesions (Luo *et al.*, 2021). Hence, the dual-task learning approach, combining shape extraction with traditional segmentation, can help in reducing false positives and is beneficial in precisely localizing the FCD lesions.

### 4.2.5 A Cascading Strategy for Coarse to Fine Segmentation.

Generally, segmentation models often employ patch-wise strategies for their notable advantages, including enhanced learning from extensive regions of interest and diminished computational expenses. Nevertheless, in the FCD dataset utilized in this study, the patch-wise segmentation approaches demonstrate minimal efficacy due to the lesion regions' inter-class homogeneity and intra-class heterogeneity. To address this limitation, the proposed model incorporates a segmentation method through a cascaded learning approach, functioning in two distinct stages.

The initial stage focuses mostly on lesion localization with high precision, wherein the parameters of the loss function are fine-tuned to minimize false positives, i.e., incorrect detection of FCD lesions. This stage does not effectively outline the full extent of the lesions due to the attempt to detect lesions across the entire brain. To address this limitation, a second segmentation stage is introduced in the proposed segmentation pipeline. A neighborhood patch of size $128 \times 128$ surrounding the detected FCD lesion regions is extracted, and these cropped shallow slices $128 \times 128 \times 16$ are passed to the subsequent segmentation model. This cropping uses an adaptive bounding box approach across the possible FCD lesions, and the coordinates of the bounding box are computed as shown in Equations 4.2-4.4.

$$[C_x, C_y] = [\frac{M_{01}}{M_{00}}, \frac{M_{10}}{M_{00}}] \tag{4.2}$$

$$[X_{min}, X_{max}] = [C_x - 50, C_x + 50] \tag{4.3}$$

$$[Y_{min}, Y_{max}] = [C_y - 50, C_y + 50] \tag{4.4}$$

where $[C_x, C_y]$ represents the centroid pixel coordinates of the binary segmentation map, $[X_{min}, X_{max}]$ are the minimum and maximum X-coordinates, and $[Y_{min}, Y_{max}]$ are the minimum and maximum Y-coordinates of the bounding box, respectively. Also, $M_{00}$ is the zeroth moment; representing the total mass or area of the shape, $M_{10}$ is the first moment about the y-axis and $M_{01}$ is the first moment about the x-axis.

The latter stage also employs dual-task learning and dual encoder-decoder systems, similar to the localization stage. Nonetheless, to accommodate the reduced spatial size of the input data, the model's complexity is receded, entailing a reduction in convolution kernels per layer and the overall depth of the CNN architecture. This approach ensures accurate lesion segmentation while maintaining computational efficiency. The loss parameters in the second stage are adjusted to get a reasonable trade-off between the recall and precision performance.

### 4.2.6 3D Attention network for Maintaining Consistency between Encoder and Decoder Pairs.

Our architecture incorporates a modified 3D CBAM attention mechanism, drawing inspiration from the CBAM model detailed by Woo et al. (Woo *et al.*, 2018). This module adeptly adjusts feature maps by dynamically accentuating or diminishing them, depending on their relevance. In our dual-encoder setup, we simultaneously process MRI slices alongside cortical thickness maps. This approach is mirrored in the dual-decoder module, where it handles both the binary ground truth and the distance map.

To ensure a seamless flow of data through the encoder-decoder pairs, our model integrates a sequence of attention blocks. The proposed dual-input 3D CBAM takes two sets of 4D feature maps $H \times W \times D \times C$ as inputs. The module calculates an attention map

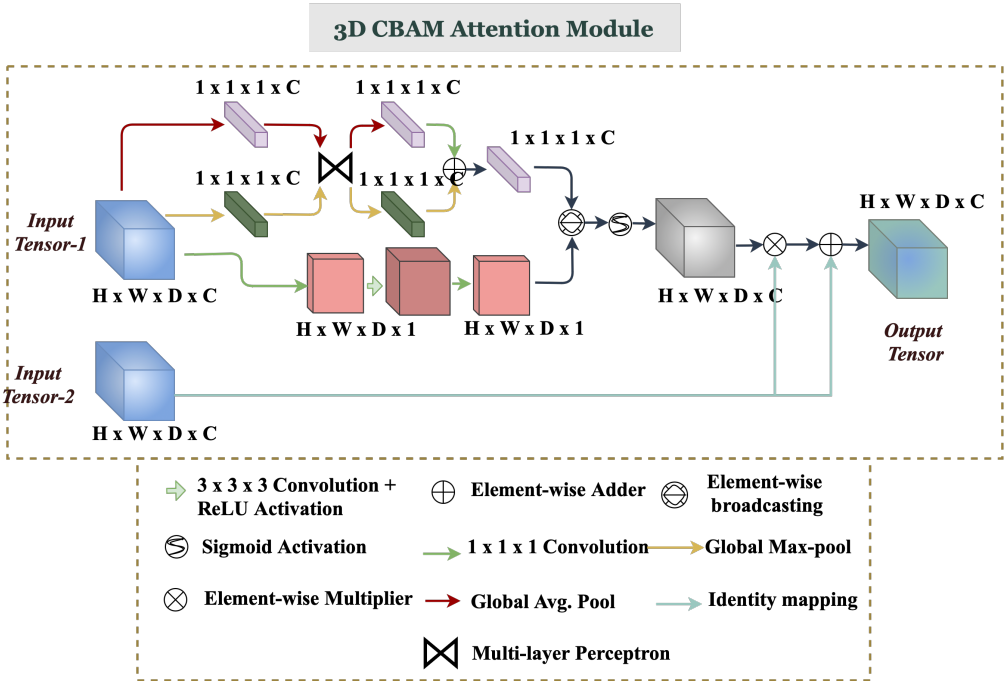



Figure 4.5: Proposed 3D CBAM attention module with dual inputs.

from the primary input and then applies this map to the secondary input, as depicted in Figure 4.5. This attention mechanism operates along two distinct pathways: channel-wise and spatial-wise. The channel attention pathway combines both global max pooling and global average pooling methods to condense the spatial dimensions to a singular point. The result of each pooling operation is a vector $1 \times 1 \times 1 \times C$ that captures essential global information for each channel. These vectors are then processed by a shared network to create a channel attention map. Once the network processes the vectors, their output feature vectors are combined using an element-wise addition approach.

The spatial attention is computed by first applying a 3D convolution with a kernel of size $1 \times 1 \times 1$ to the 4D input volume to reduce the channel dimension. This is followed by 3D convolutional layers to aggregate contextual information with a larger receptive field. The output of the spatial attention pathway is then converted to $H \times W \times D \times 1$ using $1 \times 1 \times 1$ convolution. These outputs are fused using a broadcasting element-wise addition operation, yielding a final attention map with the shape $H \times W \times D \times C$. To ensure the values are within the 0 to 1 range, a sigmoid activation function is applied. Finally,

an element-wise multiplication is performed between the attention map (generated from the primary volume) and the secondary input volume, and the result is added to the secondary input volume itself. Thus, the attention module returns a 4D feature map of the same size as the input feature space.

### 4.2.7 Network Architecture.

This research presents a cascaded 3D deep learning framework employing a dual encoder-decoder structure. Our methodology is centered around an enhanced 3D CNN architecture, which incorporates residual connections to form the backbone of our segmentation network. The detailed architecture is presented in Figure 4.6. Instead of processing the entire MRI volume as a single input, the model splits it down into smaller, overlapping 3D sub-volumes of size $256 \times 256 \times 16$. This approach leads to notable performance improvements, such as reducing model complexity and memory demands, and generating sufficient training data, eliminating the need for extensive data augmentation techniques. We performed trials involving varying depths of shallow slices (8, 16, 32, and 64) and different stride values (1, 3, and 5). Our findings indicated a favorable balance between segmentation accuracy and computational efficiency when employing smaller stride values. This can be attributed to the generation of a greater amount of training data. Consequently, based on empirical observations, we have chosen a depth and stride of 16 and 1, respectively, for the shallow sliced data.

In the traditional 3D U-Net design, each encoder level uses a pair of convolutional layers. This incurs a large quadratic computational cost with limited feature extraction at different scales. To address these issues, in the proposed network, we replace the sequential convolutions with residual modules, which leads to faster convergence and benefits multi-scale feature extraction. At each encoder depth, a 3D convolutional layer is employed, succeeded by two residual modules, each containing a pair of 3D convolutional layers. Similarly, a transposed convolution layer (Noh *et al.*, 2015) is used in the decoder phase, followed by two residual modules. All convolutional operations use 3D kernels

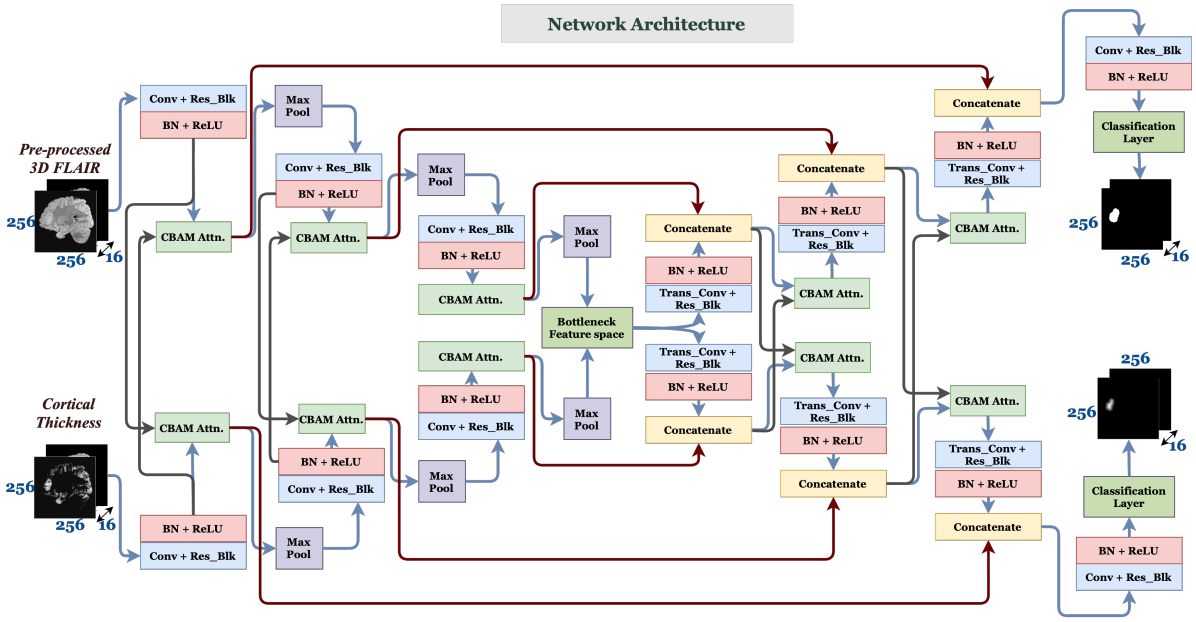



Figure 4.6: Architecture diagram of the proposed 3D CNN model for FCD segmentation

with size $3 \times 3 \times 3$, and the convolution layers are followed by batch normalization (Ioffe and Szegedy, 2015) and 'LeakyReLU' activation (Nair and Hinton, 2010). To reduce the dimensionality of the data, a $2 \times 2 \times 2$ max-pooling operation is applied after each depth, which halves the size of the feature maps. To compensate for the loss of information due to max-pooling, the number of filters in subsequent layers is also doubled.

The proposed model uses a cascade of two networks: one for the initial localization stage and the other for the fine-level segmentation phase. In the localization phase, the proposed architectural framework uses a depth of 3 and is targeted for processing input data sized $256 \times 256 \times 16$. At the initial depth, all convolutions make use of eight filters each. In the subsequent depths, the number of filters gets doubled. The decoder path uses transposed convolution to upsample feature maps and concatenates them with high-resolution features from the contracting path. To mitigate overfitting, dropout, and regularization are also incorporated. The secondary network (for fine-level segmentation) adopts a similar architecture with a depth of two, tailored to process inputs of shape $128 \times 128 \times 16$.

The decoder path compares the predicted masks with respect to the binary ground truth and the corresponding signed distance maps. The networks consider two loss functions during training: conventional segmentation loss ($L_{Seg}$) and the signed distance map loss ($L_{Dist}$). The dataset employed in this study shows class imbalance, where lesion voxels constitute less than 5% of the negative voxels. To address this significant class imbalance, the Tversky loss function (Salehi *et al.*, 2017) is employed as shown in Equation 4.6. It uses the Tversky similarity index (Equation 4.5), which has the ability to mitigate class imbalance through its generalized Dice coefficient properties.

$$\text{Tversky similarity index} = \frac{TP}{TP + (\alpha \times FP)\,(\beta \times FN)} \tag{4.5}$$

$$L_{Seg} = 1 - \text{Tversky similarity index} \tag{4.6}$$

where $TP$ represents true positives, $FP$ stands for false positives, and $FN$ denotes false negatives. The parameters $\alpha$ and $\beta$ play roles in determining the weights associated with $FP$ and $FN$, respectively.

When $\alpha = \beta = 0.5$, both $FP$ and $FN$ incur equal penalties. In such instances, the Tversky similarity index aligns closely with the Dice coefficient. Our experimentation involving loss functions like *Binary Cross Entropy* (BCE) loss and Dice loss revealed a notable disparity between Precision and Recall values. This difference arises from the class imbalance between the FCD region and the normal brain region. This imbalance not only impacts the Precision-Recall trade-off but also hinders the model's ability to effectively detect FCD lesions. To address this imbalance, we adopted the Tversky loss, which assigns higher weights to $FN$. This adjustment enhances the Recall rate while maintaining satisfactory precision performance. Through an iterative process of experimentation with various Tversky loss parameter combinations, we empirically determined the optimal values for $\alpha$ and $\beta$ to be 0.3 and 0.7, respectively.

The Mean Squared Error (MSE) between the ground truth and predicted signed distance maps is the signed distance map loss. The total loss is computed by taking

a weighted summation of the $L_{Seg}$ and $L_{Dist}$. The total loss is computed as shown in Equation 4.7.

$$\text{Total loss } (L_{Total}) = \gamma \times L_{Seg} + (1 - \gamma) \times L_{Dist} \tag{4.7}$$

where $\gamma$ is the weight factor and is empirically set as 0.5 for the initial localization phase and 0.75 in the second fine-level segmentation stage.

## 4.3 Results and Analysis

This section presents the hardware details, evaluation metrics, ablation study, and discussion involving qualitative and quantitative analysis.

### 4.3.1 Hardware Details

Experiments were carried out on an NVIDIA® DGX-1® machine running the Ubuntu operating system, with NVIDIA® Tesla® V100 GPUs with 32GB of dedicated graphics memory. The models were implemented using the Keras and TensorFlow libraries in Python.

### 4.3.2 Evaluation Metrics

Quantitative analysis employs metrics such as Precision, Recall, and the Dice coefficient. Precision measures the proportion of true positives among all predicted positives, while Recall estimates the fraction of true positives identified within the actual positive class. The Dice coefficient, a weighted average of these two, accounts for both *FP* and *FN*, making it a popular benchmark for evaluating the overlap between predictions and ground truth.

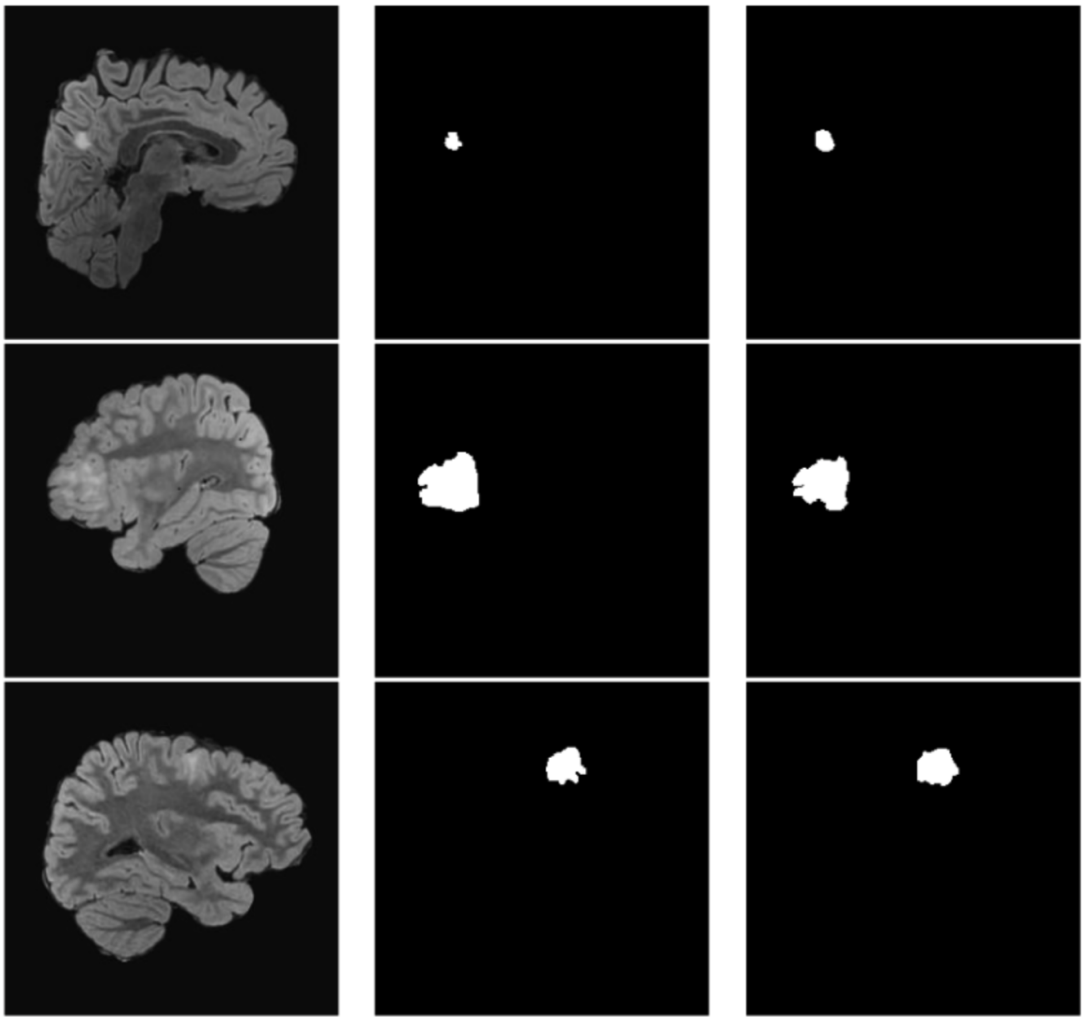
### 4.3.3 Datasets

The data for this study were sourced from the Sree Chitra Tirunal Institute for Medical Sciences and Technology (SCTIMST), Trivandrum, India, with the institutional ethics committee's approval (Ref. No.: IEC/1073). The dataset comprises 8000 FLAIR brain MR scans from 26 patients acquired at an image resolution of $512 \times 512$. These images were captured on a 3T MRI scanner (GE Healthcare, UK) in the sagittal plane, which has 320 slices per volume. The scanning parameters included a slice thickness of 1 mm, a pixel spacing of 0.5 mm, and a TR/TE/TI/flip angle of $7200ms/117.241ms/1936ms/90°$, respectively.

### 4.3.4 Training Methodology

We employed random indexing to create training, testing, and validation sets for the experiments. During this process, we meticulously ensured that these sets were segregated to prevent any data leakage across the training, testing, and validation phases. During the coarse-level segmentation phase, both the pre-processed MRI volume and cortical thickness map of shape $256 \times 256 \times 256$ are passed. The model dynamically generates shallow stacked slices of size $256 \times 256 \times 16$ across three planes: sagittal, axial, and coronal, which are subsequently passed into the 3D CNN framework for first-level segmentation. Training is designed first to undergo ten epochs using shallow stacked slices from the sagittal plane, followed by five epochs each for the shallow stacked slices from the axial and coronal planes. Cumulatively, these 20 epochs constitute a super-epoch, and the overall training process is carried out by up to 30 super-epochs.

To match the model's input size $256 \times 256 \times 16$, overlapping 3D sub-volumes of the same size were generated with one-voxel stride for training and validation sets. Test data used non-overlapping sub-volumes. Limited offline augmentation tripled training samples. Additionally, to address memory constraints and reduce the class imbalance, the training set was resampled to a 1:3 positive-to-negative ratio by randomly selecting



(a) (b) (c)

Figure 4.7: Qualitative analysis of the proposed model: (a) Brain MRI scans with FCD, (b) Ground truth, and (c) Predicted output of the proposed 3D Res-UNet.

only 33% of total negative samples.

The network's architecture was empirically optimized through hyperparameter tuning. Adam optimizer with Adagrad update rule (Kingma and Ba, 2014; Reddi *et al.*, 2018) employed for weight updates with an initial learning rate of 0.001 and batch size of 16. *He Normal* (He *et al.*, 2015*a*) initialization and L2 regularization with $l2 = 0.01$ (Cortes *et al.*, 2012) were applied to all convolutional layers. Both localization and fine-level segmentation phases were trained for 30 super-epochs with an early stopping patience

Table 4.1: Performance comparison of the proposed model with benchmark models.

| Folds | Attention-UNet (Thomas *et al.*, 2020) | | | MultiRes 3D Res-UNet (Niyas *et al.*, 2021) | | | Proposed Model | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Dice** | **Precision** | **Recall** | **Dice** | **Precision** | **Recall** | **Dice** | **Precision** | **Recall** |
| **Fold 1** | 62.55 | 59.91 | 65.63 | 69.53 | 69.91 | 69.17 | 72.67 | 75.51 | 70.04 |
| **Fold 2** | 45.67 | 73.67 | 33.12 | 46.75 | 66.13 | 36.16 | 50.04 | 71.24 | 38.56 |
| **Fold 3** | 72.54 | 68.65 | 77.00 | 73.35 | 67.80 | 79.89 | 74.87 | 70.52 | 79.78 |
| **Fold 4** | 71.49 | 66.72 | 77.33 | 72.09 | 70.81 | 73.42 | 71.97 | 71.09 | 72.88 |
| **Fold 5** | 56.81 | 71.57 | 48.05 | 59.89 | 73.25 | 50.66 | 63.61 | 76.62 | 54.37 |
| **Average** | **61.81** | **68.10** | **60.23** | **64.32** | **69.58** | **61.86** | **66.63** | **73.00** | **63.13** |

of 10 epochs. The best weights (based on validation loss) at each epoch were saved for further analysis and model deployment.

### 4.3.5 Results and Discussion

This study employs both qualitative and quantitative analyses to evaluate the performance of the proposed FCD segmentation method. We compare it against state-of-the-art 2D CNN and 3D CNN-based FCD segmentation models (Thomas *et al.*, 2020; Niyas *et al.*, 2021). Figure 4.7 showcases samples of the predicted segmentation masks used in the qualitative assessment. The close resemblance between the predictions and the ground truth demonstrates the method's effectiveness in the accurate segmentation of FCD regions.

Quantitative assessment employs a five-fold cross-validation, with three repetitions of experiments within each fold. The reported outcomes reflect the average results obtained from three instances of 5-fold cross-validations. Table 4.1 showcases the outcomes for FCD lesion segmentation, featuring a comparative analysis with cutting-edge FCD segmentation techniques. Our proposed methodology excels in identifying more boundary pixels within FCD lesions, resulting in enhanced performance in all benchmark metrics

compared to alternative approaches.

## 4.4 Summary

In this chapter, we presented an effective 3D deep learning model designed for precise segmentation of FCD lesions in MRI volumes. This model leverages an advanced multi-view dual encoder-decoder architecture underpinned by a 3D CNN framework with integrated residual connections. It uniquely processes FLAIR MRI and cortical thickness maps through a dual encoder-decoder network, enhanced by a 3D attention mechanism focusing on cortical brain regions. Key enhancements include a multi-view training technique and dual-task learning utilizing distance maps from ground truth data. The segmentation operates in two stages: an initial broad localization of FCD lesions, followed by a focused, fine-level segmentation. This dual-phase approach ensures a balance between broad localization and precise segmentation, establishing a robust method for FCD lesion detection. Our model demonstrates superior performance over existing methods in both quantitative and qualitative aspects of FCD lesion segmentation. Future work will aim to further enhance segmentation accuracy by incorporating advanced generative AI techniques.

# CHAPTER 5

# SEGMENTATION OF ISCHEMIC STROKE LESIONS FROM CT PERFUSION IMAGES USING 3D ATTENTION-DRIVEN VOX2VOX

## 5.1 Introduction

In recent years, CNN-based methodologies have emerged as the predominant techniques for medical image segmentation. Particularly, when dealing with medical volumes, 3D CNNs are capable of learning inter-slice relationships, which are significant for identifying abnormalities that spread across multiple slices. Consequently, there has been a notable shift in research focus towards 3D CNN methodologies that can effectively process volumetric cross-sectional images, such as CT and MRI. Particularly, the segmentation of ischemic stroke lesions from such medical images has attracted increasing attention in recent years, reflecting its critical role in enhancing diagnostic accuracy and treatment efficacy.

Stroke ranks as the third leading cause of death globally and is the primary cause of acquired disability (Sudlow and Warlow, 1997). The most common type of stroke, accounting for nearly 80% of cases, is ischemic stroke. This occurs when an artery is blocked, leading to reduced blood flow to the brain, tissue damage due to lack of oxygen, and, ultimately, tissue death (infarction). A quicker response time from symptom onset to treatment significantly improves patient outcomes (Jahan *et al.*, 2019). Initially, the stroke-affected areas are categorized into two main areas: the infarct core, which consists

[3]The work described in this chapter has been submitted for publication to: **S. Niyas**, Vivek A. Saraf, Ajith Abraham, Neethi A. S. and Jeny Rajan (2024). Segmentation of Ischemic Stroke Lesions from CT Perfusion images using 3D Attention-Driven Vox2Vox. Journal of Medical Imaging

of irreversibly damaged tissue, and the penumbra, which is at-risk tissue that could recover if blood flow is restored. Identifying and measuring the acute core or penumbra is crucial in clinical practice as it helps assess the potential tissue recovery with different interventions, leading to better treatment decisions.

Current screening techniques for diagnosing ischemic stroke primarily involve using Non-Contrast Computed Tomography (NCCT), CT perfusion (CTP), and MRI methods to get a clearer picture of cerebral blood flow. The decision on which screening method to use for timely detection and analysis of ischemic stroke depends on several factors, including the clinical condition, the availability of resources, and specific patient characteristics. CTP is often chosen over NCCT because it has higher accuracy in detecting areas with reduced blood flow and the ischemic penumbra and its ability to distinguish between other conditions, such as hemorrhagic stroke or non-stroke issues like migraines (Donahue and Wintermark, 2015). Moreover, CTP is preferred over MRI for the initial assessment of ischemic strokes due to its widespread availability, faster image capture, lower cost, and easier patient monitoring (Gillebert *et al.*, 2014).

In CTP imaging, a series of CT Angiography (CTA) images (4D spatiotemporal images) are captured while perfusion occurs. This process generates maps of perfusion maps such as Cerebral Blood Flow (CBF), Cerebral Blood Volume (CBV), Mean Transit Time (MTT), and Time-to-maximum flow (Tmax) (Demeestere *et al.*, 2020). These maps facilitate a more efficient detection of ischemic stroke lesions compared to relying solely on NCCT. Manual lesion segmentation from CTP is a laborious process, while automated stroke lesion segmentation results in more accurate and consistent stroke detection compared to manual screening, which may be subject to human error and variability. Moreover, using automated methods for stroke detection using CTP offers significant advantages in terms of efficiency, reliability, integration with clinical workflow, and reduced subjectivity, making it a valuable tool in modern healthcare.

In response to the growing significance of visual data, there has been a notable surge in research focused on constructing neural network models tailored for images. This

has notably led to the emergence and refinement of CNN, marking a key advancement in deep learning. The successful deployment of deep CNN models has substantially impacted a diverse range of applications within image processing and computer vision domains. Recent strides in deep learning methodologies have empowered researchers to introduce advanced techniques for addressing various challenges in medical imaging, such as segmentation, registration, and classification. This trend can be seen in the latest methodologies devised for stroke lesion segmentation from CTP images as well, showcasing the continuous evolution and applicability of deep learning techniques in medical imaging research.

The DEFUSE 3 (Endovascular Therapy Following Imaging Evaluation for Ischemic Stroke 3) (Albers *et al.*, 2018) and DAWN (Clinical Mismatch in the Triage of Wake-Up and Late Presenting Strokes Undergoing Neurointervention With Trevo) (Nogueira *et al.*, 2018) trials have underscored the remarkable efficacy of endovascular treatment for patients presenting 6 to 24 hours post-stroke onset. Notably, these trials primarily relied on CTP for patient selection, leading to the global uptick in the utilization of CTP across medical centers. The acquisition of dynamic CTP images necessitates subsequent postprocessing to gauge the extent of infarcted core and hypoperfused regions, which is crucial for treatment decision-making. Given the intricate nature of acute ischemic stroke lesion progression, leveraging data-driven machine learning techniques holds promise for enhancing core infarct estimation.

Annotated medical image datasets are essential for developing CAD methods by serving as training data for machine learning models, facilitating algorithm development and optimization, enabling performance evaluation and benchmarking, supporting generalization across diverse scenarios, and contributing to the clinical validation of CAD systems. The Ischemic stroke lesion segmentation challenge (ISLES) (Maier *et al.*, 2017), launched in 2015, encourages researchers to develop advanced stroke lesion detection tools. It tackles dataset inconsistencies and postprocessing variations by providing standardized datasets, enabling fair comparisons of emerging methods. With growing global

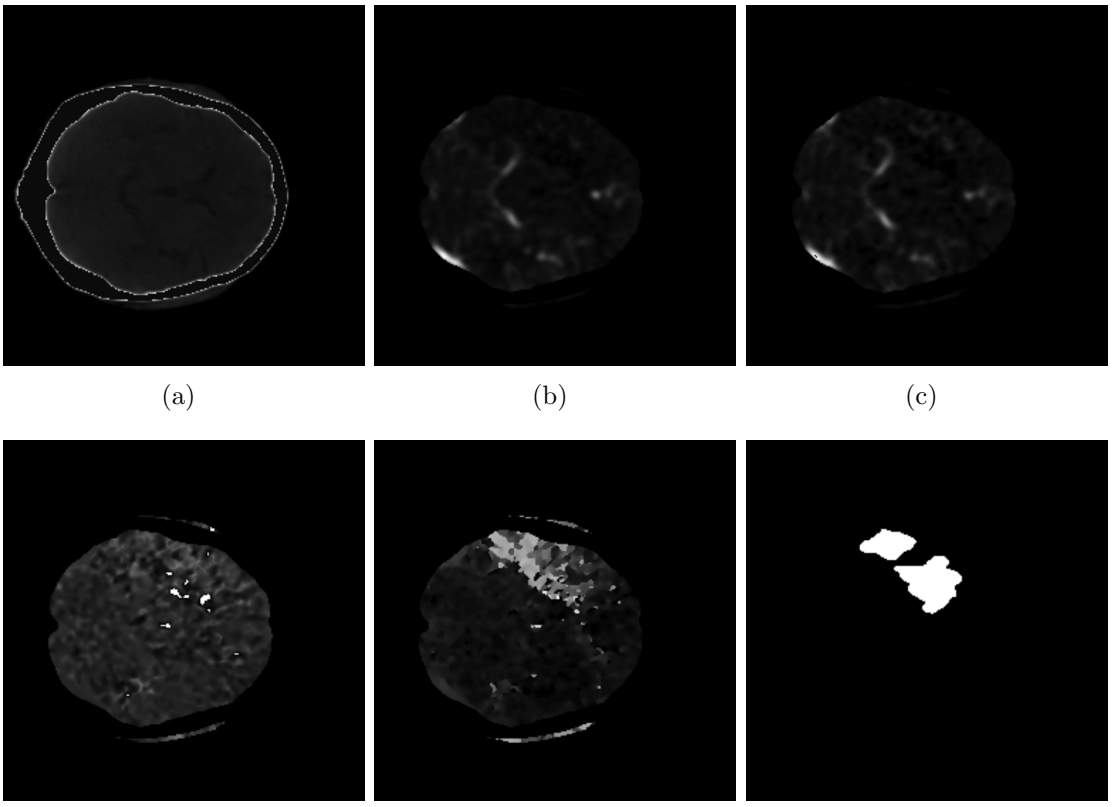


Figure 5.1: CT Perfusion maps provided in ISLES 2018 Challenge (a) Normal CT, (b) Cerebral Blood Flow (CBF) (c) Cerebral Blood Volume (CBV), (d) Mean Transit Time (MTT) (e) Time-to-maximum flow (Tmax), and (f) Ground truth.

participation, the challenge occurs annually alongside the Medical Image Computing and Computer-assisted Intervention (MICCAI) Conference. A sample of the ISLES 2018 dataset is shown in Figure 5.1.

The ISLES 2018 (isl; Cereda *et al.*, 2016; Hakim *et al.*, 2021) initiative was driven by insights from clinical trials and community input, with a primary objective of predicting the infarction from CTP imaging in patients with acute ischemic stroke. The challenge sought to segment the infarct core from CTP scans, which were annotated manually using Diffusion-Weighted Imaging (DWI) in subsequent MRI scans without any interventions or treatments between the CTP and DWI scans.

Most research efforts in ischemic stroke analysis have focused on the segmentation of stroke lesions from MR images. This is exemplified by the ISLES 2015–2017 challenges, which concentrated on multi-modal MR images, including T1, T1-contrast, FLAIR, and DWI sequences.

In recent years, there has been a growing trend toward utilizing deep learning methodologies in the analysis of Ischemic stroke, with many promising advancements documented in the scholarly literature. Dolz et al. (Dolz *et al.*, 2019) introduced a deep learning method using DWI and CTP to precisely segment ischemic stroke lesions, employing a densely connected UNet framework enriched with Inception modules. Similarly, Liu et al. (Liu *et al.*, 2021) introduced a novel deep learning architecture: 3D DAGMNet, specifically tailored for accurately detecting and segmenting lesions associated with acute and early subacute ischemic strokes in brain MRI scans. Although these methodologies have shown remarkable performance, they relied on MRI data, thus limiting the early analysis of stroke lesion segmentation using CTP images.

While research on ischemic stroke lesion segmentation from CTA or CTP perfusion parameter maps has been limited, the lower signal-to-noise ratio of CTP maps compared to DWI presents a significant challenge for automated segmentation. The ISLES 2018 dataset stands as a leading resource for CTP-based stroke segmentation, with several studies reporting successful segmentation models utilizing this data. The ISLES 2018 dataset includes baseline 4D CTP scans, along with derived CTP maps such as CBF, CBV, MTT, and Tmax. The dataset is divided into training and test sets, containing 94 and 62 cases, respectively. The images were acquired as slices with a variable number of axial slices (ranging from 2 to 22) with a standardized spacing of 5 mm. The spatial resolution of scans is $256 \times 256$.

Song et al. (Song, 2019) introduced an innovative framework for ischemic stroke lesion segmentation, comprising an extractor, a generator, and a segmentor. Initially, the extractor extracts representative feature images directly from CTP feature images. Subsequently, the generator utilizes the output from the extractor and perfusion param-

eters to generate pseudo-DWI images. Finally, the segmentor stage precisely segments the ischemic stroke lesion using the generated DWI images. Motivated by the work of Song et al. (Song, 2019), Wang et al. (Wang *et al.*, 2020*b*) proposed an improved version of a stroke lesion detection approach using a deep learning-based DWI synthesis method. To enhance DWI synthesis quality, their approach introduces a hybrid loss function that prioritizes lesion areas while promoting robust high-level contextual coherence. Subsequently, the lesion region is segmented within the synthesized pseudo-DWI using a segmentation network empowered by switchable normalization and channel calibration, thereby optimizing performance

Building on ISLES 2018 CTP data, Liu et al. Liu (2019) proposed a similar deep learning approach leveraging GANs to synthesize DWI images. This enabled them to segment ischemic stroke lesions directly on the generated DWI, employing a CNN architecture coupled with a custom loss function to balance positive and negative class imbalance during training. The challenge initially provides DWI images for the training dataset, which provides an additional advantage in generating intermediate DWI data for effective stroke segmentation. The top-ranking models mentioned above depend on both CTP and DWI data. However, access to DWI data along with CTP is not practical in real scenarios. Therefore, the challenge team restricts access to DWI data, leading to many studies in the literature that solely utilize CTP data for model development.

In a study, Chen et al. (Chen *et al.*, 2018*d*) introduced a 2.5D deep learning framework designed specifically for extracting and integrating information from various CTP modalities with high efficacy. This framework consists of an ensemble comprising multiple backbone networks such as U-net and demonstrates enhanced segmentation outcomes, thus contributing to greater segmentation performance in detecting ischemic stroke lesions. Furthermore, the study explored various data augmentation techniques to improve the overall performance of the framework. Clèrigues et al. (Clèrigues *et al.*, 2019) proposed a 2D patch-based deep learning approach designed to segment the acute stroke lesion core from CT perfusion images. The model utilizes an asymmetrical residual encoder-decoder

CNN architecture. The lesion core class constitutes approximately 5% of the brain tissue within the training dataset, and to mitigate this imbalance, the training process incorporates data augmentation using elastic deformation fields, dropout layers to introduce noise, and early stopping mechanisms.

In their study, Ghnemat et al. (Ghnemat *et al.*, 2023) introduced a mutation model integrated with a distance map within a GAN framework to generate synthetic medical data. This method employs Euclidean distance calculations to assess the intensity distance map by comparing each pixel and neighborhood. Following this, a threshold is used to identify adjacent areas with similar intensities for the mutation process. In the proposed supervised GAN, both the segmentation and discriminator components share the same CNN architecture, significantly reducing computational workload. Abulnaga et al. (Abulnaga and Rubin, 2019) introduced another 2D deep learning segmentation model based on PSPNet, leveraging pyramid pooling to incorporate global and local contextual information effectively. To capture the diverse shapes of lesions, they employed focal loss during network training, a specialized loss function that emphasizes learning from challenging samples.

The above-mentioned studies utilized 2D deep learning approaches for their model implementation due to the small training dataset, which comprises only around 500 slices from 94 cases. Generally, 3D models did not yield satisfactory results with such limited axial slices. However, this dataset provides a great opportunity for researchers to validate 3D deep learning models to work on such small yet reliable medical image datasets. Some of these 3D deep learning models for stroke segmentation have also been reported in the literature. Tursynova et al. (Tursynova and Omarov, 2021) made a pioneering attempt with a 3D deep learning model for segmenting stroke lesions from CTP images, employing a modified version of the 3D UNET architecture. However, due to the absence of optimizations needed to process the small, complex 3D data, the model's performance was limited, achieving only a 35% Dice score similarity in the test data evaluation.

Tureckova et al. (Tureckova and Rodríguez-Sánchez, 2019) introduced an improved

stroke segmentation framework that utilizes a modified 3D encoder-decoder network. They investigated the effectiveness of dilated convolutions in improving learning outcomes with a limited 3D dataset. Dilated convolutions expand the filter's receptive field, improving segmentation accuracy and increasing the Dice score to 37%. However, additional optimization is needed to tackle the segmentation challenges of small 3D medical image datasets.

Hu et al. (Hu *et al.*, 2018) introduced StrokeNet, a 3D residual framework designed for automated segmentation of ischemic stroke lesions from CTP images. This model targets voxel segmentation of stroke-affected areas within 3D perfusion CT scans. The segmentation framework employs a multi-level 3D refinement module that integrates local details and spatial-temporal context information using 3D convolutional layers, leading to substantial performance improvements. The training methodology incorporates curriculum learning, data augmentation techniques, and the Focal loss function to enhance model performance and address data imbalance challenges effectively. While the model demonstrated high accuracy on the training dataset, the article does not include the evaluation results for the test dataset.

This chapter explores a 3D segmentation model for stroke lesion segmentation which utilizes a 3D GAN. This approach leverages a 3D image-to-image translation network, specifically Attention-Vox2Vox, to achieve accurate segmentation results on volumetric data. The method uses two deep CNNs- a generator and a discriminator, that train concurrently over the CTP scans to generate 3D segmentation masks. Additionally, we optimize the generator and discriminator architectures to achieve enhanced detection and segmentation quality of stroke lesion cores from multi-channel CTP images. The key contributions outlined in this study are as follows:

1. ***Vox2Vox approach:*** The multichannel perfusion maps are processed simultaneously using a 3D GAN approach tailored for volumetric image segmentation.

2. ***Dual-task learning:*** The generator network integrates a dual-decoder stage

specifically tailored for dual-task learning, which effectively learns from binary ground truth as well as distance maps derived from ground truth masks to improve segmentation accuracy.

3. ***3D Attention network in the decoder stage:*** The proposed generator network uses a Dual Input 3D Convolutional block attention module (3D CBAM) to refine the feature maps in the decoding stage.

## 5.2 Methods

The stroke segmentation method proposed in this study utilizes a 3D image-to-image translation network known as Attention-Vox2Vox, drawing inspiration from the CNN model introduced by Cirillo et al. (Cirillo *et al.*, 2021). Attention-Vox2Vox builds upon the Pix2Pix model initially presented by Isola et al. (Isola *et al.*, 2017), extending it into the three-dimensional domain for enhanced performance in stroke segmentation tasks. Our segmentation framework incorporates two key supervised GAN components: a generator and a discriminator. This section provides a comprehensive overview of the methodology employed in our proposed Attention-Vox2Vox framework.

The proposed Attention-Vox2Vox framework leverages a tailored Vox2Vox architecture specifically crafted for addressing image segmentation challenges. The generator module is designed to process 3D slices from CTP maps and learn to generate 3D binary prediction masks that closely match the ground truth for stroke lesions. Concurrently, the discriminator module is trained to distinguish between the outputs generated by the generator and the actual ground truth data. In this segmentation model, the discriminator is trained using two distinct sets of 3D slices: one comprises pairs of 3D CTP slices and their corresponding 3D binary segmentation mask images, while the other set comprises pairs of 3D CTP slices and the 3D segmentation maps generated by the generator network.

### 5.2.1 Generator Network

In the conventional Pix2Pix and Vox2Vox models, the generator model employs the 2D and 3D versions of standard encoder-decoder networks, respectively. In contrast, our generator network incorporates an improved encoder-decoder model with residual modules and attention gates, surpassing several constraints of the traditional U-Net design. Moreover, it features a tailored loss function that addresses class imbalance concerns. Additionally, it incorporates a dual-decoder strategy for dual-task learning, enabling the model to delve into more details, such as the size and morphology of lesioned areas. These integrations significantly enhance the deep learning framework's ability for generalized learning. The proposed generator network is shown in Figure 5.2. This framework emphasizes key considerations at the network level, as outlined below.

#### 5.2.1.1 ResNet blocks to Enhance Training Efficiency.

The generator network utilizes a sequence of ResNet blocks at each depth, allowing for the extraction of multi-scale features without computational overload and enabling effective learning. Two residual modules are present in each depth, each of which contains two normal convolution layers that utilize identity mapping through short skip connections. Since there are four consecutive $3 \times 3 \times 3$ filters at each depth, it offers the advantage of assessing features across multiple scales without needing separate kernels. This approach reduces the computational burden and speeds up the learning process (Niyas *et al.*, 2021). Additionally, incorporating identity mappings addresses native challenges of deep neural networks, such as feature degradation and vanishing gradient issues, thereby improving their performance and stability.

#### 5.2.1.2 Selective and Shared Learning through 3D CBAM Attention

Our architecture incorporates a customized 3D CBAM attention mechanism. This module adeptly adjusts feature maps by dynamically emphasizing or diminishing their impor-

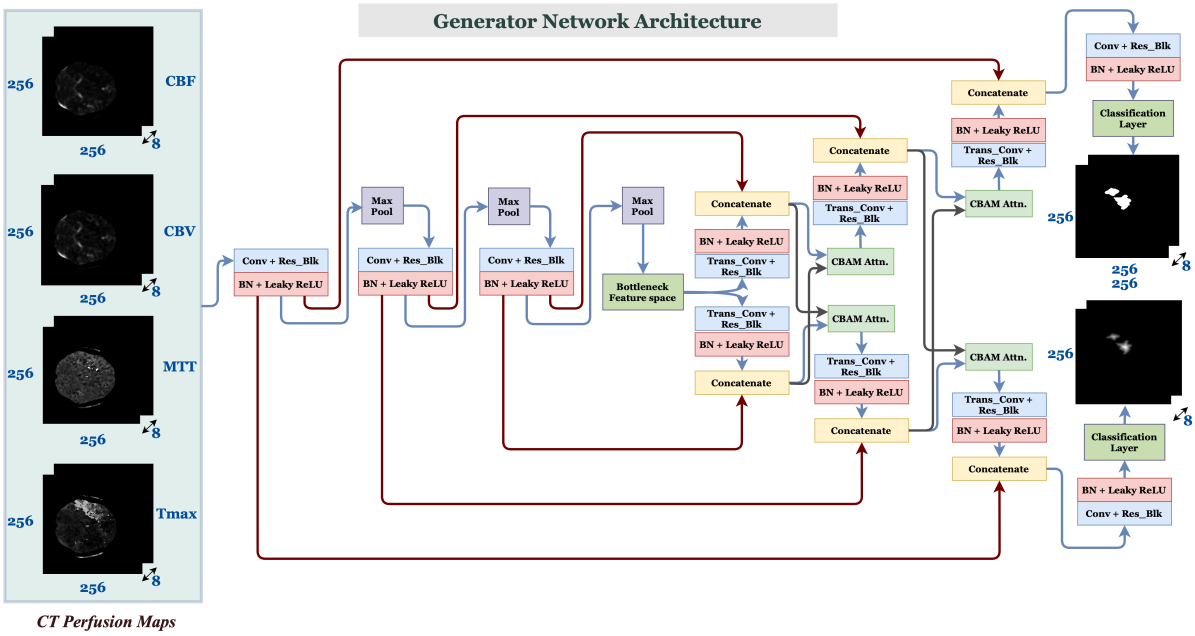



Figure 5.2: Proposed segmentation architecture for the Generator network

tance. In our dual-decoder configuration, the proposed attention blocks receive inputs from each decoder and analyze the feature maps concurrently to predict the distance maps and binary segmentation mask through shared learning.

#### 5.2.1.3 Dual-task Learning

The proposed generator network introduces a specialized dual-decoder phase designed for dual-task learning. This phase leverages distance maps generated from binary ground truth labels to enhance the learning process. The feature maps from the bottleneck layer undergo parallel processing by dual-decoder networks, with each decoder operating similarly to a conventional expanding path and receiving long skip connections from corresponding encoder layers.

The dual-task learning mechanism involves evaluating the error between the probability maps generated by each decoder layer with respect to the binary ground truth and its corresponding distance transform. The output of the first decoder is compared to the binary ground truth, ensuring equal emphasis on loss calculation across all lesion

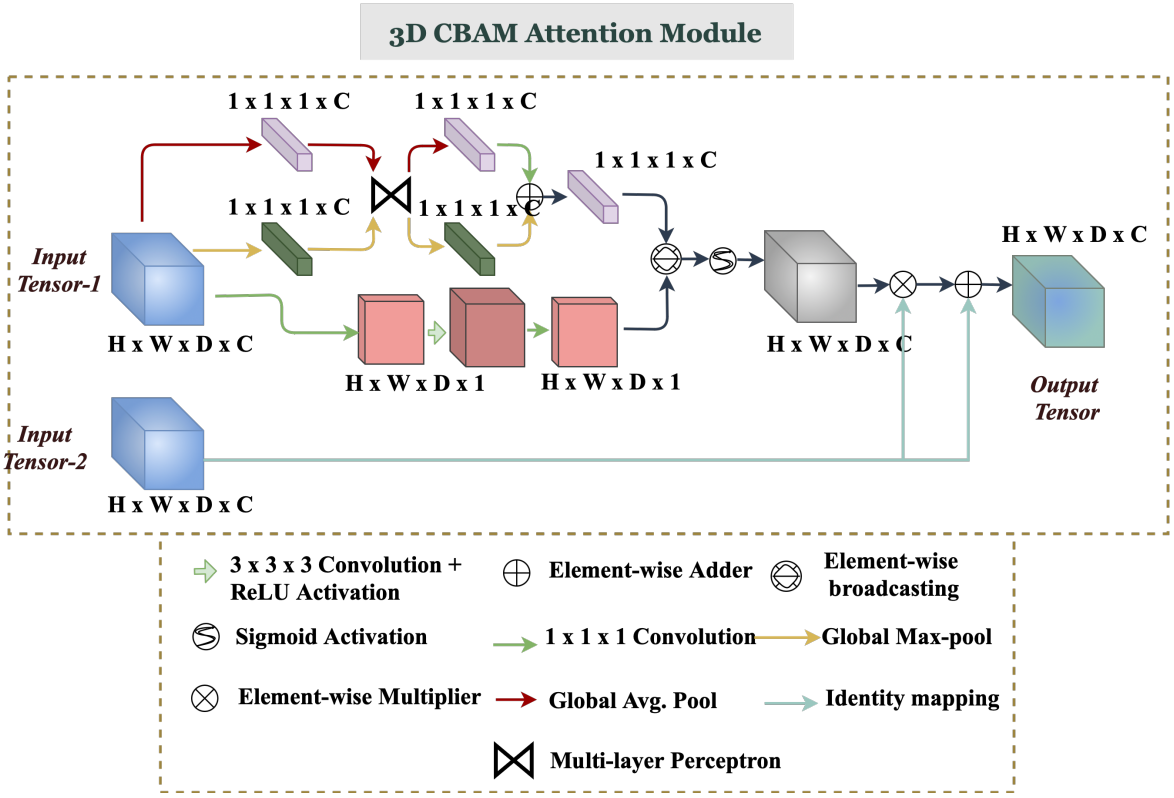



Figure 5.3: Proposed 3D CBAM Attention Module

regions. In contrast, the output of the second decoder is compared to the distance map, giving priority to the central lesion region over the lesion edges. We leverage a transformation function to generate the distance maps for dual-task learning. This function acts upon a binary input image, calculating the distance transform for each foreground (non-zero) element to its closest background pixel. The mathematical formulation for this transformation is presented in the following equation.

$$Y_{DM(i,j)} = \sqrt{\left(Y_{(i,j)} - b_{(i,j)}\right)^2} \tag{5.1}$$

where $Y_{(i,j)}$ is the binary ground truth image with pixel coordinates $(i, j)$, $b_{(i,j)}$ is the background point (value 0) with the smallest Euclidean distance to $Y_{i,j}$, and $Y_{DM}$ is the distance map image.

The proposed generator network for stroke segmentation network architecture is based

on a 3D CNN with residual connections. This enhanced architecture features a dual-decoder path, and the details are illustrated in Figure 5.2. At each encoder level, we incorporate a 3D convolutional layer, followed by two residual modules, each comprising a pair of 3D convolutional layers. Similarly, the decoder phase utilizes a transposed convolution layer, succeeded by two residual modules. All convolutional operations employ 3D kernels sized at $3 \times 3 \times 3$, and after the convolution layers, *batch normalization* and $LeakyReLU$ activation (Nair and Hinton, 2010) are applied. To reduce data dimensionality, a max-pooling operation of size $2 \times 2 \times 2$ is implemented after each level, halving the feature map dimensions. The number of filters in the subsequent encoder layers is doubled to counteract information loss from max-pooling.

In the initial layer, each convolution employs eight filters. As the depth increases, the number of filters doubles. In the decoder path, the predicted masks are compared with both the binary ground truth and the corresponding signed distance maps. The decoder path utilizes transposed convolution for upsampling feature maps, concatenating with high-resolution features from the encoder path. Dropout and regularization techniques are also integrated into the encoder and decoder layers to prevent overfitting.

### 5.2.2 Discriminator Network

Our proposed Attention Vox2Vox model incorporates a modified 3D CNN classification architecture as the discriminator network and is shown in Figure 5.4. This network takes two images as input: the generated one and the corresponding ground truth. Discriminator architecture consists of five convolutional layers, with the first three utilizing subsampling using strided convolutions. These initial layers progressively increase the number of filters, starting with 64, then 128, and finally 256. All convolutional layers employ a $4 \times 4 \times 4$ filter size. Following these layers, a single convolutional layer with 512 kernels is applied, succeeded by a convolutional layer with one kernel. The resulting feature map is taken for calculating the cost function.

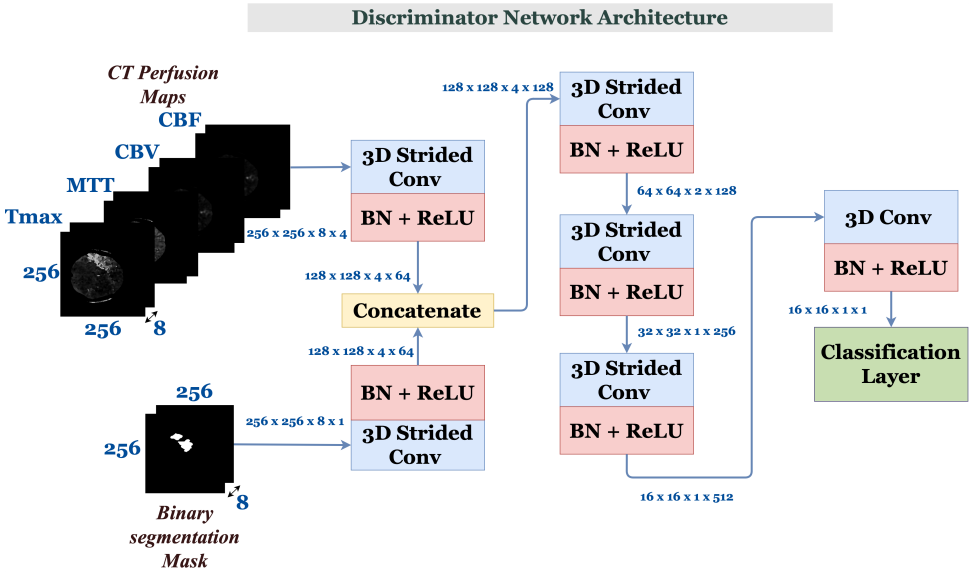



Figure 5.4: Proposed segmentation architecture for the Discriminator network

The discriminator outputs numerical values indicating the likelihood that the generated image closely resembles the actual image. The goal is to maximize the probability that the discriminator perceives the generated image as realistic. To achieve this, the discriminator undergoes training utilizing a binary cross-entropy loss function. This function quantifies the disparity between the discriminator output and the true labels, which is 0 for a fake image and 1 for a real image.

## 5.3 Results and Analysis

### 5.3.1 Hardware Details

The experiments detailed in this study were conducted on a workstation operating on the Ubuntu platform. The hardware setup included NVIDIA ® GeForce RTX 4090 ® GPU, equipped with 24 GB of dedicated graphics memory. The implementation of the models was achieved using the Keras and TensorFlow libraries and Python programming language.

### 5.3.2 Evaluation Metrics

In the ISLES 2018 challenge, the test data evaluation was conducted blindly to prevent any bias, and no ground truth information was disclosed to ensure fair evaluation and avoid any potential malpractices in result scoring. The ISLES 2018 challenge team used standard evaluation metrics such as Precision, Recall, the Dice Similarity Coefficient (DSC), Absolute Volume Difference (AVD), etc., to assess the performance of stroke lesion segmentation from CT perfusion scans. Precision measures the ratio of true positives to all predicted positives and provides insight into the accuracy of positive predictions. Conversely, Recall measures the fraction of true positives correctly identified within the whole positive class, offering a perspective on the model's ability to capture all relevant instances. The Dice coefficient is the harmonic mean of Precision and Recall, often utilized as the most important benchmark due to its incorporation of both Precision and Recall, providing a balanced assessment by considering both False Positives ($FP$) and False Negatives ($FN$). This metric is particularly valuable in evaluating the overlap between model predictions and ground truth data. Additionally, the AVD quantifies the absolute number of voxels that differ between the prediction segmentation maps and the ground truth. This metric contributes valuable information regarding the model's performance in accurately delineating structures within the data.

### 5.3.3 Datasets

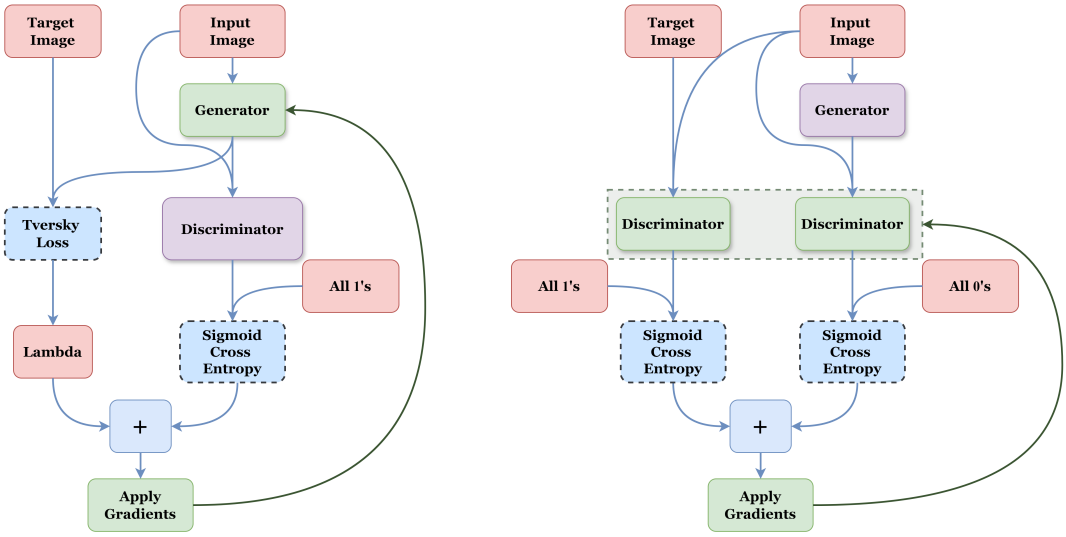
The study utilized the ISLES 2018 challenge dataset, which was organized into training and test sets consisting of 94 and 62 cases, respectively. The images consist of variable axial slices (ranging from 2 to 22) with a standardized spacing of 5 mm and a spatial resolution of $256 \times 256$. Each case includes CT, 4DPWI, CBF, CBV, MTT, and Tmax images. The train set is provided with corresponding binary ground truth in stroke lesion regions, while the test dataset is not provided with any ground truth, and the model validation on the test set is carried out blindly by the challenge team.

In this dataset, each input volume varies in dimensions from $256 \times 256 \times 2$ to $256 \times 256 \times 22$. However, our segmentation model adopts a shallow sliced approach with a data resolution of $256 \times 256 \times 8$. This approach is designed to balance segmentation accuracy and computational efficiency. During the preprocessing stage, we resize the number of slices in each volume to generate 3D slices of shape $256 \times 256 \times 8$. This is achieved through an adaptive slice repetition process. If a volume has fewer than eight slices, we repeat and concatenate the available slices to create a 3D slice with a depth of 8. Conversely, if a volume has more than eight slices, we normalize it to the nearest multiple of 8. For instance, a volume with four slices is repeated to form an 8-slice 3D object, while a volume with 20 slices is expanded to 24 slices.

We utilized random indexing within the challenge training dataset to assign training and validation subsets for our experiments. To ensure compatibility with the model's input dimensions of $256 \times 256 \times 8$, we generated overlapping 3D sub-volumes of this size with a one-voxel stride for both the training and validation sets. For the test data, non-overlapping sub-volumes were extracted from the preprocessed test dataset. The CBF, CBV, MTT, and Tmax perfusion maps are derived from the 4DPWI. Hence, to avoid redundancy, the current study uses CBF, CBV, MTT, and Tmax to segment the stroke lesion area as the lesion regions. After preprocessing and normalizing the number of slices in each case, the size of the 3D data samples has dimensions of $256 \times 256 \times 8 \times 4$, where 8 is the depth of the 3D sample, and 4 represents the four perfusion channels.

### 5.3.4 Training Methodology

Compared to a traditional CNN semantic segmentation network, our proposed Attention-Vox2Vox model introduces a significant enhancement by integrating the discriminator's function to boost the generator's performance. The Attention-Vox2Vox framework consists of two CNN networks- a generator and a discriminator, and both are trained concurrently to reinforce their learning efficiency. Figure 5.5 (a) represents the training methodology of the generator network.




(a) (b)

Figure 5.5: Training methodology of (a) the Generator, and (b) the Discriminator modules.

This generator training is characterized by the utilization of three distinct loss functions: the first one evaluates the error between the generated image (from the first decoder classification layer) and the desired target image, whereas the second loss function measures the error between the generated distance map (from the second decoder classification layer) and the distance map of the actual ground truth. The third loss function calculates the sigmoid cross-entropy loss between the generated images and a set of ones. The dataset utilized in this study exhibits a high imbalance between the foreground and background voxel classes. We adopt the Tversky loss function (Salehi *et al.*, 2017) as the $L_{Seg}$ to address this issue.

The overall generator loss is determined through a weighted aggregation of these loss components. Mathematically, this can be expressed as follows:

$$Loss_{Gen} = (\lambda_1 \times Loss_{Seg} + \lambda_2 \times Loss_{Dist} + Loss_{GAN}) \tag{5.2}$$

Here, $Loss_{Gen}$ represents the aggregate generator loss, encompassing $Loss_{Seg}$ (segmenta-

tion loss) and $Loss_{Dist}$ (distance map loss), which quantifies the Tversky loss between the generated image and ground truth and with their corresponding distance map pairs, respectively. $Loss_{GAN}$ denotes the *sigmoid cross-entropy* loss between the generated images and an array of ones.

The training pipeline of the discriminator network is shown in Figure 5.5 (b). The discriminator receives input from two distinct sets. The first set consists of perfusion maps and their corresponding segmentation maps generated by the generator network. In contrast, the second set comprises perfusion maps alongside the ground truth image. Training the discriminator involves optimizing its parameters based on the gradients of the discriminator loss, which is computed as the combination of two distinct loss functions corresponding to these input sets.

The discriminator loss $Loss_{Disc}$ computation can be expressed as follows:

$$Loss_{Disc} = (Loss_{Real} + Loss_{Fake}) \tag{5.3}$$

Here, the term $Loss_{Real}$ corresponds to the sigmoid cross-entropy loss between the discriminator output and an array of ones when processing the real image pair. Conversely, the $Loss_{Fake}$ component represents the sigmoid cross-entropy loss associated with the generated outcomes, matching the discriminator output to an array of zeros.

The proposed model is designed to process 3D perfusion slices of size $256 \times 256 \times 8 \times 4$. This configuration allows for simultaneous processing of the four channels of CT perfusion maps, enhancing the efficiency of the networks. To augment the training data and improve model robustness, we employed online data augmentation techniques, including rotation, which effectively increased the diversity of samples used for training. The choice of hyperparameters, including the network depth, the number of convolutional layers per depth, and the size and number of filters, was guided by rigorous experimentation and empirical results. We meticulously tuned these parameters to optimize the performance of our model across various evaluation metrics.

For model optimization, we utilized the *Adam* optimizer in conjunction with the Adagrad algorithm to adaptively adjust the model weights during training. Specifically, we initialized the model weights using the *HeNormal* (He *et al.*, 2015*b*) initialization and trained the model over 250 epochs. During training, we maintained a batch size of 4 and set the initial learning rate to 0.001, ensuring stable and effective convergence of the model. The parameters for each network model were chosen according to the highest validation dice score.

### 5.3.5 Results and Discussion

The ISLES 2018 dataset consists of a training dataset with annotated stroke lesions and a test set without any provided stroke annotation masks. Hence, in this study, we designed the experiments in two ways. First, we generated temporary training and validation sets from the official ISLES 2018 training dataset, which includes known annotations for stroke lesion regions. The model development and hyperparameter tuning were conducted at this stage. To achieve the optimized CNN design for segmenting stroke lesion regions, we performed a series of experiments with various architectural ablations and combinations.

Given that the primary components of the proposed model are the integration of supervised GAN-based training, dual-task learning, and dual-input 3D CBAM attention,

Table 5.1: Performance of the proposed Vox2Vox model with various ablations.

| **Method** | **Mean DSC** | **Mean Precision** | **Mean Recall** |
|---|---|---|---|
| **Baseline Vox2Vox** | 0.60 | 0.57 | 0.64 |
| **Baseline Vox2Vox + Dual-Task Learning** | 0.69 | 0.66 | 0.72 |
| **Baseline Vox2Vox + CBAM + Dual-Task Learning** | 0.73 | 0.71 | 0.76 |
| **Proposed Vox2Vox + CBAM + Dual-Task Learning** | 0.78 | 0.76 | 0.81 |

Table 5.2: Slice-wise segmentation performance of the proposed Vox2Vox model with various ablations.

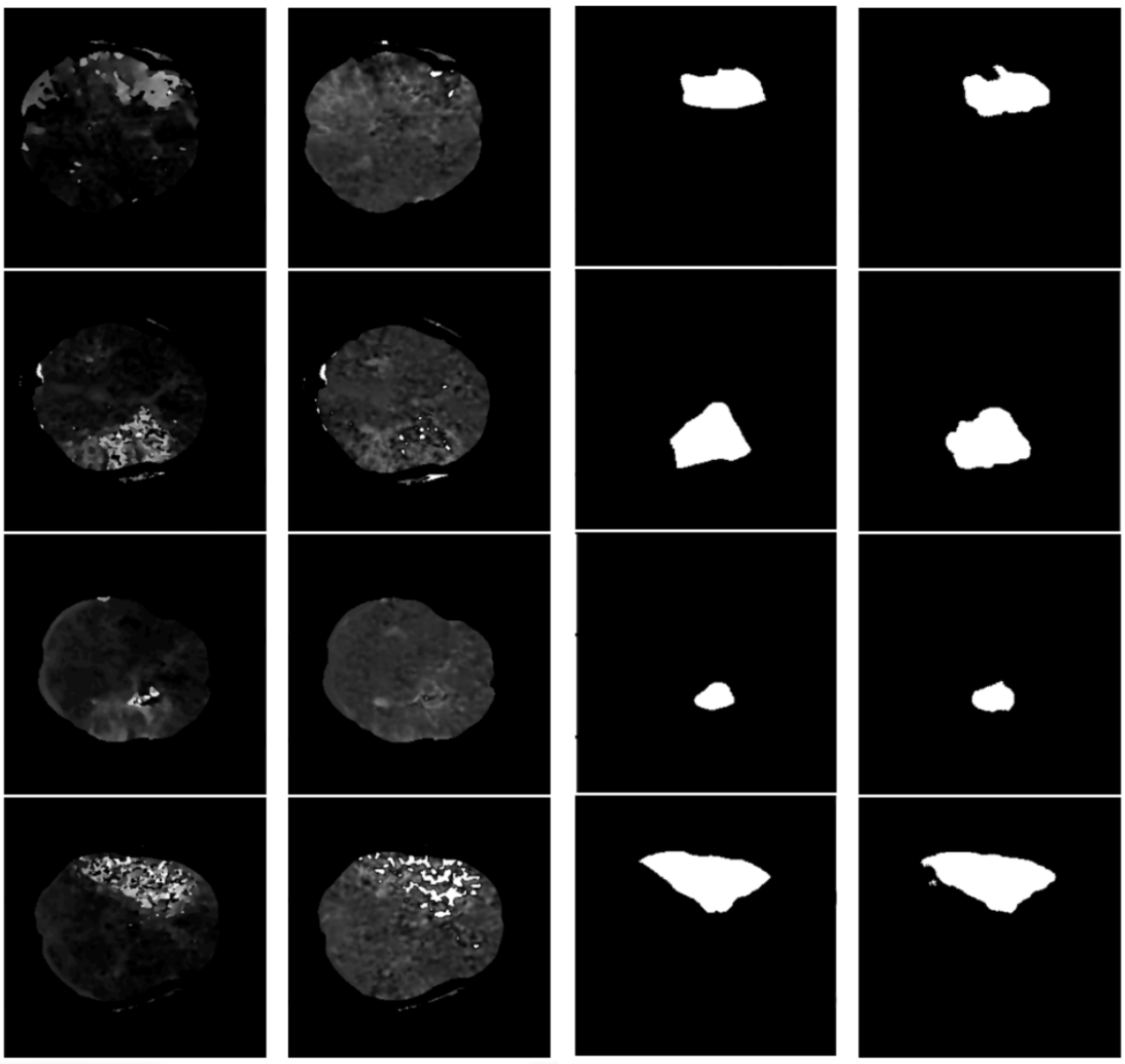
| Method | Mean DSC | Mean Precision | Mean Recall |
|---|---|---|---|
| **Baseline Vox2Vox** | 0.72 | 0.73 | 0.71 |
| **Baseline Vox2Vox + Dual-Task Learning** | 0.75 | 0.72 | 0.78 |
| **Baseline Vox2Vox + CBAM + Dual-Task Learning** | 0.83 | 0.84 | 0.82 |
| **Proposed Model (Vox2Vox + CBAM + Dual-Task Learning)** | 0.89 | 0.87 | 0.91 |

ablation studies were conducted with different combinations of these components. The pixel-wise and slice-wise segmentation performance of the top-performing ablation models is presented in Table 5.1 and 5.2.

The results shown in Table 5.1 represent the average pixel-wise segmentation performance across a 5-fold cross-validation while Table 5.2 represents the average slice-wise segmentation performance. These results substantiate the effectiveness of the incorporation of CBAM and dual-task learning for improving stroke lesion segmentation performance. The proposed model, featuring a dual-decoder generator module with residual convolutional layers, significantly enhances both recall and precision performance.

The qualitative assessment of the segmentation performance is depicted in Figure 5.6. The results demonstrate that the segmentation output of the proposed model closely aligns with the ground truth segmentation map.

The proposed model's quantitative analysis was further validated through a blind evaluation conducted by the ISLES challenge team, utilizing test data with unknown ground truth. Models that rank the first few positions in the challenge leaderboard benefit from both CTP and DWI data, giving them an edge in segmenting lesion regions with higher performance. For instance, models such as those by Song et al. (Song, 2019), Liu et al. (Liu, 2019), and Clérigues et al. (Clèrigues *et al.*, 2019) achieve average pixel-



(a) (b) (c) (d)

Figure 5.6: Qualitative analysis of the proposed model: (a) CBF, (b) CBV, (c) Ground truth, and (d) Predicted stroke lesion region using the proposed Vox2Vox model.

wise segmentation performances with Mean Dice similarity scores of 0.51, 0.49, and 0.49, respectively. However, their models use both CTP and DWI data for training, and hence, these models were excluded from direct comparison with models that rely solely on CTP data.

The pixel-wise segmentation performance of the proposed model is compared with

Table 5.3: Segmentation performance of the proposed model compared to best models reported in ISLES 2018 challenge leaderboard. The models marked with † denote 2D models developed with CTP scans, while those marked with # are the 3D segmentation models with CTP scans (without DWI data).

| Method | Mean DSC | Mean Precision | Mean Recall | Mean Absolute VD |
|---|---|---|---|---|
| **Chen et al.** (Chen *et al.*, 2018*d*) † | 0.48 | 0.59 | 0.46 | 10.59 |
| **Ghnemat et al.** (Ghnemat *et al.*, 2023) † | 0.42 | 0.48 | 0.43 | 11.49 |
| **Abulnaga et al.** (Abulnaga and Rubin, 2019) † | 0.44 | 0.49 | 0.53 | 10.18 |
| **Tursynova et al.** (Tursynova and Omarov, 2021) # | 0.35 | 0.41 | 0.38 | 13.61 |
| **Tureckova et al.** (Tureckova and Rodríguez-Sánchez, 2019) # | 0.37 | 0.44 | 0.44 | 24.95 |
| **Proposed Method** # | 0.48 | 0.56 | 0.48 | 11.02 |

state-of-the-art 2D and 3D models listed in the challenge's official leaderboard that uses only CTP scans and is given in Table 5.3. In the 2D context, there are approximately 500 samples for training, whereas developing a 3D model only has around 60 samples (converting slices to 3D slices of depth 8), which is normally insufficient for a robust 3D deep learning model. However, we addressed this issue by designing the 3D model to work effectively with such a small dataset. As a result, our model outperforms all other 3D models reported for stroke segmentation, solely utilizing CT perfusion maps without additional inputs like DWI data. We achieve competitive performance compared to state-of-the-art 2D methods and are superior to existing 3D CNN models, demonstrating the efficacy of our 3D Attention-Vox2Vox model.

## 5.4 Summary

This chapter explores the impact of a 3D CNN model on effectively segmenting complex medical tasks, such as ischemic stroke lesion segmentation from CT perfusion maps. The proposed model, a 3D Attention-Vox2Vox, leverages the advantages of supervised GAN for efficiently segmenting the lesion region. The segmentation framework incorporates several design considerations to enhance performance on small datasets. Key design contributions include generating 3D samples of shallow depth to extract inter-slice information without significant computational and GPU memory requirements, incorporating residual connections for faster multi-scale feature extraction, employing dual-decoder networks to learn from ground truth and distance maps simultaneously, integrating 3D Attention modules for maintaining consistency in the decoder pair, etc.

The proposed model demonstrated competitive performance compared to state-of-the-art 2D models incorporating additional data inputs such as DWI scans during development. Notably, our model currently ranks first among reported ischemic stroke lesion segmentation models using 3D CNNs and third among all segmentation models. Future experiments are planned to enhance segmentation performance by introducing a cascaded learning system capable of localizing and subsequently segmenting lesions with precise boundary information. Another area of future work involves integrating recent deep learning approaches, such as transformer modules, to achieve this in a 3D context.

# CHAPTER 6

# CONCLUSIONS AND FUTURE WORK

## 6.1 Summary of Contributions

This thesis has addressed critical gaps in the field of computer-aided medical image segmentation, particularly focusing on the challenges associated with conventional CNN-based approaches and the opportunities that arise with 3D deep learning techniques. The overarching goal of this research was to enhance the precision and efficiency of medical image segmentation, thereby enhancing clinical diagnostics and treatment planning processes.

A significant contribution of this thesis is the development of an efficient deep 3D CNN segmentation model that successfully harnesses the potential of 3D imaging data without the extensive computational demands typically associated with such models. By introducing a shallow sliced stacking approach and incorporating residual connections, the model achieves high segmentation accuracy while maintaining computational efficiency and reduced model complexity. Further, this work presented an advanced 3D deep learning model employing a multi-view dual encoder-decoder architecture. This model leverages multi-modality data and incorporates architectural enhancements such as end-to-end cascaded modules and 3D Attention modules, improving the detection of focal cortical dysplasia (FCD) lesions in neuro-radiology.

Additionally, the application of Generative Adversarial Networks (GANs) in the Vox2Vox CNN network represents a pioneering approach to segmenting acute stroke lesion cores in computed tomography perfusion (CTP) scans. This model highlights the potential of supervised GAN to refine the quality of segmentation outputs, further extending the capabilities of 3D deep learning in medical imaging.

## 6.2 Implications for Clinical Practice

The methodologies developed and validated in this thesis hold substantial implications for clinical practice with suitable refinement and clinical trials. By improving the accuracy and efficiency of 3D medical image segmentation, these models can assist medical professionals in making more informed decisions quicker. This is particularly crucial in emergency medical situations, such as stroke, where time is critical.

## 6.3 Directions for Future Research

While this thesis has made significant strides in the application of 3D deep learning for medical image segmentation, several things remain open for future research:

**Scalability and Generalization:** Replicating the proposed models on a different dataset while maintaining the same architectural design poses several challenges due to variations in scan types and data quality between datasets. Hence, further research is needed to test the scalability of the proposed models across larger and more diverse datasets, which would help in assessing the generalizability of the models across different medical settings and populations.

**Integration with Clinical Workflows:** Future work could focus on the integration of these advanced segmentation tools into clinical workflows, including real-time processing capabilities and user-friendly interfaces for clinical use. Additionally, extensive clinical validation with diverse datasets is necessary to substantiate the proposed models for medical applications.

**Expanding Modality Coverage:** The performance of the proposed models on different datasets may vary due to factors such as differences in data distribution, noise levels, or domain-specific features. Additionally, dataset-specific biases during model training could lead to issues such as overfitting or underfitting when applied to new data. Hence, Extending the current models to include other imaging modalities and

pathological conditions could broaden the applicability of 3D deep learning in medical diagnostics.

**Advanced Computational Techniques:** Exploring more sophisticated computational strategies, such as federated learning, could address privacy concerns and data accessibility issues, facilitating wider adoption of advanced AI models in healthcare.

## 6.4 Concluding Remarks

In conclusion, this thesis not only advances the field of medical image segmentation through innovative 3D deep learning approaches but also sets the stage for future research that could further revolutionize the capabilities of computer-aided diagnosis. The integration of these advanced technologies into clinical practice promises to enhance the diagnostic process, offering a significant impact on patient care and treatment efficacy.

# LIST OF PAPERS BASED ON THESIS

## 6.5 Journal Publications (Within the Scope of Thesis)

1. **S. Niyas**, S. Chethana Vaisali, Iwrin Show, T.G. Chandrika, S. Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan. "Segmentation of Focal Cortical Dysplasia Lesions from Magnetic Resonance Images using 3D Convolutional Neural Networks". Biomedical Signal Processing and Control, 70, 102951 (2021).

2. **S. Niyas**, S.J. Pawan, M. Anand Kumar, and Jeny Rajan, "Medical image segmentation with 3D convolutional neural networks: A Survey". Neurocomputing, 493, 397-413 (2022)

3. **S. Niyas**, Chandrasekharan Kesavadas, and Jeny Rajan. "A Dual Encoder-Decoder Multi-task 3D Deep Learning Framework for the Segmentation of Focal Cortical Dysplasia Lesions." Biomedical Signal Processing and Control, (2024) (Under Review).

4. **S. Niyas**, Vivek A. Saraf, Ajith Abraham, Neethi A. S. and Jeny Rajan. "Segmentation of Ischemic Stroke Lesions from CT Perfusion images using 3D Attention-Driven Vox-2-Vox." Journal of Medical Imaging (2024) (Under Review).

## 6.6 Supplementary Journal Publications

1. **S. Niyas**, Ramya Bygari, Rachita Naik, Bhavishya Viswanath, Dhananjay Ugwekar, Tojo Mathew, J. Kavya, Jyoti R. Kini, and Jeny Rajan. "Automated Molecular Subtyping of Breast Carcinoma Using Deep Learning Techniques." IEEE Journal of Translational Engineering in Health and Medicine 11 (2023): 161-169.

2. Neethi. A. S, **S. Niyas**., Santhosh Kumar Kannath, Jimson Mathew, Ajimi Mol Anzar, and Jeny Rajan. "Stroke classification from computed tomography scans using 3d convolutional neural network." **Biomedical Signal Processing and Control** 76 (2022): 103720.

3. Tojo Mathew, **S. Niyas**, C. I. Johnpaul, Jyoti R. Kini, and Jeny Rajan. "A novel deep classifier framework for automated molecular subtyping of breast carcinoma using

immunohistochemistry image analysis." **Biomedical Signal Processing and Control** 76 (2022): 103657.

4. Thomas, Edwin, S. J. Pawan, Shushant Kumar, Anmol Horo, **S. Niyas**, S. Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan. "Multi-res-attention UNet: a CNN model for the segmentation of focal cortical dysplasia lesions from magnetic resonance images." **IEEE Journal of Biomedical and Health Informatics** 25, no. 5 (2020): 1724-1734.

5. K. M Bijay Dev, Pawan S. Jogi, **S. Niyas**, S. Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan. "Automatic detection and localization of focal cortical dysplasia lesions in MRI using fully convolutional neural network." **Biomedical Signal Processing and Control** 52 (2019): 218-225.

## 6.7 Supplementary Conference Publications

1. **S. Niyas**, Shraddha Priya, Reena Oswal, Tojo Mathew, Jyoti R. Kini, and Jeny Rajan. "Automated Molecular Subtyping of Breast Cancer Through Immunohistochemistry Image Analysis." In Computer Vision and Machine Intelligence: Proceedings of CVMI 2022, pp. 23-35. Singapore: Springer Nature Singapore, (2023).

# REFERENCES


(). ISLES: Ischemic Stroke Lesion Segmentation Challenge 2018 — isles-challenge.org. https://www.isles-challenge.org/ISLES2018. [Accessed 23-03-2024]. 75

Fcd. 2020. https://radiopaedia.org/articles/focal-cortical-dysplasia/ (accessed 25 June 2020). 29

**Abulnaga, S. M.** and **J. Rubin**, Ischemic stroke lesion segmentation in ct perfusion scans using pyramid pooling and focal loss. *In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*. Springer, 2019. 78, 93

**Adler, S.**, **K. Wagstyl**, **R. Gunny**, **L. Ronan**, **D. Carmichael**, **J. H. Cross**, **P. C. Fletcher**, and **T. Baldeweg** (2017). Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy. *NeuroImage: Clinical*, **14**, 18 – 27. ISSN 2213-1582. URL http://www.sciencedirect.com/science/article/pii/S2213158216302674. 30

**Ahmed, B.**, **C. E. Brodley**, **K. E. Blackmon**, **R. Kuzniecky**, **G. Barash**, **C. Carlson**, **B. T. Quinn**, **W. Doyle**, **J. French**, **O. Devinsky**, and **T. Thesen** (2015). Cortical feature analysis and machine learning improves detection of "mri-negative" focal cortical dysplasia. *Epilepsy & Behavior*, **48**, 21 – 28. ISSN 1525-5050. URL http://www.sciencedirect.com/science/article/pii/S1525505015002322. 31

**Albers, G. W.**, **M. P. Marks**, **S. Kemp**, **S. Christensen**, **J. P. Tsai**, **S. Ortega-Gutierrez**, **R. A. McTaggart**, **M. T. Torbey**, **M. Kim-Tenser**, **T. Leslie-Mazwi**, *et al.* (2018). Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *New England Journal of Medicine*, **378**(8), 708–718. 74

**Alexandre Jr, V.**, **R. Walz**, **M. M. Bianchin**, **T. R. Velasco**, **V. C. Terra-Bustamante**, **L. Wichert-Ana**, **D. Araújo Jr**, **H. R. Machado**, **J. A. Assirati Jr**, **C. G. Carlotti Jr**, *et al.* (2006). Seizure outcome after surgery for epilepsy due to focal cortical dysplastic lesions. *Seizure*, **15**(6), 420–427. 28

**Antel, S. B.**, **A. Bernasconi**, **N. Bernasconi**, **D. L. Collins**, **R. E. Kearney**, **R. Shinghal**, and **D. L. Arnold** (2002). Computational models of mri characteristics of focal cortical dysplasia improve lesion detection. *Neuroimage*, **17**(4), 1755–1760. 29

**Armato, S.**, **G. McLennan**, **L. Bidaut**, **M. McNitt-Gray**, **C. Meyer**, **A. Reeves**, **B. Zhao**, **D. Aberle**, **C. Henschke**, **E. Hoffman**, **E. Kazerooni**, **H. MacMahon**, **E. J. R. V. Beeke**, **D. Yankelevitz**, **A. Biancardi**, **P. Bland**, **M. Brown**,

**R. M. Engelmann**, **G. Laderach**, **D. Max**, **R. C. Pais**, **D. Qing**, **R. Roberts**, **A. R. Smith**, **A. Starkey**, **P. Batrah**, **P. Caligiuri**, **A. O. Farooqi**, **G. Gladish**, **C. Jude**, **R. Munden**, **I. Petkovska**, **L. Quint**, **L. Schwartz**, **B. Sundaram**, **L. Dodd**, **C. Fenimore**, **D. Gur**, **N. Petrick**, **J. Freymann**, **J. Kirby**, **B. Hughes**, **A. V. Casteele**, **S. Gupte**, **M. Sallamm**, **M. D. Heath**, **M. Kuhn**, **E. Dharaiya**, **R. Burns**, **D. Fryd**, **M. Salganicoff**, **V. Anand**, **U. Shreter**, **S. Vastagh**, and **B. Y. Croft** (2011). The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics*, **38 2**, 915–31. 26

**Bakas, S.**, **M. Reyes**, **A. Jakab**, **S. Bauer**, **M. Rempfler**, **A. Crimi**, **R. T. Shinohara**, **C. Berger**, **S. M. Ha**, **M. Rozycki**, *et al.* (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*. 25

**Bergo, F. P. G.**, **A. X. Falcao**, **C. L. Yasuda**, and **F. Cendes**, Fcd segmentation using texture asymmetry of mr-t1 images of the brain. *In 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2008. 29

**Berthelot, D.**, **N. Carlini**, **I. Goodfellow**, **N. Papernot**, **A. Oliver**, and **C. Raffel** (2019). Mixmatch: A holistic approach to semi-supervised learning. *Advances in Neural Information Processing Systems*, **32**. 20

**Bui, T. D.**, **J. Shin**, and **T. Moon** (2017). 3d densely convolutional networks for volumetric segmentation. *arXiv preprint arXiv:1709.03199*. 12

**Cereda, C. W.**, **S. Christensen**, **B. C. Campbell**, **N. K. Mishra**, **M. Mlynash**, **C. Levi**, **M. Straka**, **M. Wintermark**, **R. Bammer**, **G. W. Albers**, *et al.* (2016). A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a dwi standard. *Journal of Cerebral Blood Flow & Metabolism*, **36**(10), 1780–1789. 75

**Chakraborty, S.**, **S. Chatterjee**, **A. S. Ashour**, **K. Mali**, and **N. Dey**, Intelligent computing in medical imaging: A study. *In Advancements in applied metaheuristic computing*. IGI global, 2018, 143–163. 1

**Chen, H.**, **Q. Dou**, **L. Yu**, and **P.-A. Heng** (2016). Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation. *arXiv preprint arXiv:1608.05895*. 22

**Chen, H.**, **Q. Dou**, **L. Yu**, **J. Qin**, and **P.-A. Heng** (2018*a*). Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images. *NeuroImage*, **170**, 446–455. 13

**Chen, L.**, **Y. Wu**, **A. M. DSouza**, **A. Z. Abidin**, **A. Wismüller**, and **C. Xu**, Mri tumor segmentation with densely connected 3d cnn. *In Medical Imaging 2018: Image Processing*, volume 10574. International Society for Optics and Photonics, 2018*b*. 15

**Chen, S.**, **K. Ma**, and **Y. Zheng** (2019). Med3d: Transfer learning for 3d medical image analysis. *ArXiv*, **abs/1904.00625**. 26

**Chen, W.**, **B. Liu**, **S. Peng**, **J. Sun**, and **X. Qiao**, S3d-unet: Separable 3d u-net for brain tumor segmentation. *In BrainLes@MICCAI*. 2018*c*. 25

**Chen, Y.**, **Y. Li**, and **Y. Zheng**, Ensembles of modalities fused model for ischemic stroke lesion segmentation. *In International MICCAI Brainlesion Workshop*. 2018*d*. 77, 93

**Chollet, F.** (2017). Xception: Deep learning with depthwise separable convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800–1807. 13

**Çiçek, Ö.**, **A. Abdulkadir**, **S. S. Lienkamp**, **T. Brox**, and **O. Ronneberger**, 3d u-net: learning dense volumetric segmentation from sparse annotation. *In International conference on medical image computing and computer-assisted intervention*. Springer, 2016. 18

**Ciompi, F.**, **K. Chung**, **S. J. Riel**, **A. A. A. Setio**, **P. Gerke**, **C. Jacobs**, **E. T. Scholten**, **C. Schaefer-Prokop**, **M. Wille**, **A. Marchianó**, **U. Pastorino**, **M. Prokop**, and **B. Ginneken** (2017). Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Scientific Reports*, **7**. 23

**Cirillo, M. D.**, **D. Abramian**, and **A. Eklund**, Vox2vox: 3d-gan for brain tumour segmentation. *In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6*. Springer, 2021. 80

**Clèrigues, A.**, **S. Valverde**, **J. Bernal**, **J. Freixenet**, **A. Oliver**, and **X. Lladó** (2019). Acute ischemic stroke lesion core segmentation in ct perfusion images using fully convolutional neural networks. *Computers in biology and medicine*, **115**, 103487. 77, 91

**Colliot, O.**, **T. Mansi**, **N. Bernasconi**, **V. Naessens**, **D. Klironomos**, and **A. Bernasconi**, Segmentation of focal cortical dysplasia lesions using a feature-based level set. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2005. 30

**Colliot, O.**, **T. Mansi**, **N. Bernasconi**, **V. Naessens**, **D. Klironomos**, and **A. Bernasconi** (2006). Segmentation of focal cortical dysplasia lesions on mri using level set evolution. *Neuroimage*, **32**(4), 1621–1630. 30

**Cortes, C.**, **M. Mohri**, and **A. Rostamizadeh** (2012). L2 regularization for learning kernels. *arXiv preprint arXiv:1205.2653*. 69

**Crow, T. J.**, **J. Ball**, **S. R. Bloom**, **R. Brown**, **C. J. Bruton**, **N. Colter**, **C. D. Frith**, **E. C. Johnstone**, **D. G. C. Owens**, and **G. W. Roberts** (1989). Schizophrenia as an Anomaly of Development of Cerebral Asymmetry: A Postmortem Study and a Proposal Concerning the Genetic Basis of the Disease. *Archives of General Psychiatry*, **46**(12), 1145–1150. ISSN 0003-990X. URL `https://doi.org/10.1001/archpsyc.1989.01810120087013`. 29

**Demeestere, J.**, **A. Wouters**, **S. Christensen**, **R. Lemmens**, and **M. G. Lansberg** (2020). Review of perfusion imaging in acute ischemic stroke: from time to tissue. *Stroke*, **51**(3), 1017–1024. 73

**Dev, K. B.**, **P. S. Jogi**, **S. Niyas**, **S. Vinayagamani**, **C. Kesavadas**, and **J. Rajan** (2019). Automatic detection and localization of focal cortical dysplasia lesions in mri using fully convolutional neural network. *Biomedical Signal Processing and Control*, **52**, 218 – 225. ISSN 1746-8094. URL `http://www.sciencedirect.com/science/article/pii/S1746809419301211`. 32, 44, 45, 46, 47, 48, 53

**Dice, L. R.** (1945). Measures of the amount of ecologic association between species. *Ecology*, **26**(3), 297–302. 39

**Dingledine, R.** and **B. Hassel**, A new approach for epilepsy. *In Cerebrum: the Dana forum on brain science*, volume 2016. Dana Foundation, 2016. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4938265/`. 28

**Dolz, J.**, **I. Ben Ayed**, and **C. Desrosiers**, Dense multi-path u-net for ischemic stroke lesion segmentation in multiple image modalities. *In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*. Springer, 2019. 76

**Donahue, J.** and **M. Wintermark** (2015). Perfusion ct and acute stroke imaging: foundations, applications, and literature review. *Journal of Neuroradiology*, **42**(1), 21–29. 73

**Dou, Q.**, **L. Yu**, **H. Chen**, **Y. Jin**, **X. Yang**, **J. Qin**, and **P.-A. Heng** (2017). 3d deeply supervised network for automated segmentation of volumetric medical images. *Medical image analysis*, **41**, 40–54. 12

**El Azami, M.**, **A. Hammers**, **N. Costes**, and **C. Lartizien**, Computer aided diagnosis of intractable epilepsy with mri imaging based on textural information. *In 2013 International Workshop on Pattern Recognition in Neuroimaging*. IEEE, 2013. 31

**Feng, C.**, **H. Zhao**, **M. Tian**, **M. Lu**, and **J. Wen** (2020*a*). Detecting focal cortical dysplasia lesions from flair-negative images based on cortical thickness. *BioMedical Engineering OnLine*, **19**(1), 13. ISSN 1475-925X. URL `https://doi.org/10.1186/s12938-020-0757-8`. 29

**Feng, C.**, **H. Zhao**, **M. Tian**, **M. Lu**, and **J. Wen** (2020*b*). Detecting focal cortical dysplasia lesions from flair-negative images based on cortical thickness. *Biomedical engineering online*, **19**, 1–15. 58

**Feng, C.**, **H. Zhao**, **J. Zhang**, **Z. Cheng**, and **J. Wen**, Automated localization of epileptic focus using convolutional neural network. *In Proceedings of the 2020 2nd International Conference on Big Data Engineering and Technology*, BDET 2020. Association for Computing Machinery, New York, NY, USA, 2020*c*. ISBN 9781450376839. URL `https://doi.org/10.1145/3378904.3378928`. 31, 52

**Focke, N. K.**, **M. R. Symms**, **J. L. Burdett**, and **J. S. Duncan** (2008). Voxel-based analysis of whole brain flair at 3t detects focal cortical dysplasia. *Epilepsia*, **49**(5), 786–793. 31

**Ge, R.**, **G. Yang**, **Y. Chen**, **L. Luo**, **C. Feng**, **H. Ma**, **J. Ren**, and **S. Li** (2019). K-net: Integrate left ventricle segmentation and direct quantification of paired echo sequence. *IEEE transactions on medical imaging*, **39**(5), 1690–1702. 17

**Ghnemat, R.**, **A. Khalil**, and **Q. Abu Al-Haija** (2023). Ischemic stroke lesion segmentation using mutation model and generative adversarial network. *Electronics*, **12**(3), 590. 78, 93

**Gill, R. S.**, **B. Caldairou**, **N. Bernasconi**, and **A. Bernasconi**, Uncertainty-informed detection of epileptogenic brain malformations using bayesian neural networks. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019. 31, 53

**Gill, R. S.**, **S.-J. Hong**, **F. Fadaie**, **B. Caldairou**, **B. C. Bernhardt**, **C. Barba**, **A. Brandt**, **V. C. Coelho**, **L. d'Incerti**, **M. Lenge**, *et al.*, Deep convolutional networks for automated detection of epileptogenic brain malformations. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018. 31, 53

**Gillebert, C. R.**, **G. W. Humphreys**, and **D. Mantini** (2014). Automated delineation of stroke lesions using brain ct images. *NeuroImage: Clinical*, **4**, 540–548. 73

**Glorot, X.** and **Y. Bengio**, Understanding the difficulty of training deep feedforward neural networks. *In Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 2010. 43

**Goodfellow, I. J.**, **J. Pouget-Abadie**, **M. Mirza**, **B. Xu**, **D. Warde-Farley**, **S. Ozair**, **A. C. Courville**, and **Y. Bengio**, Generative adversarial nets. *In NIPS*. 2014. 18

**Gordaliza, P. M.**, **J. J. Vaquero**, **S. Sharpe**, **F. Gleeson**, and **A. Munoz-Barrutia** (2019). A multi-task self-normalizing 3d-cnn to infer tuberculosis radiological manifestations. *arXiv preprint arXiv:1907.12331*. 17

**Guo, Y.**, **Y. Gao**, and **D. Shen** (2015). Deformable mr prostate segmentation via deep feature learning and sparse patch matching. *IEEE transactions on medical imaging*, **35**(4), 1077–1089. URL `https://www.sciencedirect.com/science/article/pii/B9780128104088000122`. 52

**Hakim, A.**, **S. Christensen**, **S. Winzeck**, **M. G. Lansberg**, **M. W. Parsons**, **C. Lucas**, **D. Robben**, **R. Wiest**, **M. Reyes**, and **G. Zaharchuk** (2021). Predicting infarct core from computed tomography perfusion in acute ischemia with machine learning: lessons from the isles challenge. *Stroke*, **52**(7), 2328–2337. 75

**Hauptman, J. S.** and **G. W. Mathern** (2012). Surgical treatment of epilepsy associated with cortical dysplasia: 2012 update. *Epilepsia*, **53**, 98–104. 28

**He, K.**, **X. Zhang**, **S. Ren**, and **J. Sun**, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *In Proceedings of the IEEE international conference on computer vision*. 2015*a*. 69

**He, K.**, **X. Zhang**, **S. Ren**, and **J. Sun** (2015*b*). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 1904–1916. 13, 90

**Heinrich, M.**, **O. Oktay**, and **N. Bouteldja** (2019). Obelisk-net: Fewer layers to solve 3d multi-organ segmentation with sparse deformable convolutions. *Medical Image Analysis*, **54**, 1–9. 24

**Hu, X.**, **W. Huang**, **S. Guo**, and **M. R. Scott**, Strokenet: 3d local refinement network for ischemic stroke lesion segmentation. *In Int. MICCAI Brainlesion Workshop*. 2018. 79

**Huang, G.**, **Z. Liu**, and **K. Q. Weinberger** (2017). Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269. 23

**Ioffe, S.** and **C. Szegedy**, Batch normalization: Accelerating deep network training by reducing internal covariate shift. *In International conference on machine learning*. PMLR, 2015. 2, 38, 65

**Isola, P.**, **J.-Y. Zhu**, **T. Zhou**, and **A. A. Efros**, Image-to-image translation with conditional adversarial networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. 80

**Jahan, R.**, **J. L. Saver**, **L. H. Schwamm**, **G. C. Fonarow**, **L. Liang**, **R. A. Matsouaka**, **Y. Xian**, **D. N. Holmes**, **E. D. Peterson**, **D. Yavagal**, *et al.* (2019). Association between time to treatment with endovascular reperfusion therapy and outcomes in patients with acute ischemic stroke treated in clinical practice. *Jama*, **322**(3), 252–263. 72

**Jin, B.**, **B. Krishnan**, **S. Adler**, **K. Wagstyl**, **W. Hu**, **S. Jones**, **I. Najm**, **A. Alexopoulos**, **K. Zhang**, **J. Zhang**, **M. Ding**, **S. Wang**, **the Pediatric Imaging, Neurocognition, and Genetics Study** , and **Z. I. Wang** (2018). Automated detection of focal cortical dysplasia type ii with surface-based magnetic resonance imaging postprocessing and machine learning. *Epilepsia*, **59**(5), 982–992. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.14064`. 29

**Kabat, J.** and **P. Król** (2012). Focal cortical dysplasia–review. *Polish journal of radiology*, **77**(2), 35. 29

**Kamnitsas, K.**, **W. Bai**, **E. Ferrante**, **S. McDonagh**, **M. Sinclair**, **N. Pawlowski**, **M. Rajchl**, **M. Lee**, **B. Kainz**, **D. Rueckert**, *et al.*, Ensembles of multiple models and architectures for robust brain tumour segmentation. *In International MICCAI Brainlesion Workshop*. Springer, 2017*a*. 15

**Kamnitsas, K.**, **E. Ferrante**, **S. Parisot**, **C. Ledig**, **A. V. Nori**, **A. Criminisi**, **D. Rueckert**, and **B. Glocker**, Deepmedic for brain tumor segmentation. *In International workshop on Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries*. Springer, 2016. 15

**Kamnitsas, K.**, **C. Ledig**, **V. F. Newcombe**, **J. P. Simpson**, **A. D. Kane**, **D. K. Menon**, **D. Rueckert**, and **B. Glocker** (2017*b*). Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, **36**, 61–78. 15

**Karpathy, A.** *et al.* (2016). Cs231n convolutional neural networks for visual recognition. *Neural networks*, **1**, 1. 34

**Kayalibay, B.**, **G. Jensen**, and **P. van der Smagt** (2017). Cnn-based segmentation of medical imaging data. *arXiv preprint arXiv:1701.03056*. 12

**Kingma, D. P.** and **J. Ba** (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. 43, 69

**Kitrungrotsakul, T.**, **X.-H. Han**, **Y. Iwamoto**, **L. Lin**, **A. H. Foruzan**, **W. Xiong**, and **Y.-W. Chen** (2019). Vesselnet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, **75**, 74–83. 23

**Krsek, P.**, **B. Maton**, **P. Jayakar**, **P. Dean**, **B. Korman**, **G. Rey**, **C. Dunoyer**, **E. Pacheco-Jacome**, **G. Morrison**, **J. Ragheb**, *et al.* (2009). Incomplete resection of focal cortical dysplasia is the main predictor of poor postsurgical outcome. *Neurology*, **72**(3), 217–223. 29

**Larsson, G.**, **M. Maire**, and **G. Shakhnarovich** (2016). Fractalnet: Ultra-deep neural networks without residuals. *arXiv preprint arXiv:1605.07648*. 15

**Li, S.**, **C. Zhang**, and **X. He**, Shape-aware semi-supervised 3d semantic segmentation for medical images. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020. 19

**Li, S.**, **Z. Zhao**, **K. Xu**, **Z. Zeng**, and **C. Guan** (2021). Hierarchical consistency regularized mean teacher for semi-supervised 3d left atrium segmentation. *arXiv preprint arXiv:2105.10369*. 20

**Li, W.**, **G. Wang**, **L. Fidon**, **S. Ourselin**, **M. J. Cardoso**, and **T. Vercauteren**, On the compactness, efficiency, and representation of 3d convolutional networks: brain parcellation as a pretext task. *In International conference on information processing in medical imaging*. Springer, 2017. 12

**Liu, C.-F.**, **J. Hsu**, **X. Xu**, **S. Ramachandran**, **V. Wang**, **M. I. Miller**, **A. E. Hillis**, and **A. V. Faria** (2021). Deep learning-based detection and segmentation of diffusion abnormalities in acute ischemic stroke. *Communications Medicine*, **1**(1), 61. 76

**Liu, P.**, Stroke lesion segmentation with 2d novel cnn pipeline and novel loss function. *In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*. Springer, 2019. 77, 91

**Long, J.**, **E. Shelhamer**, and **T. Darrell**, Fully convolutional networks for semantic segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. 12, 15

**Luo, X.**, **J. Chen**, **T. Song**, and **G. Wang**, Semi-supervised medical image segmentation through dual-task consistency. *In Proceedings of the AAAI conference on artificial intelligence*, volume 35. 2021. 61

**Maggioni, M.** and **A. Foi**, Nonlocal transform-domain denoising of volumetric data with groupwise adaptive variance estimation. *In Computational Imaging X*, volume 8296. International Society for Optics and Photonics, 2012. URL `https://www.spiedigitallibrary.org/conference-proceedings-of-spie/8296/82960O/Nonlocal-transform-domain-denoising-of-volumetric-data-with-groupwise-adaptive/10.1117/12.912109.short`. 56

**Maggioni, M.**, **V. Katkovnik**, **K. Egiazarian**, and **A. Foi** (2012). Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE transactions on image processing*, **22**(1), 119–133. URL `https://ieeexplore.ieee.org/document/6253256`. 34, 56

**Maier, O.**, **B. H. Menze**, **J. von der Gablentz**, **L. Häni**, **M. P. Heinrich**, **M. Liebrand**, **S. Winzeck**, **A. Basit**, **P. Bentley**, **L. Chen**, *et al.* (2017). Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. *Medical image analysis*, **35**, 250–269. 74

**Menze, B. H.**, **A. Jakab**, **S. Bauer**, **J. Kalpathy-Cramer**, **K. Farahani**, **J. Kirby**, **Y. Burren**, **N. Porz**, **J. Slotboom**, **R. Wiest**, *et al.* (2014). The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, **34**(10), 1993–2024. xiv, 49, 50

**Milletari, F.**, **N. Navab**, and **S.-A. Ahmadi**, V-net: Fully convolutional neural networks for volumetric medical image segmentation. *In 2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016. 12, 16

**Mlynarski, P.**, **H. Delingette**, **A. Criminisi**, and **N. Ayache** (2019). 3d convolutional neural networks for tumor segmentation using long-range 2d context. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, **73**, 60–72. 24

**Mo, J.-J.**, **J.-G. Zhang**, **W.-L. Li**, **C. Chen**, **N.-J. Zhou**, **W.-H. Hu**, **C. Zhang**, **Y. Wang**, **X. Wang**, **C. Liu**, *et al.* (2019). Clinical value of machine learning in the automated detection of focal cortical dysplasia using quantitative multimodal surface-based features. *Frontiers in neuroscience*, **12**, 1008. 31

**Moeskops, P.**, **M. A. Viergever**, **A. M. Mendrik**, **L. S. De Vries**, **M. J. Benders**, and **I. Išgum** (2016). Automatic segmentation of mr brain images with a convolutional neural network. *IEEE transactions on medical imaging*, **35**(5), 1252–1261. URL `https://ieeexplore.ieee.org/document/7444155`. 52

**Mondal, A.**, **J. Dolz**, and **C. Desrosiers** (2018). Few-shot 3d multi-modal medical image segmentation using generative adversarial learning. *ArXiv*, **abs/1810.12241**. 18

**Nair, V.** and **G. E. Hinton**, Rectified linear units improve restricted boltzmann machines. *In Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010. 38, 65, 84

**Niyas, S.**, **S. C. Vaisali**, **I. Show**, **T. Chandrika**, **S. Vinayagamani**, **C. Kesavadas**, and **J. Rajan** (2021). Segmentation of focal cortical dysplasia lesions from magnetic resonance images using 3d convolutional neural networks. *Biomedical Signal Processing and Control*, **70**, 102951. 53, 70, 81

**Nogueira, R. G.**, **A. P. Jadhav**, **D. C. Haussen**, **A. Bonafe**, **R. F. Budzik**, **P. Bhuva**, **D. R. Yavagal**, **M. Ribo**, **C. Cognard**, **R. A. Hanel**, *et al.* (2018). Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *New England Journal of Medicine*, **378**(1), 11–21. 74

**Noh, H.**, **S. Hong**, and **B. Han**, Learning deconvolution network for semantic segmentation. *In Proceedings of the IEEE international conference on computer vision*. 2015. 64

**Pace, D. F.**, **A. V. Dalca**, **T. Geva**, **A. J. Powell**, **M. H. Moghari**, and **P. Golland**, Interactive whole-heart segmentation in congenital heart disease. *In International*

*Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015. 15

**Peng, S.**, **W. Chen**, **J. Sun**, and **B. Liu** (2020). Multi-scale 3d u-nets: An approach to automatic segmentation of brain tumor. *International Journal of Imaging Systems and Technology*, **30**, 17 – 5. 13

**Rajan, J.**, **K. Kannan**, **C. Kesavadas**, and **B. Thomas** (2009). Focal cortical dysplasia (fcd) lesion analysis with complex diffusion approach. *Computerized Medical Imaging and Graphics*, **33**(7), 553–558. 29

**Rasmus, A.**, **H. Valpola**, **M. Honkala**, **M. Berglund**, and **T. Raiko** (2015). Semi-supervised learning with ladder networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, 3546–3554. 17

**Reddi, S. J.**, **S. Kale**, and **S. Kumar**, On the convergence of adam and beyond. *In International Conference on Learning Representations*. 2018. URL `https://openreview.net/forum?id=ryQu7f-RZ`. 43, 69

**Rickmann, A.-M.**, **A. G. Roy**, **I. Sarasua**, and **C. Wachinger** (2020). Recalibrating 3d convnets with project & excite. *IEEE transactions on medical imaging*, **39**(7), 2461–2471. 25

**Ronneberger, O.**, **P. Fischer**, and **T. Brox**, U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015. 5, 15

**Roth, H. R.**, **L. Lu**, **N. Lay**, **A. P. Harrison**, **A. Farag**, **A. Sohn**, and **R. M. Summers** (2018*a*). Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Medical image analysis*, **45**, 94–107. 25

**Roth, H. R.**, **H. Oda**, **X. Zhou**, **N. Shimizu**, **Y. Yang**, **Y. Hayashi**, **M. Oda**, **M. Fujiwara**, **K. Misawa**, and **K. Mori** (2018*b*). An application of cascaded 3d fully convolutional networks for medical image segmentation. *Computerized Medical Imaging and Graphics*, **66**, 90–99. 25

**Salehi, S. S. M.**, **D. Erdogmus**, and **A. Gholipour**, Tversky loss function for image segmentation using 3d fully convolutional deep networks. *In International workshop on machine learning in medical imaging*. Springer, 2017. 39, 66, 88

**Schlemper, J.**, **O. Oktay**, **M. Schaap**, **M. Heinrich**, **B. Kainz**, **B. Glocker**, and **D. Rueckert** (2019). Attention gated networks: Learning to leverage salient regions in medical images. *Medical image analysis*, **53**, 197–207. 13

**Simpson, S. L.** and **R. A. Prayson** (2014). Post-surgical outcome for epilepsy associated with type i focal cortical dysplasia subtypes. *Modern Pathology*, **27**(11), 1455–1460. 28

**Singh, S. P.**, **L. Wang**, **S. Gupta**, **H. Goli**, **P. Padmanabhan**, and **B. Gulyás** (2020). 3d deep learning on medical images: a review. *Sensors*, **20**(18), 5097. 2

**Smith, S. M.** (2002). Fast robust automated brain extraction. *Human brain mapping*, **17**(3), 143–155. URL `https://onlinelibrary.wiley.com/doi/full/10.1002/hbm.10062`. 34, 56

**Snell, J.**, **K. Swersky**, and **R. S. Zemel** (2017). Prototypical networks for few-shot learning. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4080–4090. 17

**Song, T.** (2019). Generative model-based ischemic stroke lesion segmentation. *arXiv preprint arXiv:1906.02392*. 76, 77, 91

**Srivastava, N.**, **G. Hinton**, **A. Krizhevsky**, **I. Sutskever**, and **R. Salakhutdinov** (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, **15**(1), 1929–1958. 2

**Sudlow, C.** and **C. Warlow** (1997). Comparable studies of the incidence of stroke and its pathological types: results from an international collaboration. *Stroke*, **28**(3), 491–499. 72

**Tan, Y.-L.**, **H. Kim**, **S. Lee**, **T. Tihan**, **L. Ver Hoef**, **S. G. Mueller**, **A. J. Barkovich**, **D. Xu**, and **R. Knowlton** (2018). Quantitative surface analysis of combined mri and pet enhances detection of focal cortical dysplasias. *Neuroimage*, **166**, 10–18. 31

**Thomas, E.**, **S. Pawan**, **S. Kumar**, **A. Horo**, **S. Niyas**, **S. Vinayagamani**, **C. Kesavadas**, and **J. Rajan** (2020). Multi-res-attention unet: A cnn model for the segmentation of focal cortical dysplasia lesions from magnetic resonance images. *IEEE Journal of Biomedical and Health Informatics*. 32, 44, 45, 46, 47, 48, 53, 70

**Tureckova, A.** and **A. J. Rodríguez-Sánchez**, Isles challenge: U-shaped convolution neural network with dilated convolution for 3d stroke lesion segmentation. *In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*. Springer, 2019. 78, 93

**Tursynova, A.** and **B. Omarov**, 3d u-net for brain stroke lesion segmentation on isles 2018 dataset. *In 2021 16th International Conference on Electronics Computer and Computation (ICECCO)*. IEEE, 2021. 78, 93

**Vafaeikia, P.**, **K. Namdar**, and **F. Khalvati** (2020). A brief review of deep multi-task learning and auxiliary task learning. *arXiv preprint arXiv:2007.01126*. 16

**Wang, D.**, **Y. Zhang**, **K. Zhang**, and **L. Wang**, Focalmix: Semi-supervised learning for 3d medical image detection. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020*a*. 19

**Wang, G.**, **T. Song**, **Q. Dong**, **M. Cui**, **N. Huang**, and **S. Zhang** (2020*b*). Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks. *Medical Image Analysis*, **65**, 101787. 77

**Wang, L.**, **C. Xie**, and **N. Zeng** (2019). Rp-net: A 3d convolutional neural network for brain segmentation from magnetic resonance imaging. *IEEE Access*, **7**, 39670–39679. 13

**Wang, Y.**, **Y. Zhou**, **H. Wang**, **J. Cui**, **B. A. Nguchu**, **X. Zhang**, **B. Qiu**, **X. Wang**, and **M. Zhu** (2018). Voxel-based automated detection of focal cortical dysplasia lesions using diffusion tensor imaging and t2-weighted mri data. *Epilepsy & Behavior*, **84**, 127–134. 31

**Woo, S.**, **J. Park**, **J.-Y. Lee**, and **I. S. Kweon**, Cbam: Convolutional block attention module. *In Proceedings of the European conference on computer vision (ECCV)*. 2018. 62

**Wu, K.**, **B. Du**, **M. Luo**, **H. Wen**, **Y. Shen**, and **J. Feng**, Weakly supervised brain lesion segmentation via attentional representation learning. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019. 22

**Wu, Z.**, **R. Ge**, **M. Wen**, **G. Liu**, **Y. Chen**, **P. Zhang**, **X. He**, **J. Hua**, **L. Luo**, and **S. Li** (2021). Elnet: Automatic classification and segmentation for esophageal lesions using convolutional neural network. *Medical Image Analysis*, **67**, 101838. 16

**Xie, S.** and **Z. Tu**, Holistically-nested edge detection. *In Proceedings of the IEEE international conference on computer vision*. 2015. 25

**Yang, C.**, **M. Kaveh**, and **B. J. Erickson**, Automated detection of focal cortical dysplasia lesions on t1-weighted mri using volume-based distributional features. *In 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2011. 29

**Yang, H.**, **C. Shan**, **A. F. Kolen**, and **P. H. de With** (2020*a*). Weakly-supervised learning for catheter segmentation in 3d frustum ultrasound. *arXiv preprint arXiv:2010.09525*. 21

**Yang, H.**, **C. Shan**, **A. F. Kolen**, *et al.*, Deep q-network-driven catheter segmentation in 3d us by hybrid constrained semi-supervised learning and dual-unet. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020*b*. 19

**Yu, L.**, **X. Yang**, **J. Qin**, and **P.-A. Heng**, 3d fractalnet: dense volumetric segmentation for cardiovascular mri volumes. *In Reconstruction, segmentation, and analysis of medical images*. Springer, 2016, 103–110. 14

**Zhang, Y.** and **J. Zhang** (2021). Dual-task mutual learning for semi-supervised medical image segmentation. *arXiv preprint arXiv:2103.04708*. 20

**Zhou, C.**, **C. Ding**, **X. Wang**, **Z. Lu**, and **D. Tao** (2020). One-pass multi-task networks with cross-task guided attention for brain tumor segmentation. *IEEE Transactions on Image Processing*, **29**, 4516–4529. 13

**Zhou, S. K.**, **H. Greenspan**, **C. Davatzikos**, **J. S. Duncan**, **B. van Ginneken**, **A. Madabhushi**, **J. L. Prince**, **D. Rueckert**, and **R. M. Summers** (2021*a*). A review of deep learning in medical imaging: Image traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the Institute of Radio Engineers*, **109**, 820–838. 4

**Zhou, Y.**, **H. Chen**, **Y. Li**, **Q. Liu**, **X. Xu**, **S. Wang**, **P.-T. Yap**, and **D. Shen** (2021*b*). Multi-task learning for segmentation and classification of tumors in 3d automated breast ultrasound images. *Medical Image Analysis*, **70**, 101918. 16

**Zhou, Y.**, **Y. Wang**, **P. Tang**, **S. Bai**, **W. Shen**, **E. Fishman**, and **A. Yuille**, Semi-supervised 3d abdominal multi-organ segmentation via deep multi-planar co-training. *In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019*a*. 19

**Zhou, Z.**, **V. Sodha**, **M. M. R. Siddiquee**, **R. Feng**, **N. Tajbakhsh**, **M. B. Gotway**, and **J. Liang**, Models genesis: Generic autodidactic models for 3d medical image analysis. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019*b*. 26

**Zhu, X.**, **J. Chen**, **X. Zeng**, **J. Liang**, **C. Li**, **S. Liu**, **S. Behpour**, and **M. Xu**, Weakly supervised 3d semantic segmentation using cross-image consensus and inter-voxel affinity relations. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021. 22

**Zhu, X.** and **A. B. Goldberg** (2009). Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, **3**(1), 1–130. 17

# NIYAS S.

Karakkattu Villa, Adinadu South, Karunagappally, Kerala, India - 690542
(+91) 9037208172 ⋄ niyasknpy@gmail.com

## PROFESSIONAL SUMMARY

Highly motivated and skilled researcher, specializing in medical image segmentation using deep learning techniques. Proven track record of publishing high-impact research and collaborating in multidisciplinary teams. As a dedicated researcher with a strong background in developing and implementing deep learning models for medical image analysis, I am seeking a Postdoctoral Research Fellow position to contribute to cutting-edge research in AI-driven medical image analysis.

## EDUCATION

**National Institute of Technology Karnataka - Surathkal.** *Jan 2019 - October 2024*
Doctor of Philosophy (Ph.D)
Department of Computer Science and Engineering

*Thesis Title:* 3D Convolutional Neural Network Architectures for Volumetric Medical Image Segmentation.
*Summary:* The research mainly aims at developing efficient 3D CNN models for various medical image segmentation applications.

**College of Engineering Karunagappally, Kerala (CUSAT)** *July 2012 - May 2014*
Master of Technology (M.Tech)
Department of Electronics and Communication Engineering (Signal Processing)

*Thesis Title:* An Adaptive Approach for Computer Aided Screening of Mammograms and Classification of Abnormalities.
*Summary:* Developed a computer vision based efficient computer aided decision model to automatically detect abnormalities in mammograms.

**College of Engineering Karunagappally, Kerala (CUSAT)** *June 2007 - April 2011*
Bachelor of Technology (B.Tech)
Department of Electronics and Communication Engineering

*Thesis Title:* Image Compression using Secure Arithmetic Coding..
*Summary:* In this project work, developed a secure image compression scheme using arithmetic coding.

## TECHNICAL STRENGTHS

| | |
|---|---|
| **Areas of expertise** | Image Processing, Machine Learning, Deep Learning. |
| **Softwares & Frameworks** | MATLAB, Python, Tensorflow, Keras. |

## WORK EXPERIENCE: 7 YEARS 2 MONTHS

**DetectIQ Private Limited, Kochi, Kerala.** Oct 2023 - July 2014
*Research Scientist* *(9 Months)*

**National Institute of Technology Karnataka** Oct 2020 - Aug 2021
*Senior Research Fellow* *(11 Months)*

**National Institute of Technology Karnataka** Oct 2018 - Sep 2020
*Junior Research Fellow* *(2 Years)*

**Indian Institute of Information Technology and Management, Kerala.** Dec 2015 - Apr 2017
*Research Associate* *(1 Year 5 Months)*

**Indian Institute of Information Technology and Management, Kerala.** July 2014 - Nov 2015
*Project Associate* *(1 Year 6 Months)*

**College of Engineering Karunagappally, Kerala.** Dec 2011 - June 2012
*Assistant Lecturer* *(7 Months)*

## PUBLICATIONS

1. **Niyas, S**., Ramya Bygari, Rachita Naik, Bhavishya Viswanath, Dhananjay Ugwekar, Tojo Mathew, J. Kavya, Jyoti R. Kini, and Jeny Rajan. "Automated Molecular Subtyping of Breast Carcinoma Using Deep Learning Techniques." IEEE Journal of Translational Engineering in Health and Medicine. 11 (2023): 161-169 (Impact factor. 3.4).

2. **Niyas, S**., Shraddha Priya, Reena Oswal, Tojo Mathew, Jyoti R. Kini, and Jeny Rajan. "Automated Molecular Subtyping of Breast Cancer Through Immunohistochemistry Image Analysis." In Computer Vision and Machine Intelligence: Proceedings of CVMI 2022, pp. 23-35. Singapore: Springer Nature Singapore (2023).

3. **Niyas, S**., S. J. Pawan, M. Anand Kumar, and Jeny Rajan. "Medical image segmentation with 3D convolutional neural networks: A survey." Neurocomputing 493 (2022) (Impact factor. 6.0).

4. A. S. Neethi, **Niyas, S**., Santhosh Kumar Kannath, Jimson Mathew, Ajimi Mol Anzar, and Jeny Rajan. "Stroke classification from computed tomography scans using 3d convolutional neural network." Biomedical Signal Processing and Control 76 (2022): 103720. (Impact factor. 5.076)

5. Tojo Mathew, **Niyas, S**, C. I. Johnpaul, Jyoti R. Kini, Jeny Rajan. "A novel deep classifier framework for automated molecular subtyping of breast carcinoma using immunohistochemistry image analysis." Biomedical Signal Processing and Control 76 (2022): 103657. (Impact factor. 5.1).

6. Thomas, Edwin, S. J. Pawan, Shushant Kumar, Anmol Horo, **Niyas, S**. S. Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan. "Multi-res-attention UNet: a CNN model for the segmentation of focal cortical dysplasia lesions from magnetic resonance images." IEEE Journal of Biomedical and Health Informatics 25, no. 5 (2020): 1724-1734.(Impact factor. 7.7)

7. Dev, KM Bijay, Pawan S. Jogi, **Niyas, S**. S. Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan. "Automatic detection and localization of focal cortical dysplasia lesions in MRI using fully convolutional neural network." Biomedical Signal Processing and Control 52 (2019): 218-225. (Impact factor. 5.1).

8. Lestari, Puji, **Niyas, S**, and Dikdik Krisnandi. "Depth Data based Chroma Keying using Grabcut Segmentation." In 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA), pp. 118-123. IEEE, 2018.

9. **Niyas, S**, P. Reshma, and Sabu M. Thampi. "A color image segmentation scheme for extracting foreground from images with unconstrained lighting conditions." In The International Symposium on Intelligent Systems Technologies and Applications, pp. 3-19. Springer, Cham, 2016.

10. Deepa, A. K., **Niyas, S**, and M. Sasikumar. "An adaptive approach for computer aided screening of mammograms and classification of abnormalities." In 2014 International Conference on Communication and Network Technologies, pp. 169-173. IEEE, 2014.

## INVITED TALKS

- A talk on " Techniques in Semantic Segmentation for Medical Image Analysis", in connection with the CUSAT-INID School on Medical Images Understanding, organized by the department of Computer Science, CUSAT, in association with Research Council of Norway (March 2023).
- Hands-on session on "Introduction to Machine Learning and ANN ", As a part of FDP held at St. Joseph's College of Engineering and Technology (SJCET), Palai, Kerala (August 2021)
- A talk on " Satellite Image Processing", in connection with the Technical Symposium 2018 Organized by department of Medical Electronics, BMS College of Engineering, Bangalore (September 2018).
- A talk and Hands-on session on " Image processing and its applications ", in connection with the Workshop on Image Sensing, Medical Imaging and Satellite Image Processing organised by Indian Institute of Information Technology and Management-Kerala (IIITM-K), Trivandrum, India in association with ACM Trivandrum Professional Chapter (March 2015).

## COURSE CERTIFICATIONS

- Exploratory Data Analysis with MATLAB (by Mathworks) - Coursera.
- Neural Networks and Deep Learning (by DeepLearning.AI) - Coursera.

## PORTFOLIO

LinkedIn

Upwork

## REFERENCES

Dr. Jeny Rajan
Associate Professor
Department of CSE
National Institute of Technology Karnataka
Surathkal, India
Email: jenyrajan@nitk.edu.in

Dr. Anoop B N
Assistant Professor
Dept. of Information &
Communication Technology
Manipal Institute of Technology, Manipal
Karnataka, India
Email: anoop.bn@manipal.edu

Dr. Terry Jacob Mathew
Senior Lecturer
Dept. of Computing & Engineering
University of London, Branch Campus, UAE
Email: terry.jacob@uwl.ac.ae

## DECLARATION

I hereby declare that the information furnished above is true to the best of my knowledge.

Niyas S.

Place: Mangalore
Date: 18 October 2024