Deep Learning Assisted MRI Guided Attenuation Correction in PET

A Thesis Submitted for the Degree of Doctor of Philosophy

by

Imene Mecheter

Department of Electronic and Electrical Engineering
College of Engineering, Design and Physical Sciences
Brunel University London

September 2021

Abstract

Positron emission tomography (PET) is a unique imaging modality that provides physiological and functional details of the tissue at the molecular level. However, the acquired PET images have some limitations such as the attenuation. PET attenuation correction is an essential step to obtain the full potential of PET quantification. With the wide use of hybrid PET/MR scanners, magnetic resonance (MR) images are used to address the problem of PET attenuation correction. The MR images segmentation is one simple and robust approach to create pseudo computed tomography (CT) images, which are used to generate attenuation coefficient maps to correct the PET attenuation. Recently, deep learning has been proposed and used as a promising technique to efficiently perform MR and various medical images segmentation.

In this research work, deep learning guided segmentation approaches have been proposed to enhance the bone class segmentation of MR brain images in order to generate accurate pseudo-CT images. The first approach has introduced the combination of handcrafted features with deep learning features to enrich the set of features. Multiresolution analysis techniques, which generate multiscale and multidirectional coefficients of an image such as contourlet and shearlet transforms, are applied and combined with deep convolutional neural network (CNN) features. Different experiments have been conducted to investigate the number of selected coefficients and the insertion location of the handcrafted features.

The second approach aims at reducing the segmentation algorithm's complexity while maintaining the segmentation performance. An attention based convolutional encode-decoder network has been proposed to adaptively recalibrate the deep network features. This attention-based network consists of two different squeeze and excitation blocks that excite the features spatially and channel wise. The two blocks are combined sequentially to decrease the number of network's parameters and reduces the model complexity.

The third approach has been focuses on the application of transfer learning from different MR sequences such as T1 weighted (T1-w) and T2 weighted (T2-w) images. A pretrained model with T1-w MR sequences is fine tuned to perform the segmentation of T2-w images. Multiple fine tuning approaches and experiments have been conducted to study the best fine tuning mechanism that is able to build an efficient segmentation model for both T1-w and T2-w segmentation.

Clinical datasets of fifty patients with different conditions and diagnosis have been used to carry an objective evaluation to measure the segmentation performance of the results obtained by the three proposed methods. The first and second approaches have been validated with other studies in the literature that applied deep network based segmentation technique to perform MR based attenuation correction for PET images. The proposed methods have shown an enhancement in the bone segmentation with an increase of dice similarity coefficient (DSC) from 0.6179 to 0.6567 using an ensemble of CNNs with an improvement percentage of 6.3%. The proposed excitation-based CNN has decreased the model complexity by decreasing the number of trainable parameters by more than 46% where less computing resources are required to train the model. The proposed hybrid transfer learning method has shown its superiority to build a multi-sequences (T1-w and T2-w) segmentation approach compared to other applied transfer learning methods especially with the bone class where the DSC is increased from 0.3841 to 0.5393. Moreover, the hybrid transfer learning approach requires less computing time than transfer learning using open and conservative fine tuning.

Acknowledgments

I would like to express my sincere thanks and gratitude to my supervisors Dr. Maysam Abbod, Prof. Abbes Amira, and Prof. Habib Zaidi for their continuous support, guidance, encouragement, and motivations during this research. I would also like to acknowledge PET Instrumentation and Neuroimaging Laboratory for providing the clinical data, and in particular, many thanks go to Prof. Habib. Secondly, I would like to thank my parents, my husband, my sisters, and my brother for their innumerable support and encouragement. They were my companions of this hard journey who continually pushed me up and helped me putting all pieces together. It is also impossible to forget to thank my friends and collegues: Yousra Regaya, Soumaya Allouch, Zeineb Safi, Abeer Al-marridi, and Aymen Omri for their continuous encouragement and support. Finally, I would like to thank all those who have contributed, in any way, to this research. Above all, it was with the help of Allah, that I was able to complete this work.

Author's Publications

Published Peer-Reviewed Conference Papers

- Mecheter, I., Amira, A., Abbod, M., and Zaidi, H. (2020). *Brain MR Imaging Segmentation Using Convolutional Auto Encoder Network for PET Attenuation Correction.* In Proceedings of SAI Intelligent Systems Conference (pp. 430-440).
- Mecheter, I., Amira, A., Abbod, M., and Zaidi, H. (2020). *Deep Learning based Segmentation for Multi MR Imaging Protocols using Transfer Learning for PET Attenuation Correction.* In 2020 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 2516-2520).

Published Journal Papers

- Mecheter, I., Alic, L., Abbod, M., Amira, A., and Ji, J. (2020). *MR Image-based Attenuation Correction of Brain PET Imaging: Review of Literature on Machine Learning Approaches for Segmentation.* Journal of Digital Imaging, 1-18.

- Mecheter, I., Abbod, M., Zaidi, H., Amira, A. (2022). Brain MR images segmentation using 3D CNN with features recalibration mechanism for segmented CT generation. Neurocomputing, 491, 232-243.

Submitted Journal Papers

- Mecheter, I., Abbod, M., Amira, A., and Zaidi, H. *Deep Learning with Multiresolution Handcrafted Features for Brain MRI Segmentation.* Submitted to Artificial Intelligence in Medicine (Under minor revision).

- Mecheter, I., Abbod, M., Zaidi, H, and Amira, A. *Transfer Learning from T1-weighted to T2-weighted MR Sequences for Brain Image Segmentation*. Submitted to Computer Methods and Programs in Biomedicine Update.

List of Abbreviations

- 18F-FDG: Fludeoxyglucose (18F)
- ASL: Arterial Spin Labeling
- AUC: Area Under Curve (AUC)
- bSSFP: Balanced Steady-State Free Precession
- CNN: Convolutional Neural Network
- CRF: Conditional Random Field
- CTAC: CT based Attenuation Correction
- CT: Computed Tomography
- DCE: Dynamic Contrast Enhanced
- dRHE: Dual Echo Ramped Hybrid Encoding
- DFB: Directional Filter Bank
- DSC: Dice Similarity Coefficient
- DWI: Diffusion Weighted Images
- EEG: Electroencephalogram
- EMG: Electromyographic
- ELM: Extreme Learning Machine
- FLAIR: Fluid-Attenuated Inversion Recovery

- GAN: Generative Adversarial Network
- JSC: Jaccard Similarity Coefficient
- KNN: k-Nearest Neighbors
- LGE: Late Gadolinium Enhancement
- LOR: Line of Response
- MR: Magnetic Resonance
- MLAA: Maximum Likelihood Reconstruction of Activity and Attenuation
- MLEM: Maximum Likelihood Expectation Maximisation
- MP-RAGE: Magnetization-Prepared Rapid Acquisition Gradient Echo
- NSCT: Non Subsampled Contourlet Transform
- NSST: Non Subsampled Shearlet Transform
- OSEM: Ordered Subset Expectation Maximisation
- PET: Positron Emission Tomography
- PDW: Proton Density Weighted
- PMT: Photo Multiplier Tubes
- PolSAR: Polarimetric Synthetic Aperture Radar
- ReLU: Rectified-Linear Unit
- RF: Radiofrequency
- ROC: Receiver Operating Characteristic
- SOM: Self-Organizing Map
- STE: Short Echo Time
- SVM: Support Vector Machine
- SWI: Susceptibility Weighted Imaging

- T1-c: Contrast Enhanced T1-weighted
- T1-gd: Gadolinium Enhanced T1-weighted
- T1-w: T1-weighted
- T2-w: T2-weighted
- TE: Echo Time
- TOF: Time of Flight
- TR: Repetition Time
- TSE: Turbo spin echo
- TXAC: Transmission based Attenuation Correction
- UTE: Ultra-Short Echo Time
- ZTE: Zero Echo Time

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction to PET Attenuation Correction

Positron emission tomography (PET) is an essential medical imaging modality that provides physiological and functional biomarkers using radiolabeled molecules as tracers. 18F-fluorodeoxyglucose (FDG) is the widely used tracer to capture the pathophysiological characteristics of the body. The PET acquisition suffers from photon attenuation which is absorbed by the body. This leads to artifacts and attenuation which should be corrected to obtain the full potential of PET quantification, diseases diagnosis, and treatment planning. The knowledge of the tissue-dependent attenuation coefficient is a critical step to correct the attenuation and achieve an accurate PET reconstruction. Figure 1.1 depicts the impact of the attenuation correction after PET reconstruction.

The attenuation map is typically obtained by acquiring an additional scan of computed tomography (CT). The CT image intensity which is measured by Hounsfield unit reflects the electron density of the tissue. Hence, there is a linear mapping between the CT Hounsfield value and the attenuation coefficient which makes the use of CT as the simplest way for PET attenuation correction. Commonly, the Hounsfield units in CT are mapped into the equivalent linear attenuation coefficients, which is referred to as the attenuation correction map [2, 3]. The

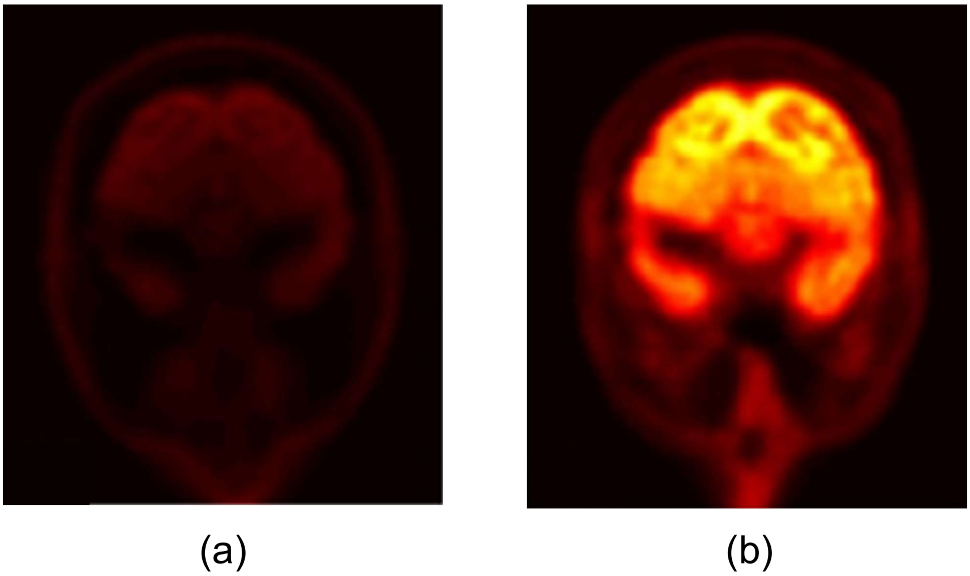


Figure 1.1: A reconstructed PET image without attenuation correction (a), and with attenuation correction (b) using the 18F-FDG radiotracer. Adapted form [1]

main drawback of CT scan is the radiation exposure which makes it an unsafe imaging option for some patients. On the other hand, magnetic resonance (MR) imaging is nowadays considered the premier imaging modality for structural and morphological brain analysis thanks to its excellent soft tissue contrast, high spatial resolution, anatomical and functional information, and absence of ionising radiation. MR images have been extensively used for visualising, analysing, diagnosis, treatment planning, and follow-up of a variety of neurological conditions.

With the various advantages of MR images, hybrid PET/MR systems were introduced and applied in the clinical routine for molecular imaging applications [4]. The main challenge of this hybrid imaging is the signal intensity in MR images has no direct mapping or correlation with the attenuation coefficient. This issue makes the process of generating the attenuation correction map from MR images an open research question [2]. For instance, bone and air tissues have similar intensity values in conventional MR images while they have quite different attenuation coefficients [3]. Figure 1.2 illustrates the difference between CT imaging-based and MR imaging-based attenuation correction.

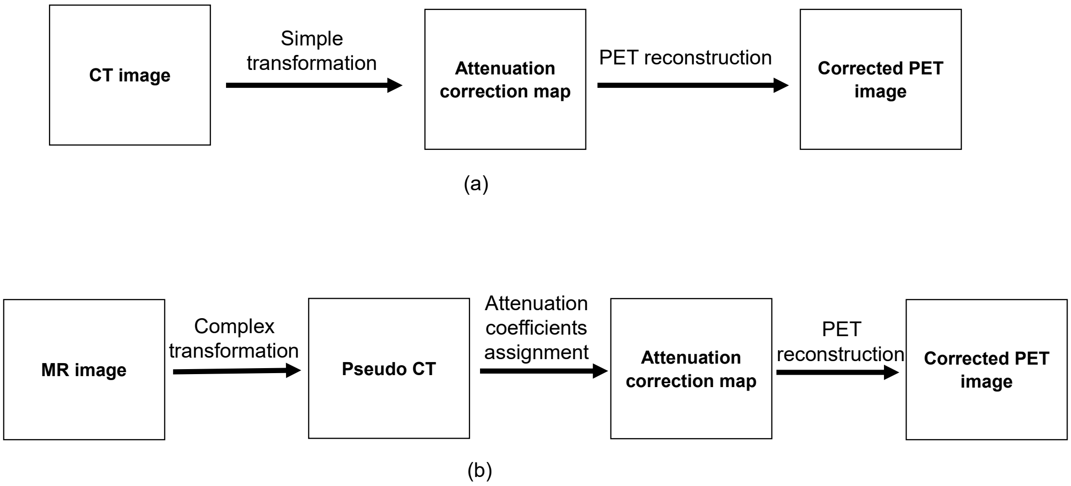


Figure 1.2: The overview process of deriving attenuation correction maps from : (a) CT images, (b) MR images

This current challenge paves the way to investigate and propose various approaches for MR images-based attenuation correction. The commonly used neuroimaging are T1-weighted (T1-w) and T2-weighted (T2-w) MR sequences for MR images-based attenuation correction. In addition to that, there are other sophisticated MR images pulse sequences which provide more details in morphological information: diffusion weighted images (DWI) [5], short echo time (STE) [6], ultra-short echo time (UTE) [7], zero echo time (ZTE) [8], dynamic contrast enhanced (DCE) [9], and magnetisation-prepared rapid acquisition gradient echo (MP-RAGE) sequences [10].

The main methods to generate pseudo-CT images from MR images are segmentation, atlas/template, and emission and transmission methods. The segmentation methods are considered the most robust and mature methods in both research and clinical domain for MR images-based attenuation correction for PET images. The first clinical hybrid PET/MR system uses the two-point Dixon gradient echo sequence which simplifies the segmentation of MR images into different tissue classes [1]. Nowadays, the commercial PET/MR systems segment images into three or four tissue classes [11], with the voxels of each tissue class assigned an approximated predefined linear attenuation coefficient [12] producing the pseudo-CT images which are used to generate the attenuation correction maps. Hence, there is a great need for a reliable segmentation method to generate the attenuation correction coefficients from MR images.

Atlas-based methods, also referred to as the registration-based methods, involve image registration between atlas/template images (MR/CT image pairs) and the target MR image using nonlinear transformation. First, the MR image of a subject is co-registered with the atlas MR image. Then, the obtained transformation combined with atlas CT image are used to create subject specific attenuation correction map by applying a learning algorithm [13]. The quality of PET reconstruction is highly dependent on the registration algorithms accuracy [14–17]. This method is time consuming and potentially have low accuracy under anatomical variations, especially in cases undergoing neurosurgery [18].

Emission and transmission-based methods have been revisited with the introduction of fully integrated clinical time-of-flight (TOF) PET/MR systems where patient specific attenuation map is estimated from TOF emission or transmission data. These methods are categorised into attenuation estimation using consistency conditions, attenuation estimation using only emission data by applying maximum likelihood reconstruction of activity and attenuation (MLAA), attenuation estimation from both emission and transmission data, and attenuation estimation from scattered coincidences [19].

Machine learning techniques are applied to both segmentation and atlas-based methods where different machine learning algorithms are applied either to perform MR images segmentation [20–25] or as a post-registration process to learn the complex mapping from MR images to CT to generate the pseudo CT images. Different machine learning techniques have been applied such as Gaussian mixture [26, 27], k-nearest neighbors (kNN) [28–30], random forest [31, 32], and deep learning [33]

The fact that the atlas-based and emission-based methods produce larger errors in attenuation correction for patients with abnormal anatomy as well as they require intensive computations makes the segmentation the winning approach in terms of robustness, simplicity, and computation time [19]. In order to ensure the accurate diagnosis and quantification, the robustness of the applied algorithm is the crucial prerequisite for the clinical adaptation of a specific MR based attenuation correction method. Since the deployment of segmentation-based on hybrid clinical PET/MRI systems, several studies have shown that the diagnosis accuracy using PET/MR is similar as the usage of the standard reference PET/CT imaging.

Moreover, extensive clinical evaluation and assessment for the atlas and emission-based methods is required to pave the way into clinical PET/MRI systems. The accuracy of the atlas-based methods relies on the database of MR/CT atlases and the performance of the machine learning algorithms. The emission-based method depends on TOF resolution, count level, and activity distribution [19].

The segmentation-based method is the method of choice which is applied on many commercial scanners such as Siemens, Philips, and General Electric. This approach is capable of generating patient-specific attenuation maps. However, it is challenging to accurately segment the brain into tissue classes due to the low image contrast between the air and bone tissues. It is worth investigating different approaches to improve the segmentation performance and exploring the possibility of applying deep learning to perform robust segmentation.

1.2 Introduction to Magnetic Resonance Imaging Segmentation

Current research utilises different machine learning techniques and various MR images data acquisition sequences to perform MR images segmentation for different medical applications. DWI sequence or a combination of sequences is the most routinely used for the diagnosis and follow-up in ischemic and hemorrhagic stroke [34–36]. For quantitative analysis in multiple sclerosis, T2-w MR images is the commonly used sequence either as a single imaging sequence [37] or in a multi-sequence approach [38–40]. T1-w MR images is frequently used to assess biomarkers of Alzheimer's disease such as hippocampal atrophy, ventricle enlargement and cortex shrinkage [41, 42]. Brain tumour segmentation of MR images received much attention over the last decade, especially for treatment planning and follow-up. A range of MR images sequences were used to perform segmentation such as single MR images sequence with [43] and without [44] contrast agent, or multi-sequence MR images with [45–48] or without contrast [45, 49].

The segmentation process is achieved by dividing the brain MR images into three or more tissue classes such as white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), air, soft tissue, bone or lesion regions. In addition to more classical segmentation methods as thresholding, edge based, and watershed based, various machine learning techniques have been applied for brain segmentation. For instance, random forest classifier has been used for white matter lesions [50], hippocampus [51], and brain tumor lesions [52] segmentation. Support vector machine (SVM) has shown its efficiency in Alzheimer disease diagnosis [41], white matter lesion [53], multiple sclerosis region [54], and brain tumor lesions [45] segmentation. Additionally, probabilistic models have been proposed to efficiently segment the brain tissue into three classes or only extract the regions of interest by applying the inverted Dirichlet mixture Model [55], Markov random field [56] and Bayesian model [57]. Furthermore, neural network such as multilayer perceptron [58], self-organizing map (SOM) network [59], extreme learning machine (ELM) [60], and cellular neural network [61] have achieved promising results in segmenting the brain into different regions. Unsupervised learning, specifically clustering techniques, have been widely used for skull striping [62], white matter lesion segmentation [63], and tumor region segmentation [44].

Recently, deep learning has proven its superiority in several brain segmentation applications using convolutional neural networks (CNNs) for skull stripping [64], stroke lesion segmentation [65, 66], multiple sclerosis segmentation [67], brain tumor segmentation [48, 68–70], prostate lesion segmentation [71], infant brain segmentation [72], and hippocampus segmentation [73]. The CNN outperforms other machine learning algorithms to perform segmentation of various medical imaging modalities and applications. The crucial prerequisites to apply deep learning methods are the availability of large datasets and powerful hardware computing resources.

1.3 Research Methodology

The process of MR based attenuation correction is addressed in this research work by the segmentation approach using the deep CNN. The general process of segmentation based attenuation correction for PET images composes of multi-components which are depicted in Figure 1.3. In this thesis, the aim and the objectives are dedicated only to the segmentation process which is highlighted in dashed lines in Figure 1.3.

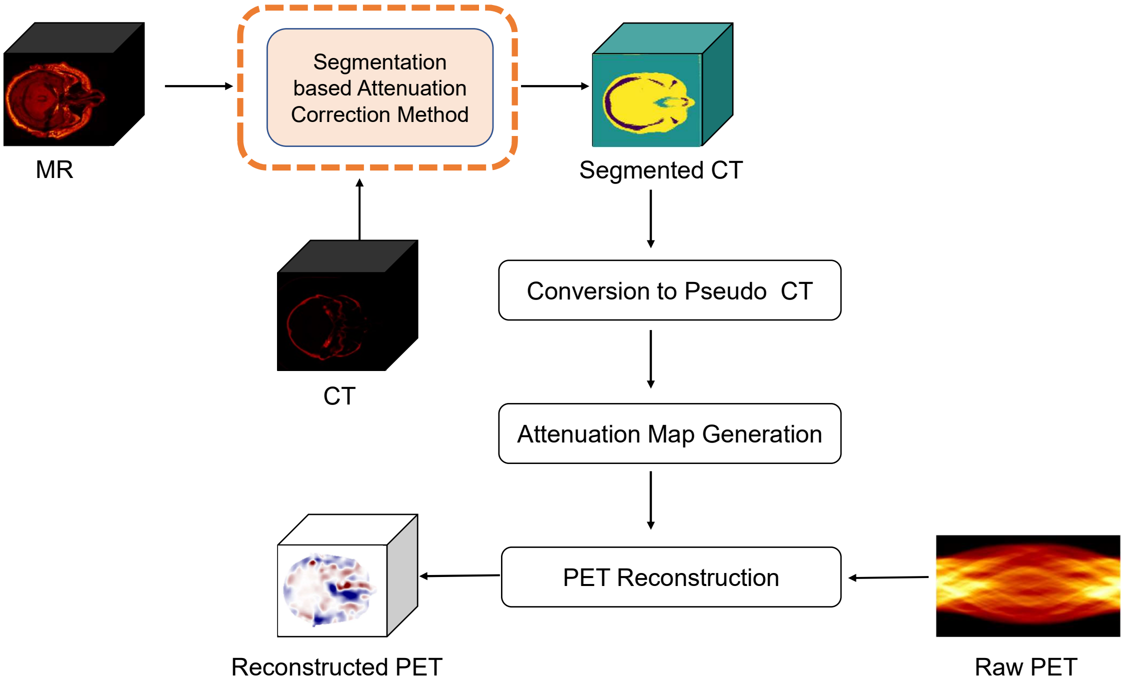


Figure 1.3: The overview process of MR segmentation based attenuation correction for PET images

In this thesis, the research work is focused on the improvement of the segmentation technique as a transformation approach from MR images to pseudo-CT images which will be used to generate the attenuation maps for PET attenuation correction. The segmentation-based MR attenuation correction is performed using deep CNN as it has shown its superiority in different medical imaging applications.

The proposed approach aims at addressing the brain segmentation in different aspects such as increasing the segmentation performance, decreasing the model complexity, decreasing the computation time and the computing resources, building a multi-modality segmentation model, and transferring the knowledge from different MR sequences.

The first stage of this research work intends to increase the segmentation performance by combining the multiresolution analysis techniques with CNN. The second stage aims at reducing the model's complexity by enhancing the CNN architecture using squeeze and excitation components and building an ensemble of CNNs that produces reliable decisions. The third stage's objective is to build a multi-modality model that is able to segment different brain MR sequences and transfer the knowledge from one sequence to another. At each stage, the proposed algorithm is evaluated using the clinical brain datasets by visual assessment and quantitative evaluation. Each segmented MR volume is compared with the ground truth CT volume by calculating the confusion matrix, precision, recall, dice similarity coefficient (DSC), and Jaccard similarity coefficient (JSC).

1.4 Research Aim and Objectives

The aim of this research is to propose a robust deep learning algorithm to enhance the bone segmentation in brain MR images for PET attenuation correction and quantification. This aim is achieved by endeavouring the following objectives:

- Explore and investigate the medical imaging modalities: MR and CT images.
- Design a 3D CNN architecture and thoroughly study and experiment the impact of different components of the network.
- Investigate the impact of combining multiresolution analysis techniques such as Contourlet and Shearlet with the CNN's features to improve the bone segmentation.
- Propose a novel segmentation CNN architecture that uses different squeezing and excitation blocks to boost the most dominant features and decrease the model complexity by relying only on the CNN's features and eliminating the handcrafted features (e.g., Contourlet and Shearlet features).

- Propose an ensemble of different CNN's architectures to build a reliable segmentation model.
- Investigate the applicability of transfer learning from different MR sequences by transferring the learning from T1-w MR to T2-w MR sequences to perform the segmentation.
- Evaluate and assess the performance of the proposed methods using clinical datasets of fifty patients with different demographics and clinical diagnosis. Each objective is evaluated by visual assessment and measurement of different segmentation metrics using the ground truth images.
- Validate the proposed models by comparing the results with other studies in the literature that applied segmentation using deep CNN to address the MR based attenuation correction for PET images.

1.5 Research Contributions to Knowledge

The main contributions of this research are summarized as follows:

- The combination of the multiresolution analysis techniques (Contourlet and Shearlet transforms) with 3D deep CNN's features to segment the brain into three tissue classes. The Contourlet and Shearlet transforms have been applied widely as handcrafted features to perform segmentation and classification tasks. However, they have not been combined with CNN's features. The design of the combination of handcrafted features with deep CNN's features has been investigated and evaluated to find the optimal approach which improves the segmentation performance.
- The proposal of a novel 3D excitation based convolutional encoder decoder network to segment the brain MR images. The excitation mechanism consists of squeezing then exciting blocks that are able to recalibrate the network features and distinguish between the most useful features on spatial and channel wise. Those blocks have been integrated

in a parallel design in other studies. However, to obtain the full potential of the squeeze and excitation blocks and to reduce the model complexity, a sequential combination of the channel wise and the spatial wise excitation blocks is proposed to perform the brain MR segmentation.

- The development of an ensemble of deep CNNs that combines three convolutional encoder decoder networks which are trained with different loss and activation functions. The application of the ensemble approach for PET attenuation correction is used for the first time in this research.
- The application of transfer learning from T1-w MR sequence to T2-w MR sequence for PET attenuation correction by proposing a hybrid transfer learning technique. The hybrid transfer learning using two conventional MR sequences is proposed in this research work for the first time. In addition to that, the application of domain adaptation by building a multi-modality convolutional encoder decoder network which is able to segment two different MR sequences (T1-w and T2-w). This type of augmentation has not been proposed in the literature.

1.6 Thesis Organisation

The remaining chapters of this thesis are organized as follows:

- A review about the applied segmentation methods for MR image based PET attenuation correction, the different CNN's architectures that are applied for segmentation tasks, the transfer learning approaches, and the medical imaging modalities is represented in Chapter 2.
- The design of 3D CNN and the combination of multiresolution analysis technique with CNN based features are detailed in Chapter 4 with extensive analysis and evaluation of the segmentation results.

- The proposal of an excitation based 3D CNN with the integration of both spatial and channel "Squeeze and Excitation" blocks sequentially as well as the ensemble of CNNs are depicted in Chapter 5 along with the segmentation results and their evaluations.
- The multi-modality model and the transfer learning from T1-w MR sequence to T2-w MR sequence are proposed in Chapter 6 with thorough evaluation.
- The general discussion about the proposed approaches with the revised research objectives are discussed in Chapter 7. In addition to that, the limitations and general recommendations of the proposed methods are discussed.

An overview of the thesis's organisation is illustrated in Figure 1.4 where each chapter is associated with its published or submitted papers.

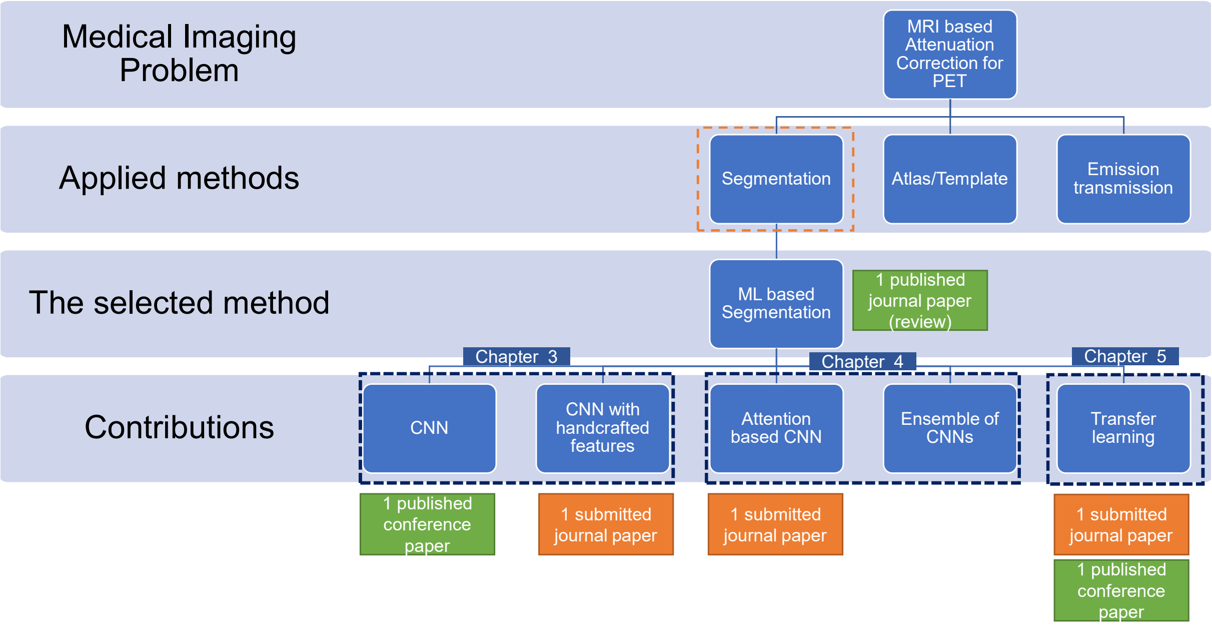


Figure 1.4: The overview organisation of the thesis where each chapter is associated with its published or submitted papers.

Chapter 2

Related Literature

2.1 Introduction

This chapter provides a brief overview of the three medical modalities: PET, MR, and CT images along with hybrid imaging systems including the image characteristics, acquisition details, reconstruction process, and clinical applications. Moreover, PET attenuation correction is introduced in this chapter by highlighting how this problem has been addressed in the literature using MR images. Additionally, MR images based attenuation correction methods are reviewed with the focus on segmentation approach. Deep CNN is one of the most efficient tools to perform image segmentation. Therefore, multiple CNN related mechanisms which can be deployed to perform and enhance medical images segmentation are also reviewed in this chapter.

Section 2.2 presents an overview of different medical imaging modalities and hybrid imaging systems. Section 2.3 introduces the MR images segmentation based attenuation correction methods. Section 2.4 covers deep CNN with the focus on attention based CNN architecture. Section 2.5 reviews the combination of multiresolution handcrafted features with deep CNN's features. Section 2.6 explores the application of the ensemble of deep CNNs to enhance the performance of various medical imaging. Section 2.7 covers the transfer learning and domain adaptation from different medical imaging modalities and multi sequences. Finally, Section 2.8

summarises the chapter findings and identifies the research gaps.

2.2 Medical Imaging Modalities: Overview

There are various medical imaging modalities which are commonly used in the clinical routine such as PET, MR, and CT images. Each modality has its specific characteristics and acquisition technique to visualise the human tissue and organs to help in the diagnosis process and treatment planning.

2.2.1 Positron Emission Tomography

PET imaging is a unique *in vivo* medical modality for measuring the physiological and functional behavior of the body at the cellular level. PET is based on radiopharmaceutical labeled biological molecules as tracers to map their distribution within the human body. The most widely used tracer is FDG which is a glucose analogue that helps to measure the consumption of the glucose *in vivo*. The sensitivity of PET images is very high compared to other conventional nuclear imaging which is based on single-photon [74].

PET depends on annihilation coincidence detection mechanism to determine the directionality of the detected photons. The PET acquisition process starts with injecting the patient with a positron emitting radiopharmaceutical such as 18F-FDG tracer. When the nucleus of the radioactive isotope emits a positron, one of the 18F atoms will decay. The positron travels randomly for few millimetres in tissue before combining with an electron. Both the positron and electron will undergo an annihilation where their masses are converted to 511 keV energy in the shape of two Gamma photons. These two photons travel in opposite directions and eventually will be detected simultaneously by two small detectors. The line of response (LOR) is the link that connects two detectors. One reason behind the high sensitivity of PET system is the use of scintillation crystals placed around the patient in the shape of a ring to ensure the detection of both Gamma photons. Figure 2.1 illustrates the ring of scintillation crystals which surrounds the patient with multiple lines of response.

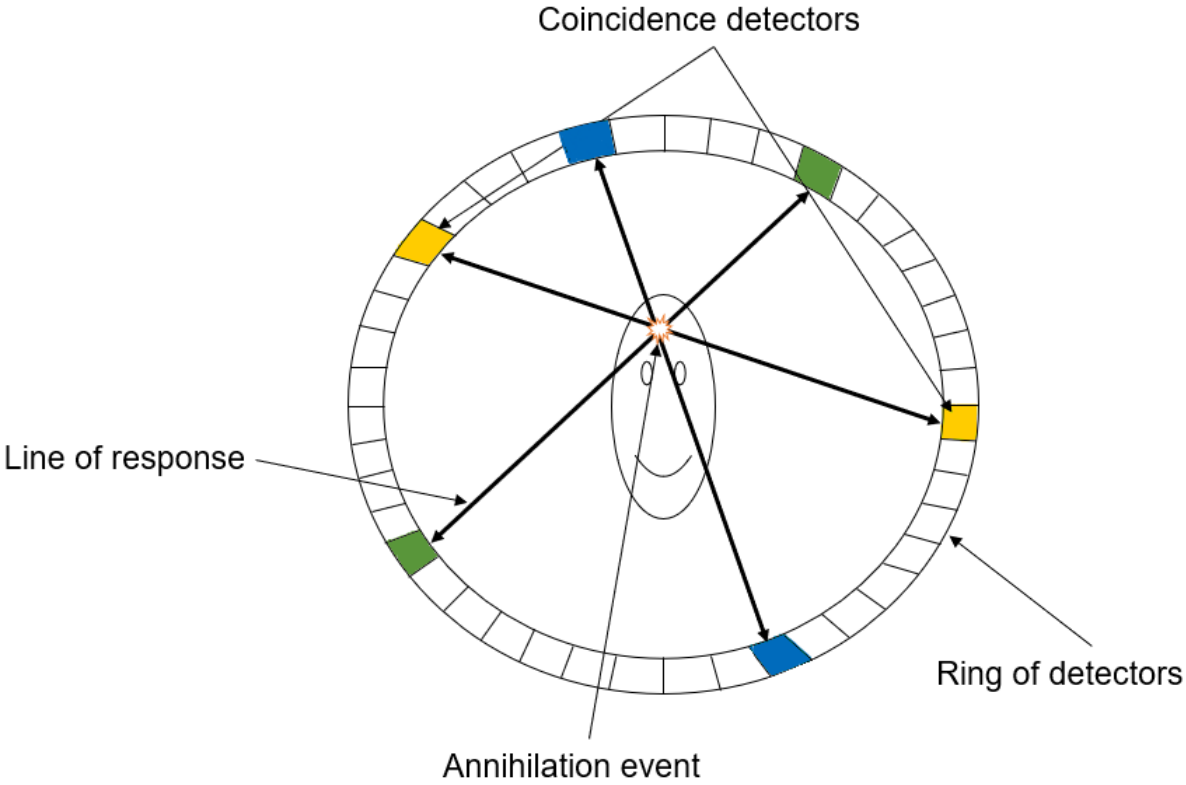


Figure 2.1: Detection scheme of PET imaging.

The scintillation crystals absorb the emitted photons and produce light from the photons energy. This light is converted into an electrical signal using Photo Multiplier Tubes (PMT). Both the spatial distribution and the magnitude of the electrical signals represent the raw PET data. A summary of the PET acquisition process is illustrated in Figure 2.2 and an illustration of the stages of the detection event and the signal generation is presented in Figure 2.3. The PET acquisition process is painless and takes around 10- 40 minutes to be completed [74, 75].

The PET reconstruction process is the conversion of the 1D projection (LOR) to 2D plane image. The raw PET images are reconstructed to take two possible formats: sinogram and list mode. The sinogram format represents the coincidence events which are grouped into projection images called sinograms. The value of each sinogram pixel represents the number of coincidence events related to that LOR. Each 2D slice is represented by a single sinogram which can be reconstructed [78]. The list mode is an event-by-event format which stores the

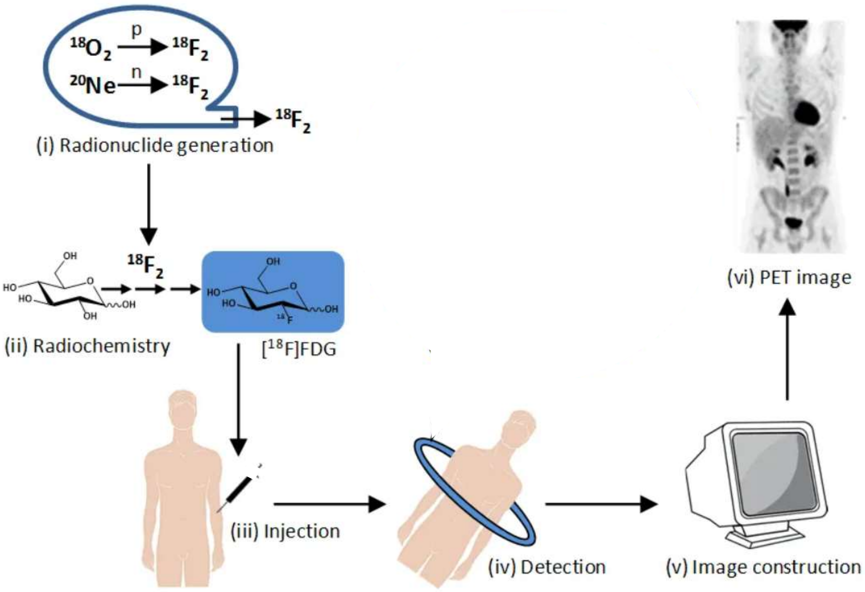


Figure 2.2: An overview of the PET imaging process. Adapted from [76].

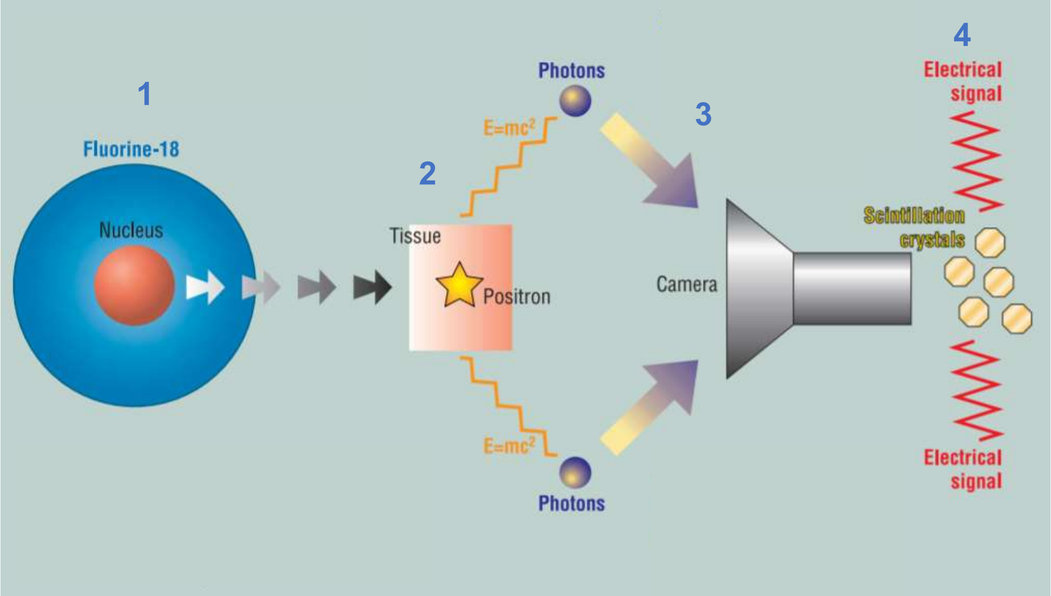
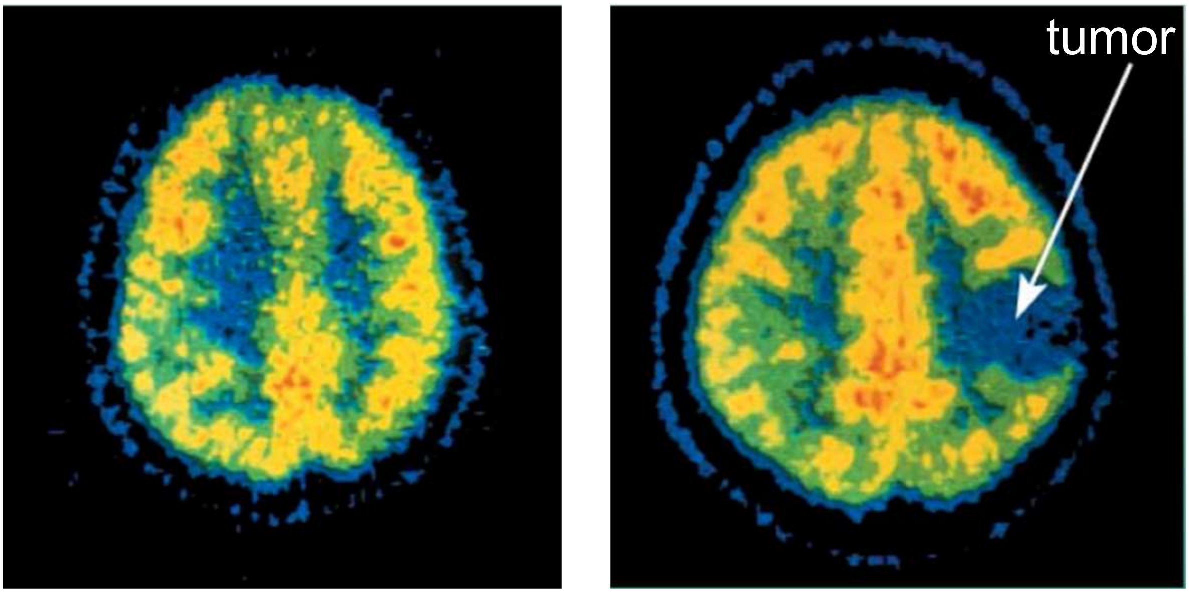


Figure 2.3: An illustration of the four stages of the PET detection process. (1) The proton emission event from the nucleus of the radioactive isotope, (2) The positron and electron's annihilation process and the conversion of their masses to 511 keV energy in the shape of two Gamma photons, (3) The PET camera detects the photons using the detectors, (4) The conversion of the photons into an electrical signal. Adapted from [77].

coordinates of the coincidence events in a sequential data stream along with other additional information such as the time and the energy of the detection event. The list mode format does not preserve the spatial localization information of each event. Yet, it is an efficient storage format with high temporal resolution and flexibility for manipulation [79].

The PET images are reconstructed on 2D and 3D approaches. The coincidences that happen within one ring can be reconstructed to 2D images. In the case of allowing the coincidences within multiple rings, 3D images can be reconstructed. The most common reconstruction algorithms of PET are the filtered back-projection [75] and iterative reconstruction algorithms such as maximum-likelihood expectation maximisation (ML-EM) and its accelerated version ordered subset expectation maximisation (OSEM) [75]. The PET images should be corrected prior to the reconstruction process to obtain the full potential of PET quantification. The PET correction includes the detector efficiency variations, attenuation correction, and scattered and random coincidences correction.

The clinical applications of PET imaging are in the domain of neurology, cardiology, and oncology. In neurology, PET imaging is commonly acquired to measure the cerebral blood flow and the brain's energy consumption which can help in the diagnosis of dementia and strokes. In addition to that, PET imaging is used to measure the glucose and tumor metabolism, the oxygen extraction, and other chemical neurotransmitters such as dopamine which helps to diagnose Parkinson's disease. In oncology, PET is used for detecting tumors and metastases and measuring the effect of treatment on tumor metabolism. Figure 2.4 shows PET images for one healthy and one cancerous brain where the poor blood flow indicates the metabolic abnormality. In cardiology, the radiopharmaceutical 13N-ammonia tracer is used to measure the myocardial blood flow while the 18FDG tracer is used for energy consumption measurement to identify the ischemic tissue. Moreover, PET imaging is used for the pre-transplantation assessment of viable myocardium [80]. The PET images of cardiac blood flow and glucose consumption are illustrated in Figure 2.5.



(a) (b)

Figure 2.4: PET images of (a) one healthy brain and (b) one brain with malignant lesion (tumor) where the poor blood flow indicates the metabolic abnormality. Adapted from [77]

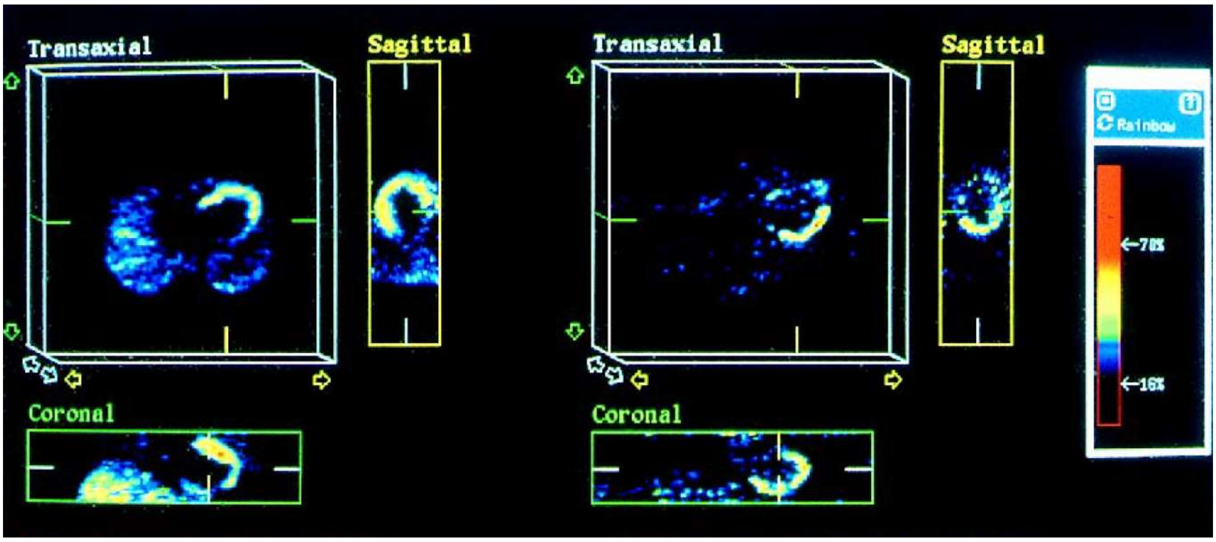


Figure 2.5: PET polar maps of cardiac blood flow (left) and glucose consumption (right) in three views (transaxial, coronal, and sagittal).Adapted from [80].

2.2.2 Magnetic Resonance Imaging

MR imaging is considered the premier tool in diagnostic radiology. It is able to capture the fine structural and morphological details of anatomy thanks to its excellent soft tissue contrast and high spatial resolution. Additionally, it is considered the safest non-invasive imaging technique because the patient is not exposed to any ionising radiation such as CT scan or X-rays [81]. MR is a versatile imaging technique where the soft tissue details can be visualised in all three planes: transaxial, sagittal and coronal as illustrated in Figure 2.6.

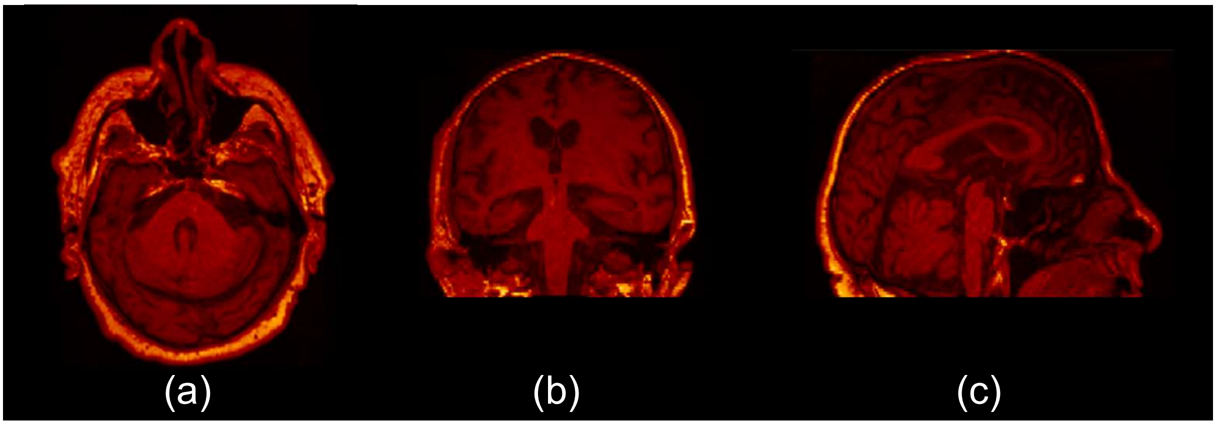


Figure 2.6: An illustration of a single slice of MR imaging in (a) axial, (b) coronal, and (c) sagittal views.

The MR imaging concept is invented with the fact that the hydrogen atoms in the human tissue act as little magnets. The hydrogen atom consists of one proton which rotates along its axis and creates the resonance effect. When the human body enters the MR scanner (hollow cylindrical magnet), it gets exposed to a strong magnetic field. With this magnetisation, the hydrogen's proton will be aligned with or against the magnetic field. An oscillating radiofrequency (RF) current is imposed which allows the protons to change their alignments as well as the direction of their magnetic field. Once the RF is off, the protons go to their original alignments with the magnetic field. This realignment of the hydrogen's protons creates an electrical signal that can be captured by a computer and transformed to an MR image. One remarkable feature of MR imaging, it is able to capture the human body with a spatial resolution of better than 1 mm [81]. Figure 2.7 shows the component of an MR scanner.

The produced MR image is described by the physical parameters of the tissue which are proton density, T1 relaxation time, and T2 relaxation time. It also depends on the RF pulse parameters such as repetition time (TR), echo time (TE) and others [83]. The relaxation time

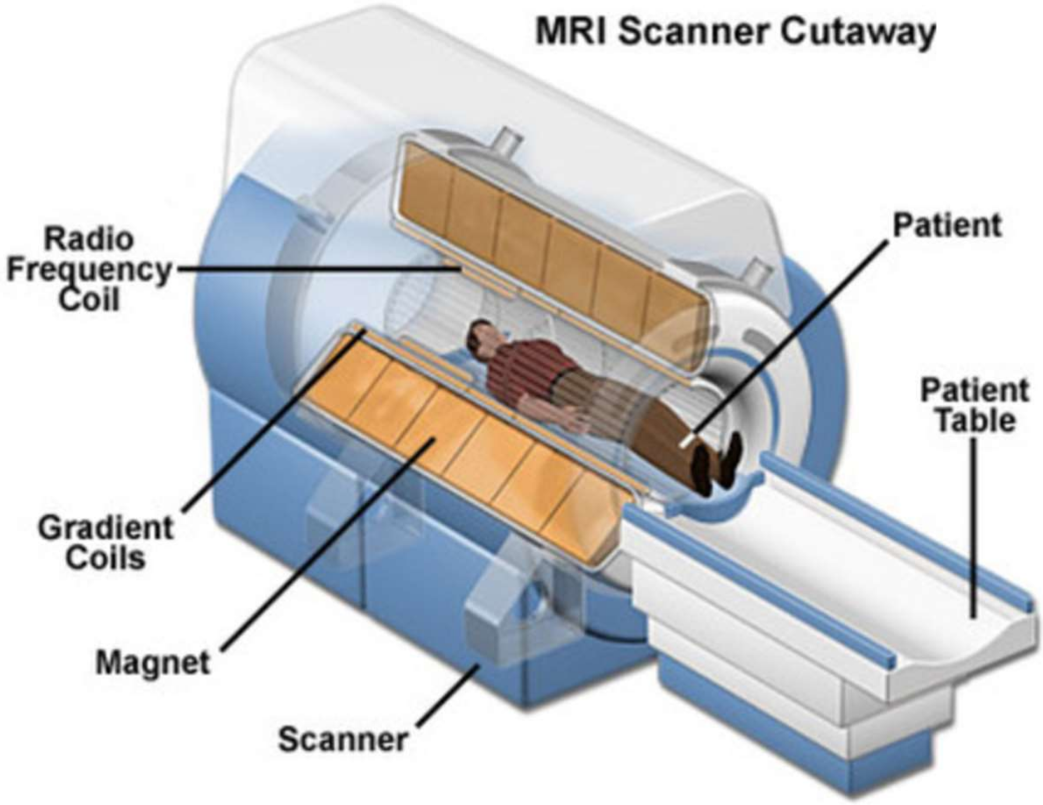


Figure 2.7: The basic design of an MR scanner which shows the components of the MR scanning system. Adapted from [82].

refers to the time required by the proton to realign and return to its equilibrium state either longitudinal (T1) or transversal (T2). The T1 relaxation time is much longer than T2 relaxation time.

Various types of MR images with different appearances are generated by varying the sequence of the applied RF pulses. The MR sequences are categorised into T1-w, T2-w, DWI, proton density weighted (PDW), fluid-attenuated inversion recovery (FLAIR), flow sensitive and others [84]. The most common sequences are T1-w and T2-w which clearly differ in CSF tissue. On T1-w, the CSF tissue is dark while it is bright on T2-w sequence. Moreover, the T1-w images are acquired using short TR and TE times. In contrary, T2-w images are scanned with longer TR and TE times. The brightness and contrast of T1-w and T2-w images are determined by the tissue's T1 properties and T2 properties, respectively [85]. Figure 2.8 shows three MR images with different contrast patterns named T1-w, DWI, and T2-w images.

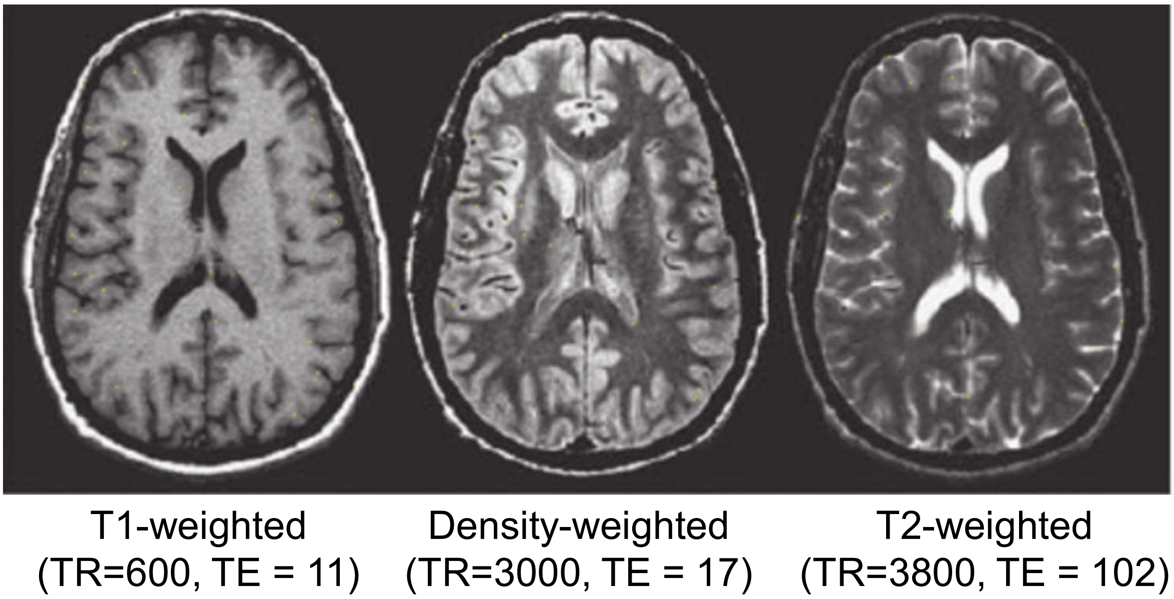


Figure 2.8: Three conventional MR sequences namely T1-w, DWI, and T2-w images. Adapted from [86].

Table 2.1 represents TR and TE values of T1-w and T2-w images and Table 2.2 illustrates the differences between the brightness and contrast of different T1-w and T2-w brain's tissues.

Table 2.1: The approximate TR and TE times values of T1-w and T2-w sequences [84].

Sequence	TR (msec)	TE (msec)
T1-w (short TR and TE)	500	14
T2-w (long TR and TE)	4000	90

Table 2.2: The brightness and contrast of different T1-w and T2-w brain tissues [84].

Tissue	T1-w	T2-w
CSF	Dark	Bright
White matter	Light	Dark gray
Cortex	Gray	Light gray
Fat (within bone marrow)	Bright	Light
Inflammation (infection, demyelination)	Dark	Bright

MR imaging is widely used in different clinical diagnosis. For instance, it is used to diagnose vascular cases including ischemic stroke, aneurysm, and venous thrombosis. It is also extensively used to detect tumor in the brain or the whole body. Additionally, it provides fine details to diagnose different infections such as cerebritis, encephalitis, and meningitis. Moreover, it is used for other applications such as detecting multiple sclerosis lesions and epilepsy [85].

2.2.3 Computed Tomography

CT is a common medical imaging based on computerised x-ray scan. This modality offers the structural view of the human body. CT images do not provide soft tissue details, yet it is an important imaging technology that helps to make urgent diagnosis for multiple diseases or treatment planning. The main drawback of CT scan is the radiation exposure which tends to limit the number of CT scanning per year for each patient. For some cases, the patient should be injected with an iodinated contrast agent before the scanning [86]. Figure 2.9 shows the first clinical CT image which is acquired at Atkinson Morley's Hospital in London (1971) [87] and another CT image which is acquired recently by Siemens Biograph mCT scanner.

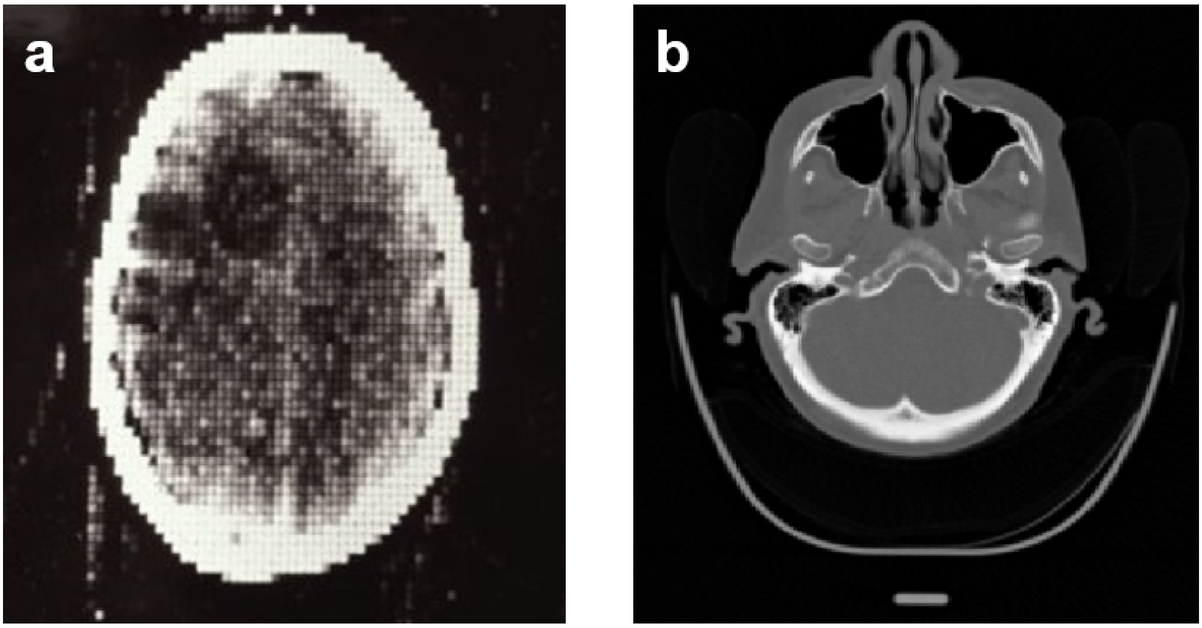


Figure 2.9: (a) The first clinical CT image which is acquired at Atkinson Morley's Hospital in London (1971) adapted from [87]
, (b) 2D CT image acquired by Siemens Biograph mCT scanner.

The acquisition of CT images depends on the X-ray transmission-based technique in which a motorised X-ray source beams that pass through the patient and received by detectors situated in the opposite side of the X-ray source. The X-ray source and the detectors are rotating around the patient's body by a particular angle to measure the tissue density with Hounsfield unit. The attenuation of the X-rays after passing through the patient are received by the detectors and transmitted to a computer [86]. Figure 2.10 illustrates an abstract view of the CT scanning system.

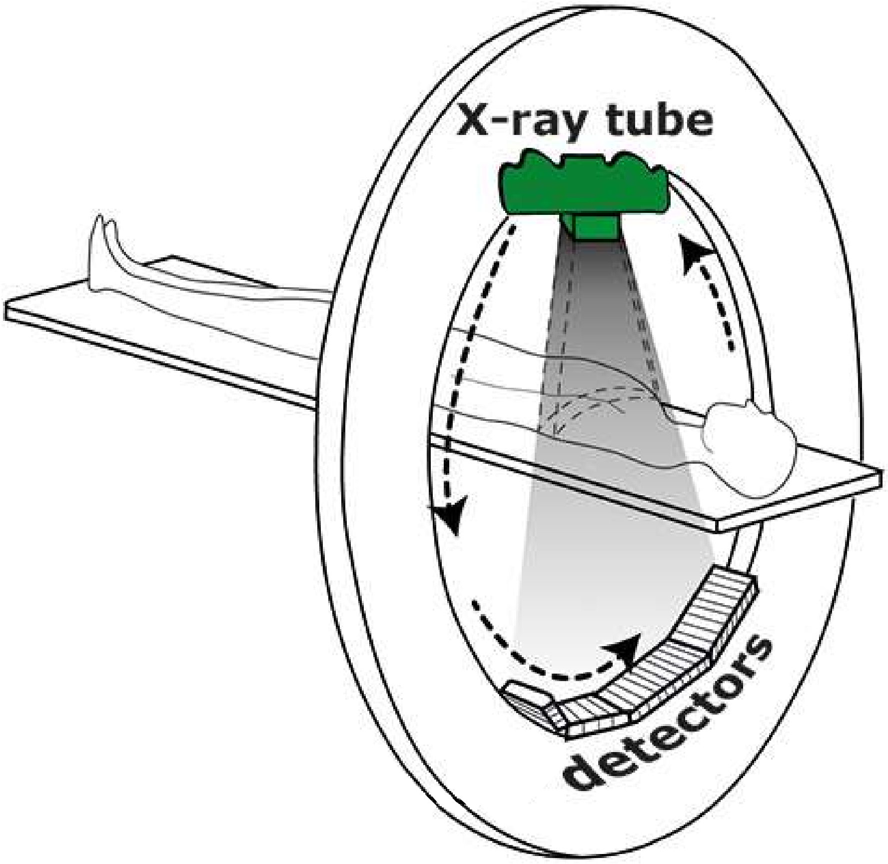


Figure 2.10: The principle components of CT scanning. A rotating X-ray source and a detector unit around the patient. Adapted from [88].

The contrast between the different body tissues is generated from the different attenuation values of X-rays in the patient's body. The 2D spatial distribution of tissue attenuation coefficients represents the signal intensities [86]. Table 5.4 shows the Hounsfield values for different tissues where the denser structures are represented by larger Hounsfield values.

The design of CT scanners is constantly improving where a CT scanner with two X-rays sources is developed to offer further information and decrease the time of the acquisition process.

Table 2.3: Hounsfield values for different tissue classes of the human body.

Tissue	Hounsfield
Air	-1000
Lung	-600 to -400
Fat	-100 to -60
Water	0
Muscle	10 to 40
Blood	30 to 45
Soft tissue	40 to 80
Bone	400 to 3000

Another development is the increase of the number of CT slices acquired simultaneously to cover a wider view. Volumetric CT can be acquired with the invention of helical CT scanner that enables the continuous rotation of the gantry and the patient's bed [87].

The X-rays projections are reconstructed to form 2D slices which can be stacked on top of each other to build a 3D CT image. The commonly used CT reconstruction algorithms are filtered back-projection technique and algebraic reconstruction algorithms [87]. The algebraic reconstruction algorithms are much simpler than filtered back-projection, yet they are less accurate. The availability of numerous projections with uniform distribution are two main requirements for the filtered back-projection technique. In case these two requirements are missing, the algebraic reconstruction algorithms can be applied alternatively. The algebraic reconstruction algorithms include simultaneous algebraic and iterative reconstruction techniques [86].

CT imaging is used for various clinical conditions including neurological, pulmonary, and abdominal diseases. For instance, CT images help to identify the possible brain fractures, hemorrhage, and edema. The hemorrhage area is indicated with high signal intensity while the edema is indicated with low signal intensity on CT images. CT is also outstanding at showing the brain tumor and can be used to analyse the bone structure changes. The detection of pulmonary diseases is very challenging using MR imaging. Yet, CT imaging is useful to detect abnormalities in the lung images such as pulmonary malignancies, emboli, silicosis, fibrosis, and emphysema. Additionally, 3D CT imaging with the use of contrast agent is used to detect abdominal tumours and ulcerations in the liver [86].

2.2.4 Modalities Comparison

Table 2.4 summarises the main differences between the three modalities: PET, CT, and MR images in terms of imaging technology, scope of application, acquisition time, safety level, and cost [89, 90].

Table 2.4: Comparison between the three modalities: PET, CT, and MR images.

	PET	CT	MR
Imaging technology	Radioactive tracers at the cellular level	X-ray	Magnets and radiowave
Scope of application	Biological process at the molecular level	Bone detection accurately	Large variety of medical conditions
Acquisition time	Very long (hours)	Short (few minutes)	Long (around 30 minutes)
Safety	Moderate to high radiation	High radiation	Free radiation
Cost	Much higher than regular CT	Less than MR	More than CT

2.2.5 Hybrid Scanners

The hybrid imaging is a powerful tool that integrates functional and structural information from different modalities and provides the full potential of quantitative analysis. PET imaging has a unique capability to show the functional and physiological details of the tissue. However, it has low spatial resolution and non-detailed anatomical information which significantly help to detect multiple diseases. These limitations have paved the way to find another imaging partner for PET images to complement each other's such as CT or MR images [91].

A PET/CT Scanner

The hybrid PET/CT scanner is the first hybrid imaging system that was implemented and evaluated in the clinical routine. It has been proved to be a successful hybrid imaging that provides anatomical information of body's tissue by CT images with their corresponding metabolic functions by PET images. The acquisition process of both imaging can be sequential or simultaneous.

Figure 2.11 illustrates brain CT, PET, and fused PET/CT images.

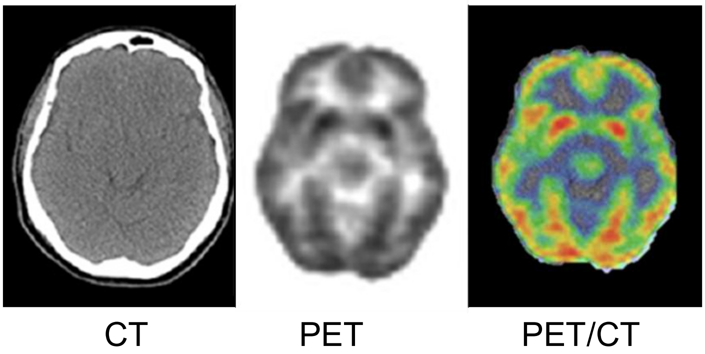


Figure 2.11: Clinical brain CT, PET, and fused PET/CT images which are adapted from [92].

This hybrid imaging has made an important revolution in the diagnosis of various diseases in oncology, neurology, treatment planning, and cancer staging. The availability of PET/CT imaging has gradually eliminated the conventional PET scanning. Figure 2.12 shows a hybrid GE Discovery 690 PET/CT scanner. The main drawback of this scanner is the high cost [92].

Figure 2.12: GE Discovery 690 PET/CT scanner [93].

B PET/MR Scanner

Although the PET/CT is a successful integrated imaging system, it still lacks the excellent soft-tissue contrast which can be provided by MR imaging. Meanwhile, MR imaging is not able to capture the tissue's physiology at the molecular level. The combination of PET acquisition process with MR imaging allows to integrate the structural information with high spatial resolution of MR soft tissue and PET functional details. Both imaging can be acquired sequentially or simultaneously [92]. Figure 2.13 represents the sequential PET/MR scanner of the Philips whole-body Ingenuity system at Geneva University Hospital with MR, PET, and PET/MR images. Although the role of PET/MR imaging in clinical setting is still controversial and there are ongoing efforts to address many related challenges properly, brain and whole-body PET/MR imaging has various clinical applications. Hybrid PET/MR images are able to capture the fine details of patient-specific tumor as well as they provide information for personalised therapy. This hybrid imaging is also capable of detecting neurodegenerative disorders, dementia, epileptic foci , focal brain ischemia, and inflammatory disorders [91].

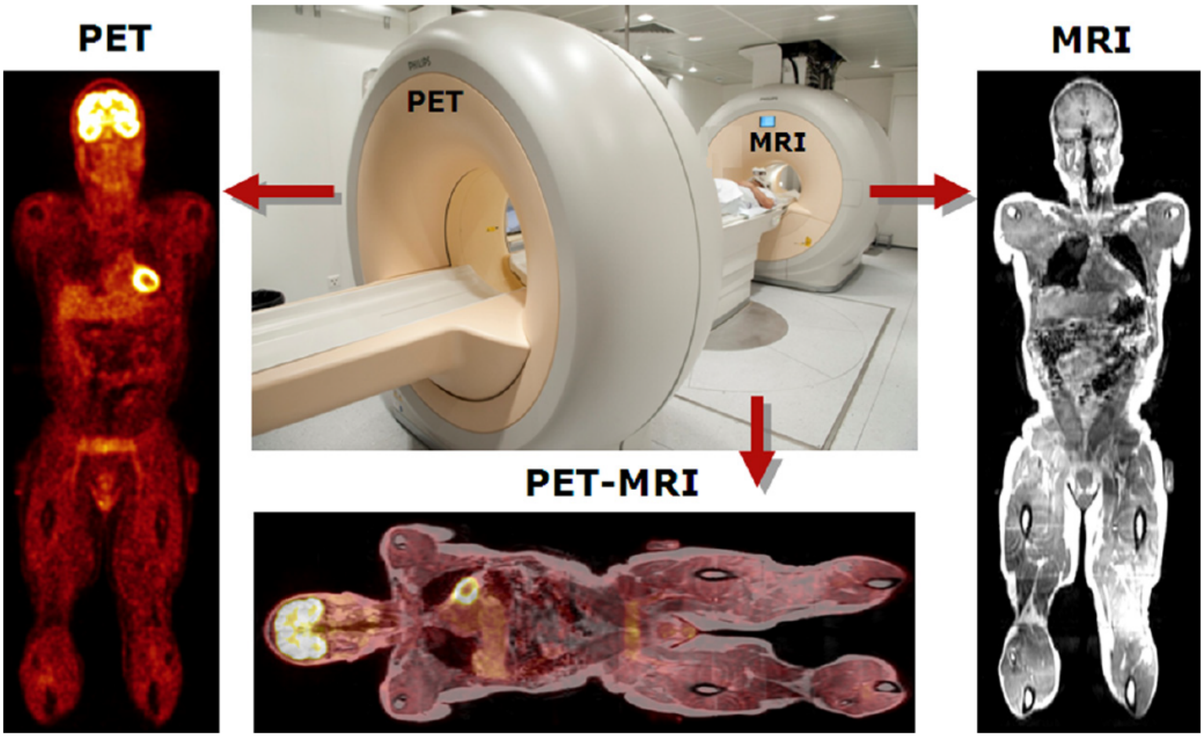


Figure 2.13: The sequential PET/MR scanner of the Philips whole-body Ingenuity system at Geneva University Hospital with MR, PET, and PET/MR images. Adapted from [92].

2.3 MR Images Segmentation-based Attenuation Correction

The segmentation of MR images to perform the PET attenuation correction task has been applied in the literature using different approaches starting with simple techniques such as level set [94, 95], thresholding [11, 96–101], and radon transform [102] till more complicated techniques such as clustering [6, 20, 22, 23], classification [1, 21, 24, 25] and deep learning [103, 104]. Table 2.5 summarises the different segmentation methods applied on different MR images sequences.

Table 2.5: The different segmentation methods applied on different MR sequences

Segmentation technique	Reference	MR sequence
Level set	[94]	STE
	[95]	UTE
Thresholding	[96]	Dual echo UTE
	[97]	UTE
	[98]	UTE and out-of-phase
	[99]	ZTE
	[100]	Dixon
	[101]	UTE
	[11]	UTE
Radon transform	[102]	T1-w
Clustering	[6]	STE and Dixon
	[20]	UTE and Dixon
	[22]	STE and Dixon
	[23]	T1-w
Classification	[1]	UTE
	[21]	UTE
	[24]	DCE, MP-RAGE, T1-w
	[25]	T1-w and Dixon
Deep learning	[103]	T1-w
	[104]	UTE, out-of-phase

2.3.1 Machine Learning based Segmentation Methods for PET Attenuation Correction in Brain Imaging

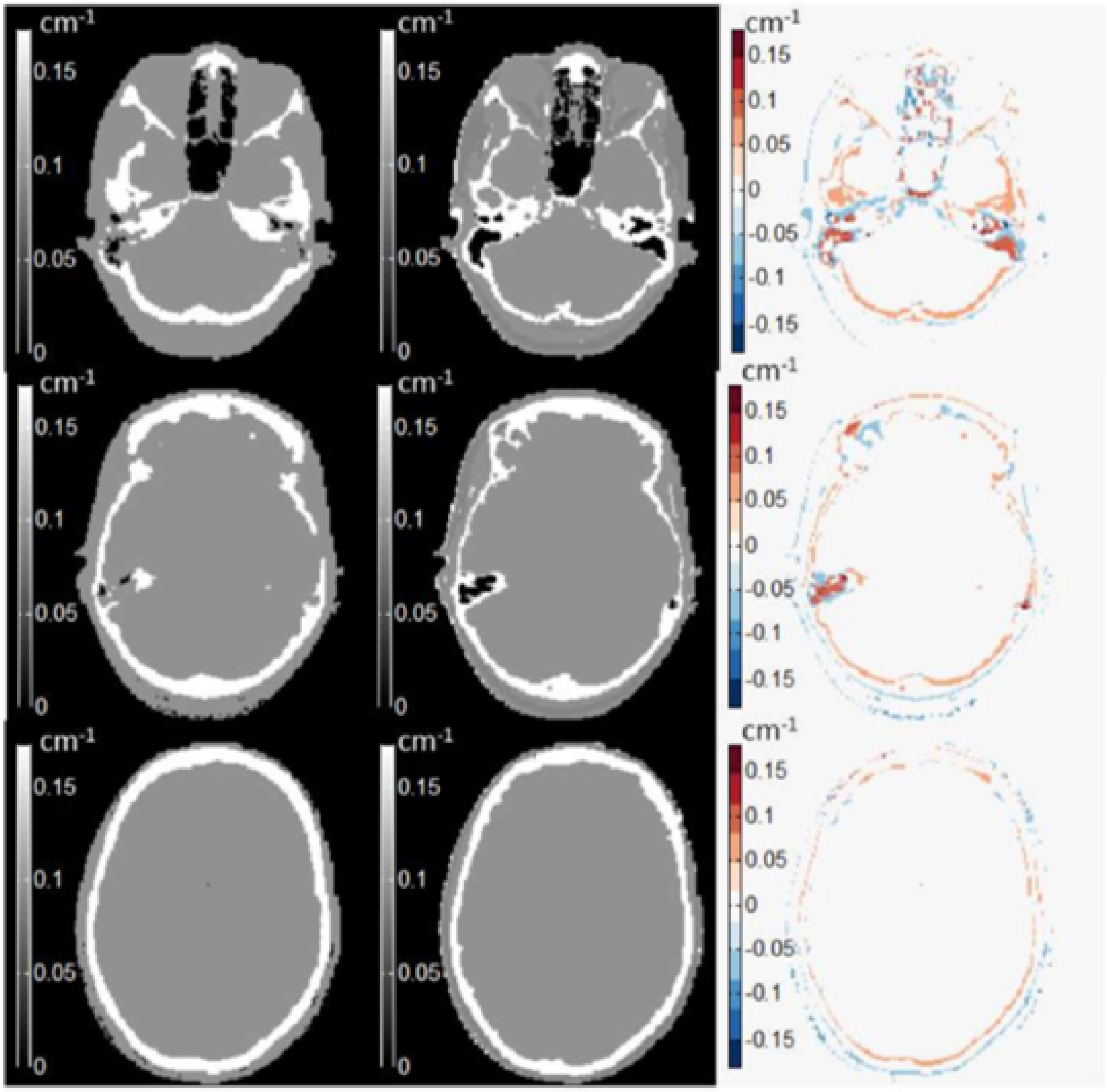
Table 2.6 summarises the three main categories of machine learning techniques proposed for MR images segmentation for PET attenuation correction: clustering, classification, and deep learning.

A Clustering

Su et al. [20] have used a multi-sequences approach for brain segmentation which are UTE transverse relaxation time map, Dixon-fat, and Dixon-water maps. They have used fuzzy C-means clustering to segment the head into five different tissue classes including brain, air, fat, fluid, and bone. Trajectory delay correction and global resizing factor have been used to evaluate the segmentation results. The obtained pseudo CT images are compared with measured CT images visually and objectively by calculating the CT histogram and the mean absolute predication deviation (MAPD).

Khateri et al. [22] have used a combination of STE sequences with 2-point Dixon technique and applied fuzzy C-means clustering to segment bone tissue. Later, the same team (Khateri et al. [6]) have applied fuzzy C-means clustering with morphological operations as post-segmentation technique to slightly enhance the segmentation of the brain into three tissue classes. They have measured the segmentation improvement by calculating the accuracy, sensitivity, and specificity. Figure 2.14 shows the MR images segmentation results with reference CT and the difference maps between the two modalities [22].

Fei et al. [23] have developed a multiscale segmentation approach using radon transform of T1-w MR images to segment the head into skull, scalp, and brain tissue. The segmented results are processed with anisotropic diffusion filter to construct a multiscale image series. The segmentation is refined by a modified fuzzy C-means technique to segment the brain into three classes: GM, WM, and CS. Afterwards, predefined attenuation coefficients are assigned to each tissue class. The segmentation results have been evaluated using DSC. The quality of PET images is compared with the results of transmission (TX)-based attenuation correction by visual inspection followed by quantitative measurements which are relative difference, mean squared error (MSE), and peak signal-to-noise ratio (PSNR).



(a) (b) (c)

Figure 2.14: MR images segmentation results achieved by [22] using clustering technique with the reference CT images (a), the segmented MR images (b), and the difference between the two modalities (c).

B Classification

Chan et al. [24] have proposed a segmentation method based on tissue classification to differentiate bone from air. Brieman's Random forest classifier has been trained using a set of features (gradient, statistical, and contextual) extracted from DCE, T1-c, and MP-RAGE MR sequences along with uncorrected PET images. The segmentation result of MP-RAGE images is shown in

Figure 2.15. With the use of CT images as ground truth, DSC for each tissue class, accuracy, area under curve (AUC), receiver operating characteristic (ROC) curve, and visual assessment have been used to evaluate the segmentation performance.

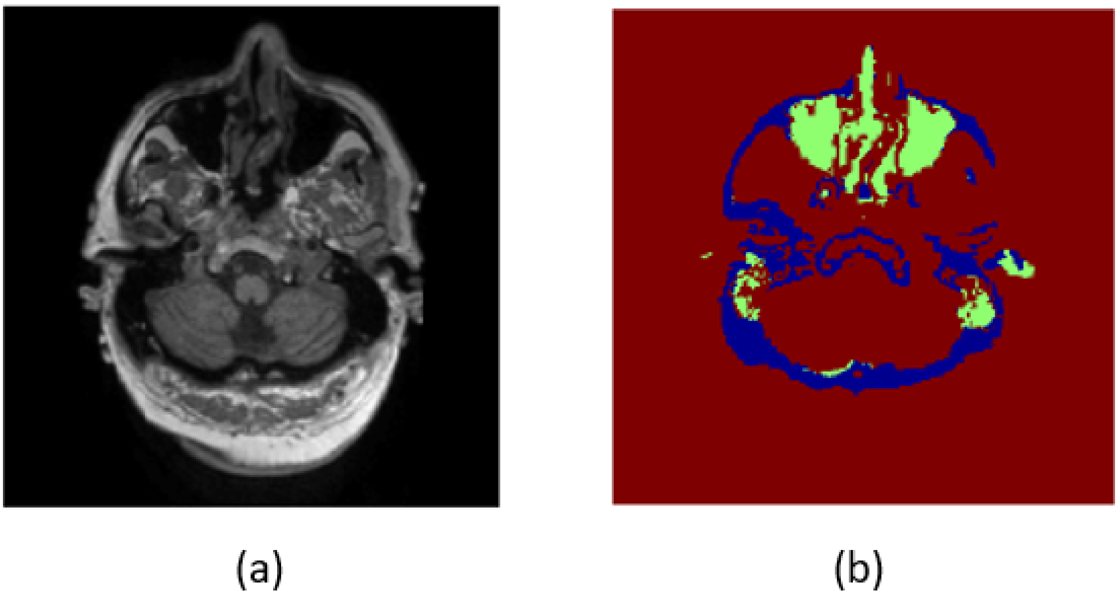


Figure 2.15: MR images segmentation results achieved by [24] using voxel classification to differentiate bone from air. MP-RAGE MR image (a) and segmented MP-RAGE MR image (b).

Koesters et al. [25] have proposed Adaboost classifier to extract bone tissue from T1-w MR sequences. Afterwards, the bone attenuation coefficient has been added to the manufacturer's Dixon-based attenuation map that reflects four tissue attenuation coefficients but not bone tissue. This method has been evaluated by comparing the standardised uptake value (SUV) estimation from CT, Dixon, and segmented images by the proposed model for whole-brain and regional analyses.

Shi et al. [1] has proposed a bone refinement method for existing attenuation correction maps. This method started with an existing attenuation correction map from the vendor or obtained from any segmentation-based method which are refined gradually by learning the relationship between the MR image and the attenuation correction map of the same patient. The learning process has been performed using multiresolution regional learning approach by applying support vector machine (SVM) classifier to refine the attenuation correction coefficients.

The resulting attenuation correction map was compared with vendor and CT based attenuation correction maps by measuring the bone recovery rate, DSC, voxel-wise error, and region-wise error.

Santos et al. [21] have performed a learning-based segmentation approach of the skull using probabilistic neural network. UTE sequences have been used to segment the brain into four classes: air, brain + soft tissue, CSF, and bone. This method has been compared with CT based attenuation correction map by calculating the accuracy and DSC as well as the visual assessment.

C Deep Learning

Deep learning is an emerging technology in machine learning which represents an advanced and more complex forms of neural networks. Deep networks are self-learning structures capable of learning high level image features and modeling a non-linear mapping between different image spaces through a convolution process. This methodology has shown its superiority in several medical applications, which paved the way to explore its efficiency to perform MR based PET attenuation correction.

Recently, there are various studies that have applied different deep network architectures to learn the mapping from MR to CT images to generate pseudo CT with continuous values such as generative adversarial network [105–109], U-Net [110], residual U-Net [111], and HighRes3DNet [112]. However, there are only few studies in the literature that have applied deep learning to perform brain MR images segmentation for PET attenuation correction [103, 104]. Each of these proposed studies have used different MR images sequences and network architectures.

Liu et al. [103] have applied a deep convolutional encoder decoder network called Segnet [113] to segment T1-w MR images. This work required a co-registration between CT and MR images before the training process and the creation of reference mask. To train the network, the reference labels have been generated from segmented CT images into three classes (air, soft tissue, and bone) using intensity-based thresholding technique. An attenuation coefficient value is assigned to each segmented class to generate the pseudo CT images illustrated in Figure 2.16. The mean DSC for each class has been calculated to evaluate the segmentation results. The PET quantification has been measured using the PET reconstruction error and compared with Dixon-based soft-tissue and air segmentation in addition to anatomic CT-based template registration. The main drawback of this architecture is the use of cross entropy loss as an objective function which does not take into account the class unbalancing issue. The number of bone voxels is much less than the air and soft tissue voxels in the brain volume; Hence, the classifier will be biased towards the majority class. There is a need to consider using an objective function which deals with the class unbalancing issue while training the network. This method shows promising results, yet the results are not cross validated and the segmentation results need further validation and testing.

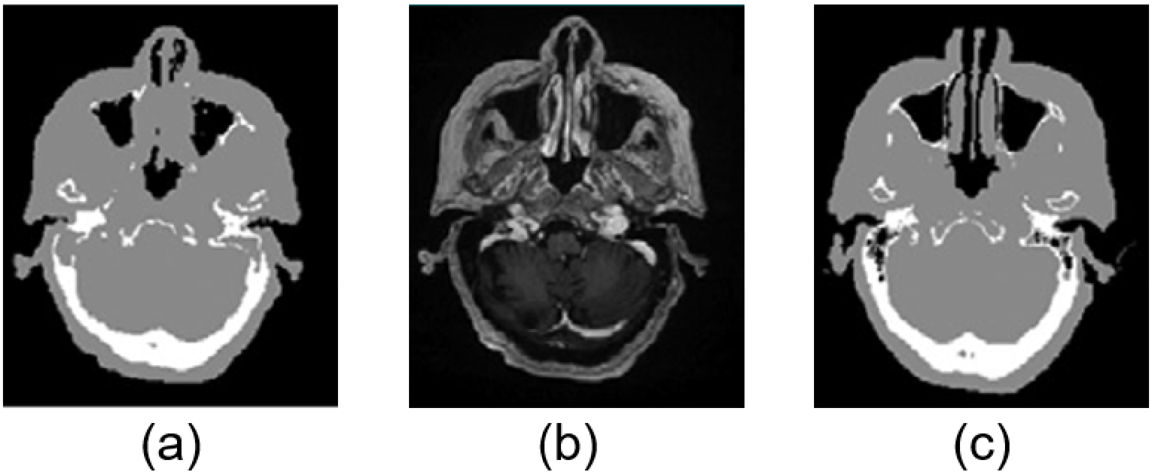


Figure 2.16: Pseudo CT image (a) obtained by segmenting T1-w MR image (b) with the use of CT image (c) as a reference mask [103].

Jang et al. [104] have used UTE and out of phase (fat and water) MR sequences which have been acquired using dual echo ramped hybrid encoding (dRHE) to segment the brain into three classes: air, soft tissue, and bone. The UTE sequences have been used as an input to retrain a pretrained deep network [103] that has been trained with T1-w MR sequences. Transfer learning has been applied to adopt the knowledge learnt from T1-w to UTE sequences in order to improve the learning and obtain a reliable training. The obtained segmented MR images have been processed using conditional random field technique to refine the segmentation results. Furthermore, the out of phase images have been used to segment the soft tissue into fat and water components using two-point Dixon based segmentation. The segmented labels from deep learning and Dixon based water and fat images have been integrated to generate the pseudo CT images. The proposed method has been compared with three other MR based attenuation correction methods while using CT based attenuation correction map as the standard reference. The DSC between the segmented images and the CT ground truth images has been calculated to evaluate the segmentation results. The relative PET error has been calculated to evaluate the corrected images. Although this method shows good results, it consists of complicated aspects such as the use of UTE sequences and the application of a post processing step which increases the method complexity.

2.3.2 MR images Segmentation based Attenuation Correction: Challenges and Opportunities

The MR imaging is the most preferred imaging modality in different medical applications. However, there are common challenges associated with MR images segmentation task including the misclassification of the pixels which determine the boundaries between different tissues. This may lead to false negatives where the lesion regions are classified as healthy and false positives where the healthy regions are identified as lesions. Hence, robust and accurate segmentation techniques are required to be used in the clinical routine.

Machine learning techniques such as neural networks, clustering, random forest, and SVM have shown promising segmentation results using different set of features. The combination of different techniques for different medical applications leads to better performance such as combining random forest classifier with Markov random field to segment white matter lesions in contrast enhanced FLAIR MR images [50] and combining regularised SVM with *k*NN classifier for hippocampus segmentation [114].

The comparison between the performance and efficiency of each technique is generally applied within the same context and same medical application. For instance, there are several studies that compare the segmentation results using different classifiers as represented in [40, 115–119]. The comparison study between different types of classifiers has shown the superiority of deep neural networks and especially CNN in the segmentation task. The conventional machine learning workflow composes of features extraction and selection then classification. However, the set of features control the performance of the segmentation methods rather than the classifier's type. Therefore, more attention on the development of the feature extraction techniques should be carried out with the employment of simple classifiers.

Although the deep learning approach has shown to be superior than any other machine learning techniques, it needs a good estimation of the numerous trainable parameters of the network and requires excessive training data and time to obtain the optimal weights which are used for prediction.

Hybrid PET/MR scanners have introduced the complementary nature of MR and PET images to clinical applications. They have also improved the PET quantification for diseases diagnosis and treatment planning by providing different tissue characteristics. Furthermore, they have reduced the acquisition time in the case of simultaneous acquisition. The merge of MR images modality into the field of PET attenuation correction and quantification has raised new challenges and difficulties.

MR images-based segmentation method is the most recommended for attenuation correction [120]. This method is easy to implement with low computational cost. However, it suffers from misclassifying the bone pixels due to the low T2 relaxation time that causes similar values for bone and air tissue in convent.ional MR sequences. Consequentially, the lack of good bone segmentation leads to inaccurate attenuation coefficient map which produces a strong spatial bias of the PET activity. For instance, the ignorance of bone attenuation coefficients in the head can lead to 20% underestimation of PET activity [121].

Table 2.6: Segmentation based MR Attenuation Correction Methods for Brain Imaging

Ref	Segmentation technique	MR sequence	Ground truth	Evaluation metrics
[6]	Fuzzy C-means clustering and morphologic operations	STE, Dixon	CTAC	Visual comparison Accuracy and specificity in the cortical bone, soft tissue, and air= 95% Sensitivity= 75% - 92% for 3 classes Correlation coefficient (R2) > 0.95
[22]	Fuzzy C-means clustering	STE, Dixon	MRI	Sensitivity, Specificity and Accuracy > 90% for the bone segmentation
[20]	Fuzzy C-means clustering	UTE , mDixon	CTAC	Mean estimated trajectory delay = 0.774 µs Average resizing factor =0.979 MPAD = -22 ± 29 HU and 130 ± 16 HU CT histogram
[23]	Fuzzy C-means clustering	T1-w	TXAC	DSC= 85.2 ± 2.6% Average relative difference= 4.2% Mean and standard deviation of the PSNR are 44.5 ± 6.9
[24]	Random forest	DCE, T1-w, MP-RAGE	NA	Visual comparison DSC (air)= 0.83±0.06 DSC (bone)= 0.98 ± 0.01 Accuracy= 0.96 ± 0.02 Area under curve (AUC)=0.9875±0.0002 p-value=0.85 ROC
[25]	Adaptive boosting	T1-w	CTAC	Whole-brain SUV estimation bias =95%

Table 2.7: Segmentation based MR Attenuation Correction Methods for Brain Imaging (Continued)

Ref	Segmentation technique	MR sequence	Ground truth	Evaluation metrics
[1]	SVM	UTE	CTAC	Bone recovery rate = 179 ± 57% Dice coefficient= 70 ± 34% Voxel-wise error = 27.8 ± 8.3% Region-wise error= 49.7 ± 6.2%
[21]	Probabilistic Neural Network	UTE	CTAC	Visual comparison Accuracy =92% DSC (for the whole head)= 0.66± 0.07
[103]	Deep learning: Convolutional encoder decoder	T1-w	CTAC	Mean DSC (air) = 0.971 ± 0.005 Mean DSC (soft tissue) = 0.936 ± 0.011 Mean DSC (bone) = 0.803 ± 0.021 PET reconstruction error = - 0.7 ± 1.1
[104]	Deep learning: Convolutional encoder decoder	UTE, out of phase	CTAC	DSC (air) = 0.76 ± 0.03 DSC (soft tissue) = 0.96 ± 0.006 DSC (bone) = 0.88 ± 0.01 Relative PET error = < 1%

2.3.3 Emerging Techniques

Deep learning-based methods have been proposed for pelvic and prostate PET attenuation correction. For instance, Bradshaw et al. [122] have applied a 3D CNN called DeepMedic [123] to segment pelvis T1-w and T2-w MR images for PET attenuation correction. Beside segmentation-based methods, there are other works that have applied deep learning for attenuation correction for brain [124] and pelvic [125] to learn the relationship between MR and CT images then generate pseudo CT images. Deep learning has shown its superiority to classical techniques for MR images segmentation. However, the application of deep learning to perform brain MR images segmentation for attenuation correction is limited.

There are other studies that have proposed deep learning-based methods for MR images guided radiation therapy and treatment planning for brain tumor [126–128], prostate/pelvic region [129] and other different whole-body tissues [130]. These studies have applied different deep models such as CNN, dilated CNN, and generative adversarial networks (GAN). A comparison study [126] between one segmentation, four atlas, and one deep learning methods to evaluate MR images-based radiotherapy planning in the pelvic region study has shown the outperformance of deep learning method in terms of segmentation accuracy, CT generation accuracy, and dosimetric evaluation.

In terms of MR sequences, currently there is room for improvement of the PET attenuation correction performance by using more sophisticated MR sequences with high signal intensity for bone such as UTE, ZTE, or Dixon sequences. These sequences are able to capture very short T2 values [99, 101] which enable accurate detection of bone tissue and, in turn, generate an accurate attenuation correction map. Another elegant way is acquiring UTE sequences at different echo time to extract more information. Moreover, Dixon sequences provide easy access to four tissue classes which are soft tissue, air, fat, and lung.

The usage of the special sequences such as UTE, ZTE, and Dixon images along with robust segmentation techniques and continuous linear attenuation coefficients enables the performance of PET attenuation correction. However, the major drawback of these sequences is the long acquisition time which potentially hampers the clinical flow.

2.4 Datasets

MR based attenuation correction task requires three different modalities which as MR, CT, and PET. In the literature, each research team is using their own non-public datasets to accomplish the task. Due to the lack of public datasets that can used specifically for MR based PET attenuation correction, some public datasets that are used to perform MR images segmentation are reported in Table 2.8. The contributions of this thesis can be evaluated with this list of datasets.

Table 2.8: Public MR images segmentation datasets.

Dataset	Task	MR sequences
Brain Tumour Segmentation (BraTS 2012-2020) [131]	Tumor Segmentation	T1-w, T1-c, T1-gd, T2-w, T1 Flair, T2 Flair
Medical Segmentation Decathlon [132]	Hippocampus segmentation	T1-w MPRAGE
BrainMetShare [133]	Metastases segmentation	T1 spin echo pre/post-contrast, T1 gradient-echo post, T2 Flair
OASIS [134]	Subcortical WM and deep GM structures segmentation	T1-w, T2-w, FLAIR, TSE, SWI, DWI, ASL, resting state bold, field maps, TOF
The University Children's Hospital Zurich [135]	Multi-tissue fetal brain segmentation	T2-w
Calgary-Campinas [136]	Skull stripping, WM, CM, and CSF segmentation	T1-w

2.5 Deep Convolutional Neural Network

The design of CNN follows the neural network architecture with the addition of convolutional layers to the existing fully connected layers. The development of different CNN architectures has been evolved rapidly after the big success of AlexNet [137] on 2012 which is considered the first deep CNN implemented on GPU to perform classification. Afterwards, many CNN's architectures have been proposed such as LeNet [138] which is a deep CNN developed by Google, ResNet [139] which employs residual blocks to enhance the features extraction operation, and the very deep VGG network [140] that has been trained using millions of natural images. For segmentation tasks, the CNN is adapted to fully convolutional network (FCN) [141] which omits the fully connected layers and retrieves the output with the same resolution as the input images. Later, dilated network which addresses the limitation of FCN has been proposed by [142].

In the medical domain, the segmentation task has been widely addressed using the well-known 2D convolutional encoder decoder architecture named U-Net [143] which employs paths from high resolution to low resolutions features maps. The 3D nature of medical images opens the door to propose 3D CNN architectures such as V-Net [144].

Recently, the classification and segmentation tasks of 2D and 3D images have been addressed using attention based networks which apply features recalibration mechanism to boost the features extraction process. Hu et al. [145] were the first team who proposed a 2D channel attention based CNN to address classification problems by adaptively recalibrate the channel wise features. They have proved that this recalibration process brings significant enhancement in classifying public natural images from LSVRC 2017 datasets. They have also evaluated the performance of this attention block using different CNN architectures to perform classification.

Later, this architectural unit which excites the features in channel wise has been embedded into segmentation based CNN. For instance, Yu et al. [146] have proposed BiSeNet, a 2D segmentation CNN which applies the channel attention block. This work has shown the capability of this block to improve the segmentation problems in computer vision domain. BiSeNet has been tested using Cityscapes, CamVid, and COCO-Stuff datasets. The same team have also proposed another 2D segmentation architecture [147] which takes the low and high level features as an aggregated input to the channel attention mechanism aiming at recalibrating the network features. This approach has been tested using PASCAL VOC 2012 and Cityscapes datasets. Moreover, Cheng et al. [148] have applied the channel attention block to propose two gated structures which can be integrated with CNN for 2D images segmentation. This network has been evaluated using CamVid and VOC public datasets.

The success of the channel attention mechanism has inspired the team of Bian et al. [149] to investigate the possibility to reduce the model complexity which is increased by adding the channel attention to CNN. They have proposed a dual channel attention network that reduces the number of additional parameters by using a lightweight channel attention module where the fully connected layer is replaced with 1D convolutional layer. The performance of this approach has been evaluated by performing multiple tasks such as classification, segmentation, and object detection using CIFAR-100 and MS-COCO datasets.

In the medical field, this mechanism has also shown its capability to boost the 2D medical images features and improve the breast cancer classification [150] and prostate zonal segmentation [151]. Another improvement in the medical images segmentation is proposed by Roy et al. [152, 153]. They have firstly suggested a spatial attention block then explored the concurrent spatial and channel attention within the FCN architecture. They have shown that the spatial attention is more useful for segmentation tasks rather than the channel attention using MR and CT images. The channel and spatial attention mechanisms are combined on parallel within the network architecture to recalibrate the features concurrently. They have observed consistent improvement of performance across three different 2D architectures (DenseNet, SD-Net, and U-Net). Afterwards, Wu et al. [154] have applied the concurrent channel and spatial attention mechanism to segment the skin lesions using an adaptive dual encoder network with dual attention module.

Due to the necessity of applying 3D CNN to segment medical volumes of different modalities, the 3D attention mechanism has also been proposed. For instance, Liu et al. [155] have suggested a modified V-net architecture and integrated the 3D channel and spatial attention blocks on parallel to segment the brain tumor using multi sequences MR images. Additionally, Hou et al. [156] have applied the 3D attention block to segment white matter hyperintensities lesions in MR images using the parallel combination of channel and spatial attention blocks.

Far from the medical domain, the 3D channel attention block has been applied in the remote sensing of sea ice field by Han et al. [157] who have designed a 3D channel attention based CNN to classify the ice sea areas using satellite images. They have also shown that this mechanism is efficient to accomplish this classification task.

2.6 Handcrafted Features with CNN

The process of automatic extraction of features using CNN has shown its superiority in various applications to perform different tasks. The CNN based features remove the burden of handcrafted features task which includes features extraction and selection. However, in some problems and with specific types of data, the handcrafted features are critical to enhance the performance [158, 159].

The combination of handcrafted features with CNN's features is becoming recently a popular approach to address various problems in different domains such as image scene geometry recognition [160], classification of working condition in froth flotation [161], signal gesture recognition [162], land cover content-based image retrieval [163], and pedestrian detection [164]. Moreover, it has been widely used in the medical field to perform classification, segmentation, detection, and others tasks. Additionally, this approach has been applied to both images and signals such as MR images [165], histology images [166], iris images [167], electromyographic (EMG) signals [162], and electroencephalogram (EEG) signals [168].

The most common handcrafted features reported in the literature which are combined with CNN's features are statistical and textural features from the source images. For instance, Shang et al. [165] have combined radiomics statistical, textural, and shape features with CNN's features to perform soft tissue tumor classification, Khan et al. [160] have applied a different set of statistical features to segment the brain tumor, and Choudhary et al. [167] have combined another set of statistical features with CNN's features to perform iris spoofing detection.

Multiresolution analysis techniques such as Contourlet and Shearlet transforms provide multiscale and multidirectional features by decomposing the image into low and high frequency sub-bands. Unlike Wavelet transform, they are shift-invariant and can capture rich information with different scales and directions in the spectral domain. The CNN is able to extract spatial features; Yet, the directional details in the spectral domain are neglected. The addition of handcrafted features that capture multiscale and multidirectional information is proposed in the literature to improve the performance of different tasks as described in the following subsections.

2.6.1 Contourlet Transform

The Contourlet transform is combined with CNN's features to perform multiple tasks including medical images fusion and multi-modality merging [158, 169, 170], natural images classification [171, 172], remote sensing classification [173], breast cancer classification [174], Seismic images denoising [175], and bone age assessment [176].

Li et al. [173] have proposed a complex domain CNN which uses Contourlet filter banks for polarimetric synthetic aperture radar (PolSAR) image classification. They have suggested to redefine the operations of the conventional CNN's layers in complex domain to provide more useful features. They have also employed non subsampled Contourlet (NSCT) filter with complex CNN filters to obtain multiresolution and multidirectional information. The proposed network has been tested on three PolSAR data sets and has shown its efficiency to capture abstract features in different directions and frequency bands. Additionally, Liu et al. [171] have proposed a Contourlet based CNN for SAR image despeckling. They have designed multiple multidirectional and multiscale subnetworks. Each independent subnetwork takes a different Contourlet sub-band as input and trains the network until it reaches a clean image with minimal noise. The inverse Contourlet transform is applied to the output of all subnetworks to obtain the clean image. They have applied a different CNN's architectures for each sub-band to fit the properties of each sub-band. The proposed approach has been evaluated using synthetic and real SAR images. This approach has been compared with the state of the art methods and has shown its superiority in suppressing speckle with minimal required runtime.

In the medical domain, the employment of the Contourlet transform with CNN is proposed by Nahid et al. [174] to classify histopathological images. They have conducted a comparison in the performance between CNNs that take different sets of inputs including raw images, statistical information of the Contourlet sub-bands, histogram information, local binary pattern features, and other frequency-domain features. They have found out that the statistical information of the Contourlet coefficients combined with histogram information outperform other approaches in terms of different evaluation metrics. The bone age assessment using X-ray images has been addressed by Liu et al. [176] where they have employed the NSCT coefficients of the raw images as inputs to the CNN. Each NSCT sub-band is fed to a single CNN then the output of all networks are merged to perform the regression task. By comparing this approach with other approaches that adopt spatial domain images as network input, the proposed method is superior for the bone age assessment task.

2.6.2 Shearlet Transform

The Shearlet transform has been recently employed with CNN to perform various tasks in different domains such as RGB images classification [177], remote sensing images denoising [159], panchromatic and multispectral images fusion [178], and classification of video quality [179]. In the medical field, the Shearlet coefficients have been widely combined with CNN's features to perform CT and MR images fusion [180, 181], multi-modality MR images fusion [182], cancer classification [183, 184], Alzheimer's disease classification [185], prostate Gleason grading [183], and image enhancement and denoising [170].

For instance, Rezaeilouyeh et al. [183] have applied classification of histopathology images using the magnitude and phase of Shearlet coefficients along with raw images as input to CNN. They have employed one CNN for each raw RGB image, Shearlet sub-band phase and magnitude. The outputs of all networks are concatenated using one fully connected layer. This method has been evaluated on performing the classification of breast cancer and Gleason grading tasks. They have also shown that this method performs better than other methods which use different set of handcrafted features. Similarly, Liang et al. [184] have applied the same approach of combining the raw histopathology images with their phase and magnitude coefficients of the

Shearlet transform to identify colon cancer. They have also shown the capability of Shearlet coefficients to enrich the CNN's features and improve the classification performance. Jabason et al. [185] have combined the PET images with their Shearlet coefficients as an input to the CNN to classify the different types of Alzheimer's Disease. Moreover, Yang et al. [177] have proposed a sparse representation CNN layer which employs Shearlet and Wavelet transforms to boost the feature extraction operation. The sparse layer generates non-trainable low pass and high pass feature maps during the back propagation process. They have evaluated the proposed approach to perform classification on five different natural images datasets (FOOD-101, CIFAR10/100, DTD, Brodatz and ImageNet). The results have shown that the multidirectional representation is able to enhance the classification performance where the best classification accuracy of different classes is achieved by either the Shearlet or the Wavelet transforms.

2.7 Ensemble of Deep Networks

The combination of multiple machine learning models is an effective approach to improve the performance of the targeted task and build a reliable decision. This aggregation is well-known by ensemble method which consists of weak individual models that are aggregated on parallel such as bagging algorithms or sequentially such as boosting algorithms [186].

This concept has been widely applied in different applications to perform multiple tasks such as classification [187], segmentation [188], and object detection [189]. Recently, with the success of deep neural networks in various domains and applications, the ensemble method has been extensively and efficiently applied with deep neural network to solve multiple problems [190–193].

The combination of multiple deep networks has been applied to address transportation problems such as autonomous driving [194], rail surface cracks detection [195], and traffic sign recognition [196]. This concept has been also introduced to the field of agriculture and food by aggregating multiple deep models to classify citrus pests [197] and insects [198] and recognise food images to reduce obesity [199]. The performance of the application of face recognition has also been improved by combining multiple CNNs to detect the subject drunkenness using thermal infrared images [200] or detect fake images [201]. Additionally, the face images quality assessment, hallucination, and reconstruction have been performed using ensemble models [202–204]. Beside that, the ensemble method has shown its efficiency in the task of human action recognition while classifying the human activities using different types of images [205, 206]. Moreover, the ensemble of deep networks has been applied widely in the field of remote sensing scene by classifying and segmenting radar and hyperspectral images [191–193, 207–209].

The focus of this review is the application of the ensemble of deep networks in the medical domain with different medical imaging modalities and applications. In the last three years, the ensemble of deep networks has been applied extensively to perform classification and moderately to perform segmentation. The utilisation of deep networks ensemble has been employed in the diagnosis of different diseases using different imaging modalities. The diagnosis of COVID-19 virus has been addressed by classifying the X-ray images to detect pneumonia using ensemble of deep networks [210–215]. Moreover, the detection of skin cancer and lesions has been performed by introducing the deep network ensemble using dermoscopic images [216–220]. The employment of the deep networks ensemble has shown its ability to improve the classification using different imaging modalities such as ultrasound [221, 222], mammography [223, 224], microscopic [225, 226], CT [227], MR [228, 229], and retinal [230–232] images to diagnose multiple diseases like Alzheimer, ocular diseases, breast cancer, thyroid disorder, and fat assessment. Other different tasks such as the estimation of the brain and bone age [233, 234], detection of region of pneumonia in the chest [235], and medical image enhancement [236] have been accomplished efficiently with the application of deep networks ensemble.

Recently, the application of the ensemble of deep networks to address the segmentation task of medical images has not been thoroughly investigated. For instance, Coupe et al. [237] have proposed an ensemble of one hundred twenty five 3D U-Net models where each takes different overlapped sub-volumes as inputs. Each single network performs local segmentation, then by aggregating the networks using the majority voting, the global segmentation of the volume is accomplished. This approach has shown good results in segmenting the brain MR images.

Karthik et al. [238] have designed an ensemble of deconvolutional networks in the decoder part of the CNN by combining multi-scale features from different encoder levels. Each decoder has different number of convolutional layers depending on the resolutions of the features that are coming from the encoder path. The majority voting for each class is applied to combine the deconvolutional networks. This approach has shown superior results in segmenting the brain 3D MR images with four different modalities, namely T1-w, T2-w, DWI, and FLAIR.

Hou et al. [156] have proposed a 3D MR images segmentation approach for brain white matter hyperintensities using an ensemble of two U-Net architectures which employ different loss functions (dice loss and focal Tversky loss) to achieve a trade-off between specificity and sensitivity. The two CNNs are aggregated by taking the average of each model's probability map.

Mei et al. [239] have applied the concept of ensemble to segment CT images for the nasopharynx cancer diagnosis. They have used three different sampling approaches to overcome the class imbalance issue with the employment of a 2.5D Unet for each sampling strategy resulting in an ensemble of three CNNs of multi scale features. Finally, the three single models are combined by applying the averaging operation.

Goyal et al . [240] have used 2D dermoscopy images to build an ensemble of two CNNs: Mask R-CNN and DeeplabV3. The DeeplabV3 is an encoder decoder that follows the Xception65 CNN architecture while the mask R-CNN is a segmentation network which is inspired by the R-CNN architecture for object detection. Both networks are pretrained with public natural images then fine tuned to segment the skin lesions. The networks are combined using three methods which are the addition of two outputs, the selection of the largest area, and the selection of the smallest area . They have found out that the aggregation method of selecting the smallest area improves the segmentation specificity while the two other methods improve the model's sensitivity.

Finally, Tan et al. [241] have performed 2D segmentation for skin and blood cancer images using an ensemble which consists of three dilation based CNN networks. The CNNs are combined using the majority voting technique. Each base model is optimised using the proposed enhanced particle swarm optimization approach.

2.8 Transfer Learning

The transfer learning refers to either transferring knowledge or features from one domain to another or performing a different task within the same domain. Recently with the wide use of deep learning, transfer learning has been applied extensively by sharing the learned weights or the set of extracted features of the pretrained model while changing the number of neurons in the output layer depending on the target task. Another view of transfer learning is training the deep model using the data of one domain then testing it with the data of another domain. Another approach is training and testing the deep model with the data of two different domains while performing the same task. The focus of this section is the application of transfer learning with deep neural networks to perform different tasks on medical images.

Despite the disparity between different domains images, the transfer learning from one domain to another has several advantages. It enables the training of deep networks with small datasets, reduces the overfitting, and decreases the training time. Transferring the knowledge from a pretrained network provides a ready set of generic features such as edges, boundaries, and colors which are useful for any image analysis task.

There are different approaches of transfer learning. First, the knowledge can be transferred by using only the pretrained weights to initialise the new model instead of random weights initialisation which tends to cause overfitting with the training data [242]. Second, the pretrained model's weights can be fine tuned using the new datasets. The fine tuning process requires freezing some layers of the network while retraining others. The fine tuning process can be limited only to retrain the last fully connected layer to perform the new task or work with the new domain data.

Transfer learning can be applied to perform the same task but using data from a different domain. This type is called domain adaption such as using datasets from different hospitals, different scanners, and different scanning protocols. On the hand, transfer learning can convey the knowledge to perform different tasks from different domains. This approach has been widely applied by using the pretrained models with natural images to perform medical images classification and segmentation tasks.

In the medical domain, the application of transfer learning to perform the same task with datasets from different domains has been applied extensively to segment MR images for different diseases. For instance, Kushibar et al. [243] have used brain MR volumes which are acquired with different scanners and protocols to segment sub-cortical brain structures. They have applied transfer learning by training the model with one dataset then fine tuning the fully connected layers to segment the images of the second dataset. They have concluded that transfer learning yields better results than the network which is trained from scratch using the same number of training images. They have also found out that the obtained results with less training images using transfer learning are similar to training a network from scratch.

The application of domain adaptation to segment different MR sequences has been investigated to segment left ventricle. For instance, Vesal and his team [244] have used three different cardiac MR sequences: late gadolinium enhancement (LGE), T2-w, and balanced steady-state free precession (bSSFP). Firstly, they have trained a network from scratch using two MR sequences (T2-w and bSSFP) then used this network to initialise another network that aims at segmenting the third MR sequence (LGE) using less number of images. Although this approach has has shown to be efficient on 2D images, it has not performed well on 3D images due to the limited number of slices. Moreover, the results have shown the efficiency of domain adaptation with augmentation techniques to improve the segmentation results of the left ventricle.

Ghafoorian et al. [245] have used only two MR sequences: T1-w and FALIR images. They have conducted experiments to study the optimal size of the new domain data. They have also performed several experiments to study the relation between the size of the new domain training data and the number of the network's layers that should be retrained to achieve the optimal segmentation results. They have found out that with only few training images of the new domain, the pretrained model with another MR sequence can be fine tuned by just retraining the last fully connected layer to achieve an accurate domain adaptation model.

Additionally, Chen et al. [246] have performed left ventricle segmentation on cine MR images by transferring the knowledge from a pretrained model on a public human cine MR dataset. The results of DSC have shown that the trained model with transfer learning outperforms the trained model from scratch.

Kessler et al. [247] have applied transfer learning from a different domain to carry out knee tissue segmentation using three datasets consist of different MR sequences acquired by multiple scanners (GE, Siemens, Philips, Toshiba, and Hitachi). They have found out that the transfer learning helps to improve the segmentation accuracy of few knee joint tissues of the new domain data while it preserves the model's capability to segment the source domain data.

Zhao et al. [248] have used transfer learning from different MR sequences (T2-w and DWI) from four medical centres for detection and segmentation of lymph nodes. Hippocampus segmentation of two different T1-w MR datasets using domain adaptation has been proposed by Ataloglou at el. [249]. The deep network is initially trained with one dataset then fine tuned using only fifteen samples of the second dataset. They have shown that the transfer learning with fine tuning from one dataset to another improves the DSC of hippocampus segmentation. Besides the segmentation tasks, the domain adaptation has been successfully applied to perform MR images synthesis [250], reconstruction [251], and classification [252],

On the other hand, transfer learning can be applied to transfer the knowledge from models trained to perform different tasks on different domains. Many medical imaging tasks are conducted by transferring the knowledge from pretrained models on natural images such Imagenet dataset which consist of fourteen millions of 2D images.

For instance, Grimm et al. [253] have performed segmentation of brain volume and cerebrospinal fluid tissue on simulated T2-w MR images by transferring the knowledge from the pretrained VGG16 network. An encoder-decoder Segnet architecture is preinitialised with layers and weights from a pretrained VGG 16 model then fine tuned by reducing the learning rate.

Brain tumor segmentation using transfer learning has been applied by Cui et al. [254] who have proposed a FCN which takes four input channels of four different MR sequences named FLAIR, T1-w, T2-w, and contrast enhanced T1-w (T1-c). The initial weights of the network are initialised using the weights of the pretrained model with ImageNet. Yet, the fourth input channel is initialised with the average weight of the three input channels. The model is fine tuned afterwards using the SGD approach. They have found that this method obtained promising segmentation results with faster computation time.

Kuzina et al. [255] have proposed a novel transfer learning approach which is based on deep weight prior concept. They have used a U-Net based generative Bayesian prior network which takes the trained convolutional filters with the source dataset to learn the prior on the target dataset. They have evaluated this method to perform segmentation for two different diseases: brain tumor and multiple sclerosis using two different datasets of MR images. They have found out that the trained network with the proposed approach outperforms the network that is initialised randomly or from a pretrained network.

Recently, brain tumor classification of MR images has been addressed widely with transfer learning from models trained with natural images such as AlexNet [182, 242, 256], Vgg16 [257], Vgg19 [258], ResNet34 [259], ResNet 101 [256], ResNet 50 [256], GoogLeNet [256], and SqueezeNet [256]. Transfer learning from natural images to accomplish medical images tasks has also been tested on MR images enhancement [260], cine MR images super-resolution [261], and detection of meniscus region on MR images [262].

2.9 Summary

This chapter has represented a brief overview of the most common and useful medical imaging in the clinical routine nowadays: PET, MR, and CT imaging. PET imaging provides the tissue function and physiology at the molecular level in the shape of a map of radioactive tracer concentration. CT imaging is a tomographic X-ray based modality that represents the X-ray absorption coefficient and considered a useful tool to detect the bone tissue. MR imaging is an excellent modality to capture the soft tissue contrast using a map of the local transverse magnetization of the hydrogen nuclei. The hybrid imaging of PET/CT and PET/MR are enabling the sequential or simultaneous acquisition of both modalities to bridge the gap between molecular and systems diagnosis. In this research, MR and CT images are used to develop the segmentation algorithm to address the application of MR based attenuation correction for brain PET images.

MR based attenuation correction methods for PET images have been introduced in this chapter with the focus on the segmentation approach. The application of segmentation using deep learning technique has been reviewed where various enhancement aspects have been discussed including the combination of handcrafted features with deep learning features, the attention based CNN, the ensemble modelling, and the transfer learning from different medical imaging modalities or sequences.

The MR images segmentation for PET attenuation correction is a simple and robust method which has been adopted on commercial scanners for a long time. However, the segmentation of bone in the brain images is very challenging due to the similarity of air and bone intensity values and the minority of bone pixels compared to other tissue types.

MR segmentation based attenuation correction has been addressed using deep learning technique, Yet, the existing literature have applied the existing CNN's architecture without any adaptation. The combination of handcrafted features with CNN's features is one possible enhancement for MR images segmentation for PET attenuation correction which has not been proposed yet in the literature. The use of multiresolution analysis features such as Contourlet and Shearlet transforms to enhance the segmentation task in the medical domain have shown the efficacy of the addition of multidirectional and multiscale features to the CNN's features. However, the proposed methods in the literature have employed complex solutions including multiple cascaded CNNs for different multiresolution sub-bands. Moreover, they have used all resultant coefficients without any dimensionality reduction. A less complicated design is required to combine the multiresolution features with CNN's features.

Recently, the attention based CNN has shown its capability to enhance the segmentation tasks of different natural and medical images using two approaches of attention namely spatial and channel excitations. This excitation technique recalibrates the extracted features of the CNN's layers by emphasising on the most useful ones to accomplish the desired task. Some studies have employed the parallel combination of both mechanisms of excitation to segment skin lesions, white matter hyperintensities lesions, and brain tumor. Yet, it has not been applied to tackle the MR segmentation based attenuation correction problem. The parallel combination results on increasing the number of the CNN's parameters, the model's complexity, and the computation time and resources. There is a need to explore different ways to integrate the excitation blocks with CNN to enhance the network performance with minimal increase of the model complexity.

Classifiers ensemble or committee machine is another technique which has been used for a long time in the field of machine learning to enable the generalisation of the model, enhance the model performance, and produce reliable results. This approach has been applied with deep networks where each base model represents a single deep network. Many studies have proposed the deep network ensemble to segment different medical images (MR, CT, and dermoscopy) for medical applications such as cancer detection, ischemic lesion segmentation, and brain tumor. The MR segmentation based attenuation correction has not been addressed yet using the ensemble of CNNs. Additionally, the existing studies have applied multiple CNNs with the same architecture. However, the induction of variation between the ensembled CNNs is an important factor to generalise the model's performance and create diversity between the base CNNs.

Finally, the application of transfer learning has became a crucial step to build deep networks especially with the availability of trained networks with millions of images. The pretrained networks with natural images have been used as a starting point to build new networks to perform various medical tasks. They have also applied transfer learning using MR images which are acquired with different scanners or from different institutions. Moreover, the transfer learning has been applied for different MR sequences such as transferring the knowledge from T1-w to FLAIR MR sequences. There are multiple MR sequences which worth investigating their applicability of transfer learning from different sequences especially the most commonly used in the clinical routine (T1-w and T2-w) to perform the same or different tasks.

Chapter 3

3D Deep Convolutional Neural Network with Handcrafted Features for MR Segmentation

3.1 Introduction

This chapter describes an approach to improve the brain MR images segmentation to generate pseudo CT images for PET attenuation correction. This approach involves deep CNN technique that segments the brain into three tissue classes namely: air, soft tissue, and bone. The main objective is to improve the bone segmentation performance by conducting the following contributions:

- Designing a 3D deep CNN by studying the impact of different components of the network.
- Applying dice coefficient as an objective function to train the 3D CNN to solve the issue of class unbalancing.
- Introducing additional handcrafted features such as directional coefficients of non-subsampled Contourlet (NSCT) and non-subsampled Shearlet (NSST) and aggregating them with

CNN's features.

The brain T1-w MR images of fifty patients are used to evaluate the proposed methods by calculating the precision, sensitivity (recall), DSC, and JSC of each class of the brain. Moreover, the visual assessment of the segmentation results of some randomly selected slices from the testing data is performed. This chapter is structured as follows: the next section gives detailed information about the utilised clinical brain images which include T1-w MR and CT images. The description of the proposed deep CNN as the baseline model is provided in section 3.3. The methodology of employing handcrafted features with different aggregation methods with the CNN are provided in section 3.4, 3.5, and 3.6. The obtained results are revealed in section 3.7 with their quantitative and qualitative evaluation. Finally, the summary with the limitations are represented in section 3.8.

3.2 Brain Datasets

3.2.1 Data Acquisition

The brain dataset that is used for the experimental work consists of fifty patients which undertook 18F-FDG PET/CT and MR scanning. The patient's demographics and clinical characteristics are summarised in Table 3.1. The research protocol and the utilisation of the data are firstly approved by each patient then by the research ethics committee of the College of Engineering, Design and Physical Sciences at Brunel University London.

Table 3.1: The datasets demographics and clinical characteristics.

Gender	Age (Mean ± SD)	Clinical diagnosis
28 women and 22 men	61±12 years	44 neurodegenerative disease, 3 epilepsy, and 3 brain tumors

The MR images were acquired using a 3 T MAGNETOM Skyra (Siemens Healthcare, Erlangen, Germany) with a 64-channel head coil. The MR scans are 3D T1-w (MP-RAGE) sequence that were acquired using the following parameters: TE= 2.3 ms, TR= 1900 ms, T1= 970 ms, flip angle 8°, NEX = 1, and matrix dimension of 255 × 255 × 250 with voxel size of 0.86 × 0.86 × 1 mm. The 18F-FDG PET/CT scans were acquired using two different Siemens scanners (Siemens Healthcare, Erlangen, Germany): Biograph mCT scanner and Biograph 64 True Point scanner. The matrix size of the CT images is 512 × 512 × 150 with voxel size of 0.97 × 0.97 × 1.5 mm. An example of a single slice of T1-w MR image, CT image, and MR overlaid on CT image are illustrated in Figure 3.1 in transaxial, coronal , and sagittal views.

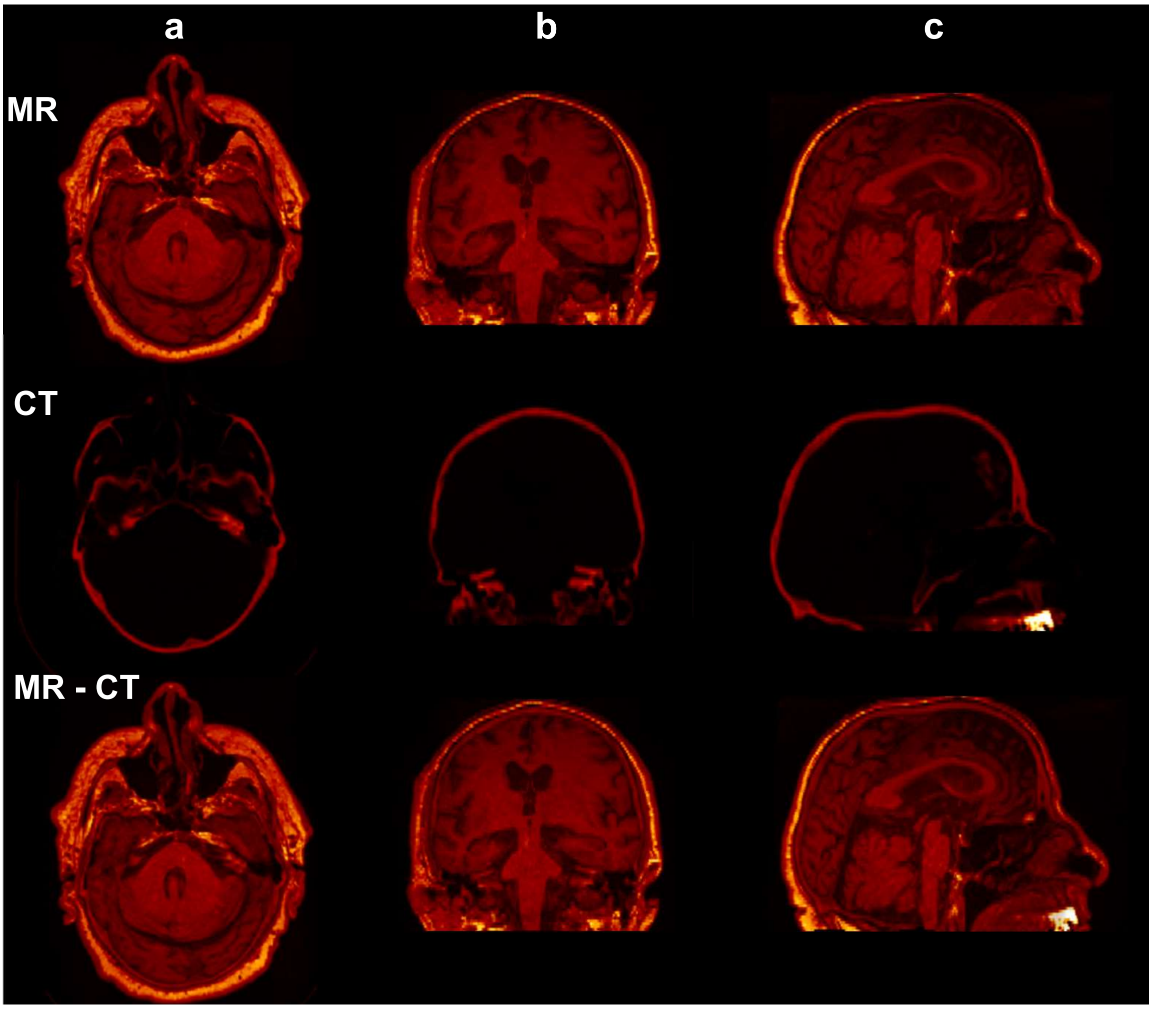


Figure 3.1: An illustration of a single slice of MR, CT, and MR overlaid on CT in (a) axial, (b) coronal , and (c) sagittal views.

There is a class imbalance of the three brain tissues where each brain volume consists of higher number of pixels that belong to soft tissue and air classes compared to bone class. For instance, the percentage of bone, soft tissue, and air pixels of 10 selected patients are 7.17%, 52.68%, and 40.15%, respectively.

3.2.2 Data Preprocessing

Although the raw data is in 3D format, the preprocessing of both T1-w and CT images is performed in 2D fashion by firstly identifying the region of interest by selecting 48 slices from each patient's volume. Each patient's volume consists of different number of slices with a maximum of 165 slices and a minimum of 82 slices. The slices that reside at the beginning and the end of each volume are blur and not providing any useful features to learn from. Based on a manual observation and checking process, 48 consecutive slices of each patient are selected by removing the blurred and useless slices.

The dimensions of each volume are reduced by removing some pixels of the background where each slice is cropped into 256 × 256 × 48 matrix dimension. Due to different resolutions, some of the volumes are resampled to 300 × 300 matrix using bilinear interpolation to include the whole brain area before cropping. The resampling size 300 × 300 is selected experimentally. Local contrast normalization technique is applied on MR images to downscale the range of the pixel's values.

3.2.3 Ground Truth Generation

CT images are used to generate the ground truth to conduct the supervised learning approach. Each MR slice is co-registered with its corresponding CT slice to obtain a common coordinate system and overcome the temporal gap between the acquisition of MR and CT images (2-3 days). The registration process starts with the application of rigid Euler transformation followed by the non-rigid B-spline transformation using a well-known medical imaging registration tool named Elastix [263]. Afterwards, the labelling of CT images is performed by applying simple pixel intensity-based thresholding to segment the brain into three tissue classes which are air, bone, and soft tissue. The thresholds of the CT Hounsfield values for each tissue class are illustrated in Table 3.2.

Table 3.2: The CT Hounsfield values threshold for generating the ground truth images.

Hounsfield value threshold	Tissue class assignment
HU > 600	Bone
HU < -500	Air
Others	Soft tissue

3.3 3D CNN for MR Segmentation

3.3.1 The Baseline CNN Architecture

The baseline CNN follows the well-known 2D U-Net architecture [143] that has been applied widely in various medical applications [264]. This network is a fully convolutional architecture that takes the shape of encoder-decoder with the inclusion of additional paths that links the high-resolution with the low-resolution features. The additional paths increase the resolution of the output and improve the localisation.

The proposed network is amended by decreasing the number of down-sampling layers which cause loss of useful features. Additionally, the number of convolutional layers is also increased by setting three convolutional layers at each depth. The 3D CNN consists of twenty one 3D convolutional layers where each three convolutional layers are followed by batch normalization layer, rectified-linear unit (ReLU) activation function, and 3D maxpooling layer with stride size of two. The kernel size of all convolutional layers is $3 \times 3 \times 3$. The number of filters of the first layer is sixty four and increases by the double as the depth of the network increases. The decoder path is the mirror of the encoder path by replacing the maxpooling layers with deconvolution layers. The application of deconvolution layers is another modification to the U-Net architecture to allow the network to learn the weight of the upsampling process. The last layer of the network is a fully connected layer that consists of three neurons that represent the number of classes with a multiclass sigmoid activation function. Figure 3.2 illustrates the baseline proposed CNN's architecture.

Another modification to the U-Net is the application of the dice coefficient loss as an objective function during the network training. The segmentation of brain MR images suffers

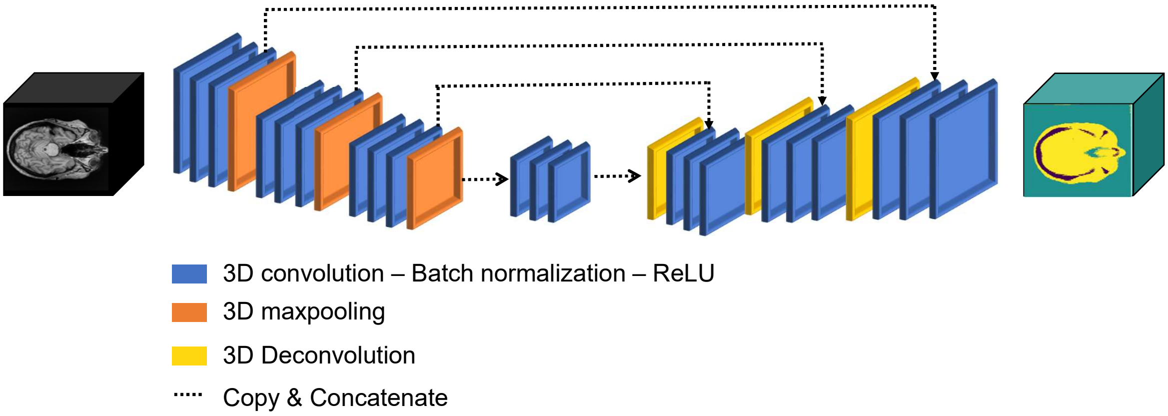


Figure 3.2: The convolutional encoder decoder network that follows the U-Net architecture.

from the problem of classes unbalancing which leads to a biased classification towards the majority classes. The cross-entropy loss function that is applied in U-Net is not suitable to address this problem. The dice coefficient loss for multi-class segmentation is defined as:

$$L_{dice} = \frac{1}{|C|} \sum_{c=1}^{C} \frac{2 \sum_i p_c^i g_c^i}{\sum_i (p_c^i + g_c^i)} \tag{3.1}$$

where p is the probability of each voxel to belong to class c, g is the ground truth of class c, and C is the number of classes.

3.3.2 Model Input

The input size of the designed CNN is 256 × 256 × 48 where 48 is the number of selected slices per patient. This huge number of pixels overwhelm the GPU memory and creates limitations when designing the network architecture and selecting the hyperparameters such as the number of filters per convolutional layer and the batch size. To overcome this situation, each volume is divided into overlapped patches to minimize the input size and increase the size of the training datasets. The employment of overlapped patches preserves the spatial contextual information for each volume. Each volume is divided into nine overlapped patches with the size of 128 × 128 × 48 where each patch covers half the area of the adjacent patch.

3.3.3 Model Training

The deep network is trained from scratch using thirty patients for training, ten patients for validation, and ten patients for testing. The network weights are initialised using Xavier/Glorot Uniform initialisation scheme which relies on the numbers of input and output neurons to automatically determine the scale of initialisation. The network is trained using Adam optimizer with an initial learning rate that is set to 0.0001 then reduces by a factor of 0.75 on plateau mode (i.e., when the training accuracy is not increasing for five continuous epochs). The momentum and the weight decay hyperparameters are set to 0.9 and 0.0005, respectively. The batch size is two and the training process stops after two hundred epochs or before that when the training accuracy stops decreasing for ten continuous epochs. All experiments are conducted on Tesla V100 GPU with 16 GB RAM. Keras with the backend of TensorFlow APIs are used to implement the proposed model.

3.3.4 Evaluation Metrics and Validation Methods

A Evaluation Metrics

The segmentation results are evaluated quantitatively using a set of metrics which have been widely applied in the literature to evaluate the segmentation tasks. The segmented MR image with the generated ground truth from CT images are used to calculate the following evaluation metrics for each tissue class *c*:

- *Precision (PRE)*

$$PRE_c = \frac{TP_c}{TP_c + FP_c} \tag{3.2}$$

- *Recall (REC)*

$$REC_c = \frac{TP_c}{TP_c + FN_c} \tag{3.3}$$

- *Dice similarity coefficient (DSC)*

$$DSC_c = \frac{2 \times TP_c}{(2 \times TP_c) + FP_c + FN_c} \tag{3.4}$$

- *Jaccard similarity coefficient (JSC)*

$$JSC_c = \frac{DSC_c}{2 - DSC_c} \tag{3.5}$$

where TP is true positive, FP is false positive, and FN is false negative. These metrics are calculated for all slices of the testing dataset and per patient.

3.3.5 Validation Methods

The proposed methods are compared with other segmentation studies that applied deep learning techniques using conventional MR sequences such as T1-w MR images for PET attenuation correction application. The existing methods which are listed below are re-implemented then trained and tested with the existing datasets:

- A deep CNN method which is proposed by [103] and follows the SegNet architecture is applied to identify and segment air, bone, and soft tissue in 3D brain MR images. The network is trained with T1-w MR images and evaluated using CT images. The same network architecture is re-implemented with some slight changes to cope with the size of the existing datasets using the same training parameters and hyperparameters.

- A deep CNN method which is proposed by [104] applies transfer learning technique by using a pre-trained model with T1-w MR images. The pretrained model is retrained using MR UTE images as input to the network. 3D fully connected conditional random field (CRF) is applied to the segmentation results as a post-processing and refinement technique.

3.3.6 Experimental Setup

A The Influence of Weight Initialization

The selection of the weight initialisation scheme of the convolutional layers is experimented by evaluating three different schemes: (1) Xavier uniform [265], (2) He normal [266], and (3) Random normal. The selection of the best initialisation scheme is based on calculating the DSC of each tissue class then applying the majority voting. While changing the initialisation strategy for each experiment, all other training settings and hyperparameters remain the same.

According to the results illustrated in Figure 3.3, the Xavier uniform initialisation scheme leads to the best segmentation performance for both soft tissue and air classes. However, the best DSC for bone class is He normal. The performance of the He normal is slightly better than the Xavier uniform strategy with the bone class. This slight difference is ignored since the performance of soft tissue and air classes is better with Xavier uniform. Hence, the Xavier uniform initialisation scheme is adopted for the network weights initialisation. In general, the bone segmentation performance is lower than soft tissue and air classes due to the difficulty of differentiating between bone and air classes as there is a similarity in the intensity values of bone and air classes. Another challenge is the variation of MR image intensity values from one patient to another [267].

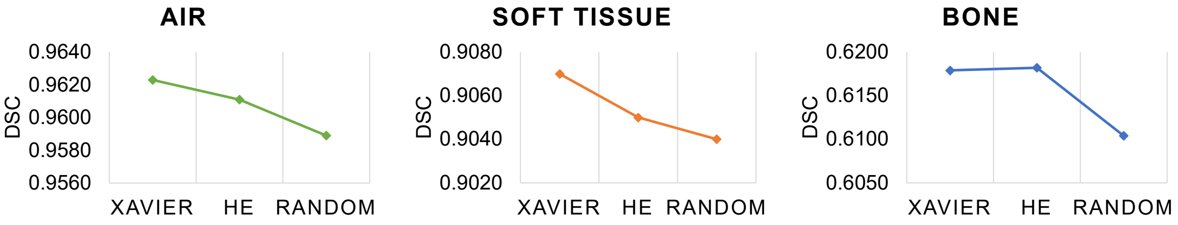


Figure 3.3: The DSC of the segmentation results using different weight initialisation schemes.

B The Influence of Pooling Layers

The pooling layer is an essential component of any CNN's architecture to reduce the dimensions of the feature maps which consequentially reduces the number of parameters and the amount of computing resources. The pooling layer also represents an abstract level of each feature map which produces invariant feature representation. On the other hand, the pooling layer leads to spatial information loss which effects the performance of the segmentation results. One way to overcome this issue is the use of skip connections to link the high-level features from high-resolution layers with low-level features from low-resolution layers to maintain some of the lost information.

There are mainly two types of pooling layers which are used in CNN's architecture: (1) Maxpooling layer and (2) Average pooling layer. The maxpooling layer selects the maximum value of each covered region of the features map while the average pooling layer represents each region by its average value. In other words, the two pooling layers is the maxpooling layer generates the most prominent features while the the average pooling represents the average of features. The selection of the best pooling type for the segmentation problem is tested by comparing the segmentation results of using two CNNs with different pooling layers by calculating the DSC of each class.

The illustrated results in Figure 3.4 show that maxpooling layer outperforms the average pooling layer for both the bone and air classes. Based on the majority voting, the maxpooling layer is adopted for the design of the pooling layers in the proposed CNN. The design of the CNN without using any pooling layer is also tested. However, the network consists of huge number of parameters which cannot be handled by the existing GPU memory.

C The Influence of Unpooling Layers

The unpooling layer is the opposite of the pooling layer which enlarges the features maps by performing the resizing operation. There are two types of unpooling operations which are commonly used in the CNN's architectures: (1) Upsampling layer and (2) Deconvolution layer.

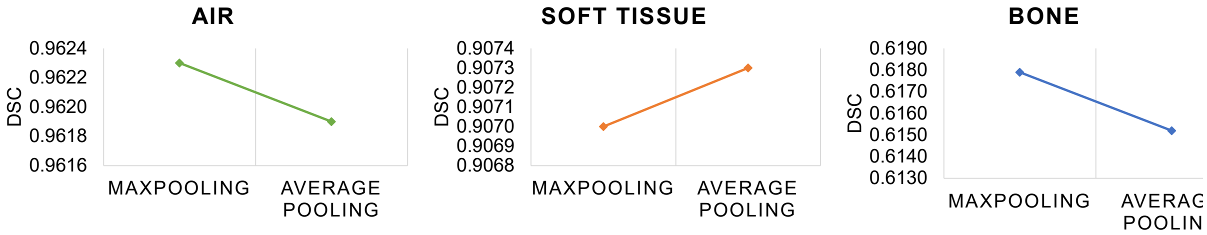


Figure 3.4: The DSC of the segmentation results using different pooling methods.

The upsampling layer resizes the feature map by means of interpolation. The upsampling layer is not trainable where it only uses the mathematical definition of the selected interpolation method. On the other side, the deconvolution layer is trainable where it tends to learn how to upsample during the training phase of the network. Hence, this makes the deconvolution layer data specific layer. Two experiments are conducted to investigate the best unpooling layer for the segmentation task. All other settings of the CNN's architecture are remained the same while changing the unpooling layers in the network design.

The demonstrated results in Figure 3.5 prove that the deconvolution layer is better than the upsampling layer by comparing the DSC of each CNN's architecture. Hence, the deconvolution is selected as the unpooling layer for the proposed network.

D The Influence of Activation Functions

The activation functions are the mathematical expression which determine the output of the neural network. Each neuron in the network is attached with an activation function which activates it at a certain condition or threshold. The activation functions also help to normalise the neuron output to a specific range. There are some activation functions which are associated with convolutional layers such as ReLU, leaky ReLU, and parametric ReLU. Additionally, there are other activation functions that are associated with the neurons at the fully connected layer such as softmax and sigmoid. The CNN mainly use the non-linear activation functions to allow the model to create sophisticated mapping and relationship between the inputs and outputs. This nonlinearity is essential for learning the complex data such as high dimensional images.

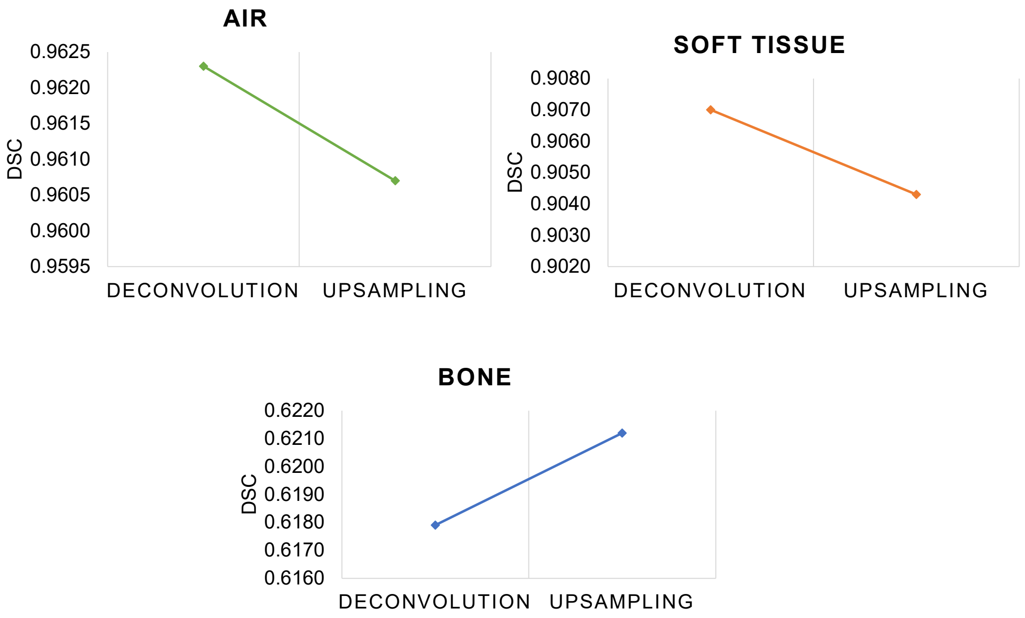


Figure 3.5: The DSC of the segmentation results using different unpooling methods.

In the next experiments, the activation functions of the fully connected layer which consists of neurons that represent the number of classes are explored to find the optimal function that achieves the best segmentation performance. The investigated activation functions are: (1) Sigmoid and (2) Softmax. All parameters of the network are fixed while only the activation function of the fully connected layer is changed.

The shown results in Figure 3.6 reveal that the sigmoid function with bound values between zero and one leads to the best segmentation performance by comparing the DSC of the three classes with the softmax function. The performance of the softmax function is very bad as the network has not been able to reach the convergence status during the training.

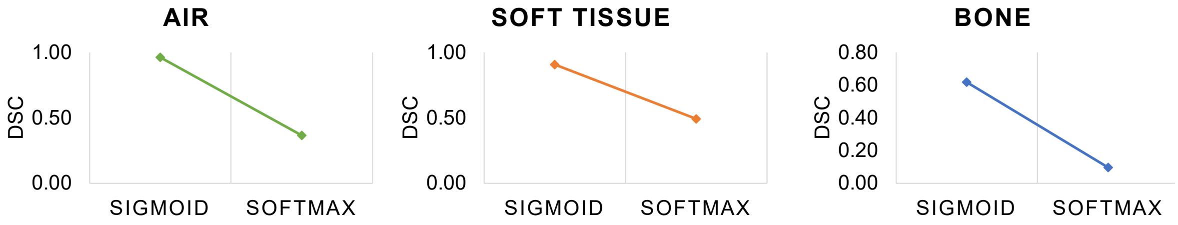


Figure 3.6: The DSC of the segmentation results using different activation functions.

3.4 Non-Subsampled Contourlet Transform Features with CNN based Features for MR Segmentation

3.4.1 Non-subsampled Contourlet transform

Contourlet is a multiscale and multidirectional transform which decomposes an image using the Laplacian pyramid and the directional filter bank (DFB). Due to the application of down-samplers and up-samplers in Contourlet filters, Contourlet transform is shift variant. The Contourlet addresses the drawback of wavelet which operates only in 2D without the capability to capture the multidirectional information. The Contourlet has the ability to encapsulate the directionality and anisotropy at various scales.

Non-subsampled Contourlet (NSCT) transform is a multiscale, multidirectional, and shift invariant transform which is proposed for the first time by Da Cunha el al. [268] to address essentially the image denoising and enhancement. This transform is a shift invariant version of the Contourlet transform with better frequency selectivity and regularity. This type of transforms include redundancy which enriches the details representation.

The Laplacian pyramid mechanism decomposes the image into one low pass sub-band and one band pass sub-band. The band pass represents the difference between the input image and the low pass image. Afterwards, the resultant low pass image is decomposed iteratively with the same mechanism at each stage. The band pass image is processed by the DFB to extract the directional details which generate the Contourlet coefficients. The DFB generates 2^k directional coefficients or sub-bands at k scale levels where k is a positive integer.

In this work, the MATLAB non-subsampled Contourlet toolbox [269] is utilised to generate the NSCT coefficients. The toolbox takes the 2D image, the number of directional decomposition levels (k), the pyramidal filter, and the directional filter as inputs. Then, it generates k+1 coefficients where the first coefficient is the low pass sub-band and the remaining ones represent different directional band pass sub-bands. The number of the directional filter decomposition levels is defined by vector K:

$$K = [1, 1, 2, 2] \tag{3.6}$$

Where the K levels generate C Contourlet coefficients which are calculated by:

$$C = \sum_{k=0}^{k} 2^k + 1 \tag{3.7}$$

The actual generated NSCT coefficients are thirteen sub-bands (twelve directional sub-bands + one low pass sub-band). The NSCT produces coefficients with the same dimensions as the input image. Figure 3.7 shows an example of some NSCT directional sub-bands at different scales in addition to the low pass sub-band for a single slice of a randomly selected patient from the brain dataset.

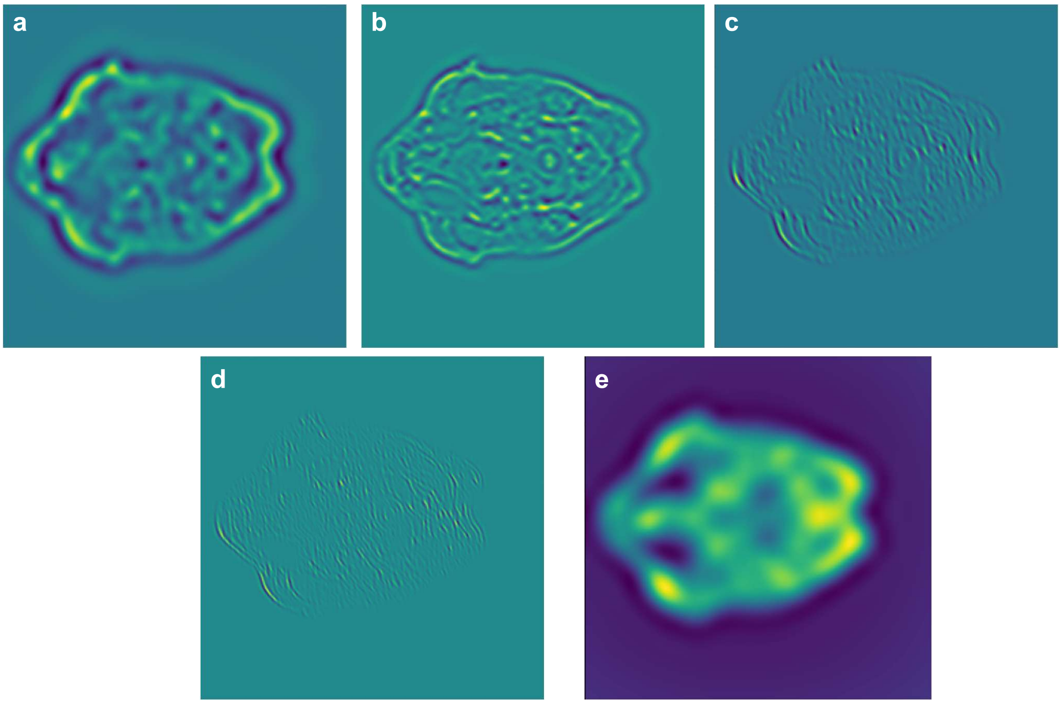


Figure 3.7: Four directional NSCT sub-bands from four scales (a-d) and one low pass sub-band (e) for a single slice of a randomly selected patient from the training dataset.

3.4.2 Non-Subsampled Contourlet Coefficients Selection

The application of the NSCT decomposition generates twelve redundant coefficients for each slice. The redundant directional information is reduced by selecting the coefficients that include more amount of information than others. The coefficients selection is performed by calculating the entropy score for each NSCT coefficient of each slice. The entropy score is a statistical measure of the information content and randomness to capture the texture characteristics of the image [270]. The entropy is calculated as:

$$Entropy = -\sum_{i} c_i log_2(c_i) \tag{3.8}$$

where c_i is the histogram count of i^{th} bin. The entropy values are calculated using the MATLAB toolbox [271].

Figure 3.8 illustrates the entropy values of the top ten selected NSCT directional sub-bands coefficients of one randomly selected slice (slice 44) with a minimum value of 3.028 and a maximum value of 3.772. Another ten randomly selected slices are illustrated in Figure 3.9 and

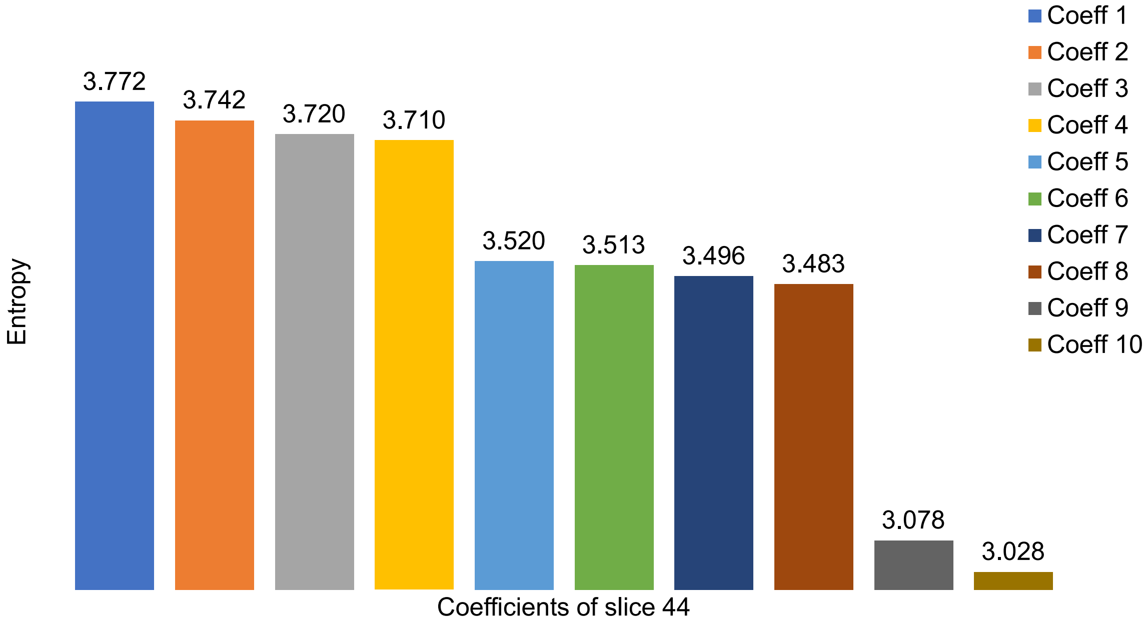


Figure 3.8: The entropy values of the top ten selected NSCT directional sub-bands coefficients of one randomly selected slice (slice 44) of patient 1.

all slices of one patient (forty eight slices) are shown in Figure 3.10. The figures demonstrate

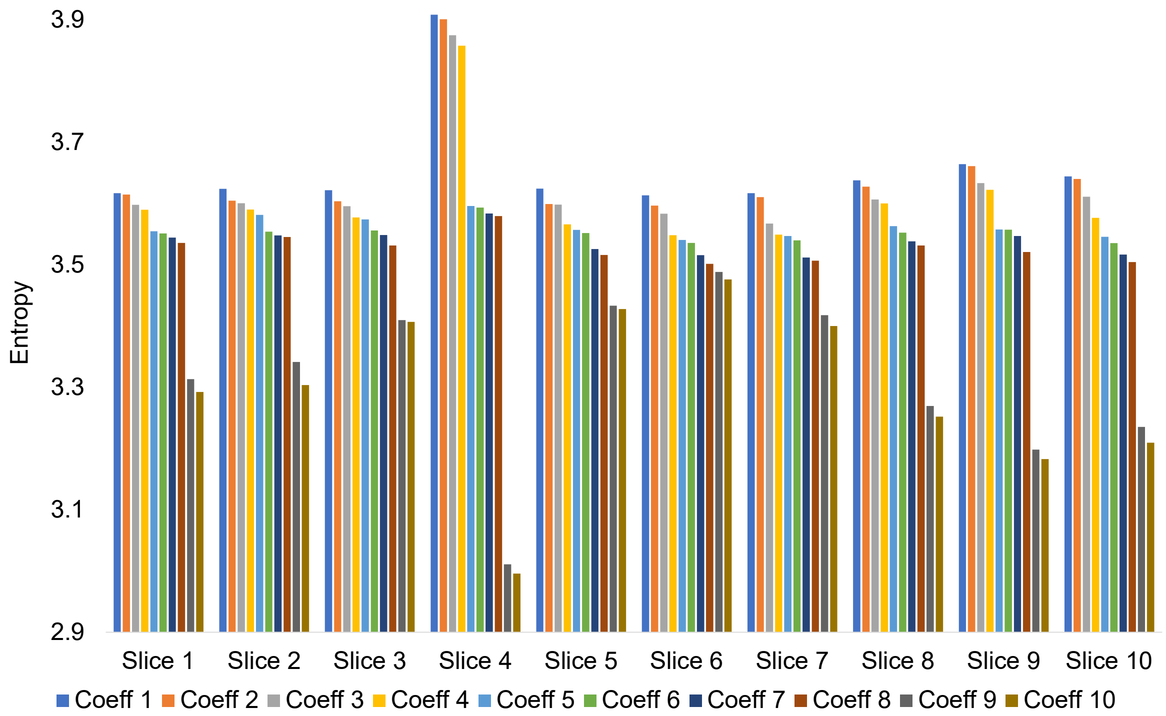


Figure 3.9: The entropy values of the top ten selected NSCT directional sub-bands coefficients of ten randomly selected slices of one patient.

that there are variations on the entropy values within and among each slice. The variation indicates there are some NSCT coefficients which are more useful than others. Each NSCT coefficient with higher entropy value indicates the inclusion of more information which can add more directional features to help improving the segmentation performance.

Out of twelve directional sub-bands, the top ten coefficients with the highest entropy values are selected as handcrafted features which will be combined with CNN's features. Figure 3.11 represents the histogram of the index of the ten selected NSCT coefficients for each slice of the fifty patients used in this study. It is observed from this figure that the directional sub-bands which are located between indices 9 and 12 represent the highest entropy values which are generated at the third and fourth decomposition levels of the NSCT. The directional sub-bands of the first and second decomposition levels are not revealing dominant features for most of the slices. It is worth noting that the more decomposition levels of NSCT coefficients,

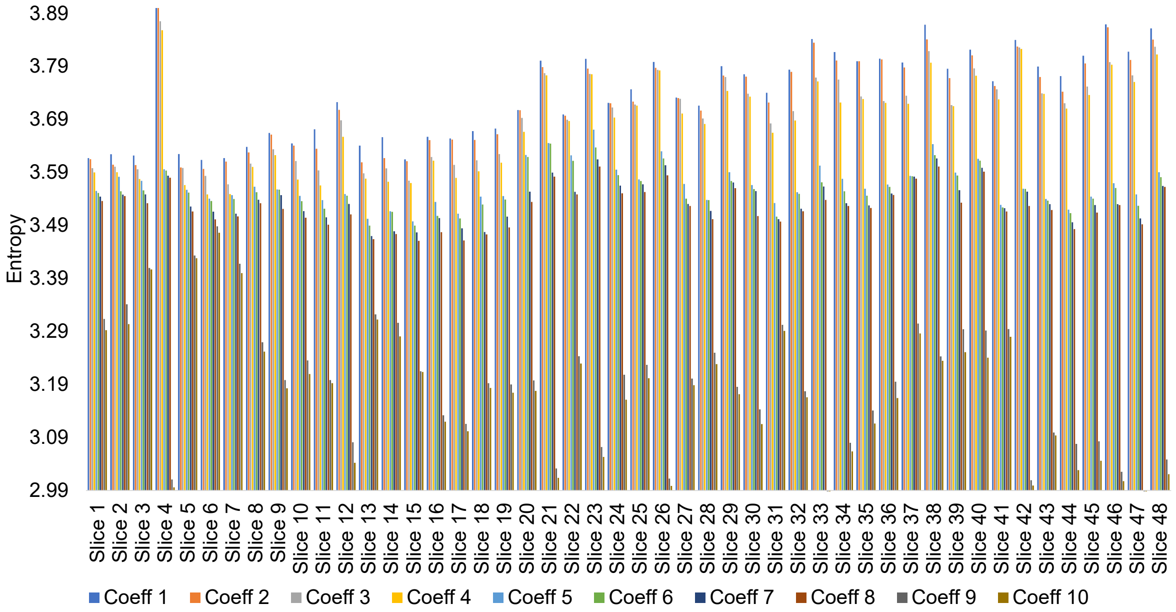


Figure 3.10: The entropy values of the top ten selected NSCT directional sub-bands coefficients of all slices of one patient (forty eight slices).

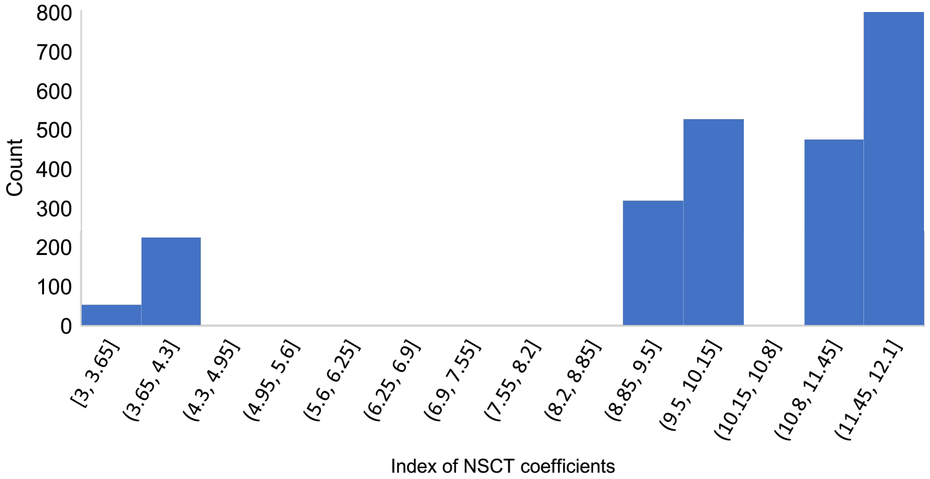


Figure 3.11: The histogram of the index of the ten selected NSCT coefficients for each slice of fifty patients.

the more useful directional features are produced.

3.4.3 Fusion of Non-Subsampled Contourlet and CNN based Features

The proposed method explores the addition of NSCT coefficients with CNN based features. The essential objective of adding the NSCT features is to introduce the directionality information and enrich and complement the CNN's features. The selected NSCT coefficients are concatenated to create an input image with size 256 × 256 × 48 ×10. As described in Section 3.3, the original image size is divided into nine overlapped patches with size 128 × 128 × 48 × 10 to firstly overcome the memory size limitation and secondly to preserve the contextual information.

The handcrafted NSCT features are concatenated with the CNN based features using three different mechanisms. The first mechanism which is illustrated in Figure 3.12 aims at concatenating the NSCT coefficients with the input MR images. The second mechanism concatenates the NSCT features at the end of the features extraction part of the CNN and just before the classifier layer as shown in Figure 3.13. The third mechanism adds another dimensionality to the input MR images by adding the low pass sub-band of the NSCT coefficient as a new channel. The input size of this network is 128 × 128 × 48 × 2. Moreover, at the end of this network, one or more directional sub-bands with the highest entropy value are concatenated with CNN based features to add some directionality information. This mechanism is depicted in Figure 3.14.

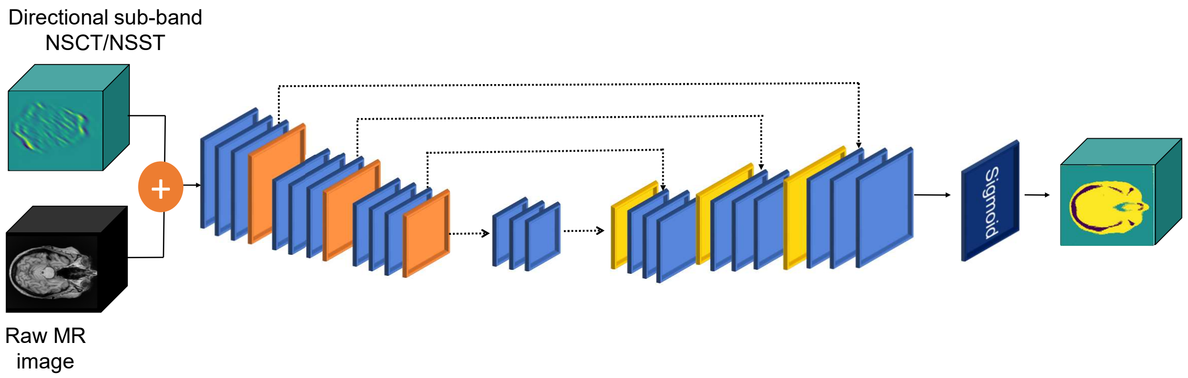


Figure 3.12: The network design of concatenating the NSCT directional coefficients with the input MR images.

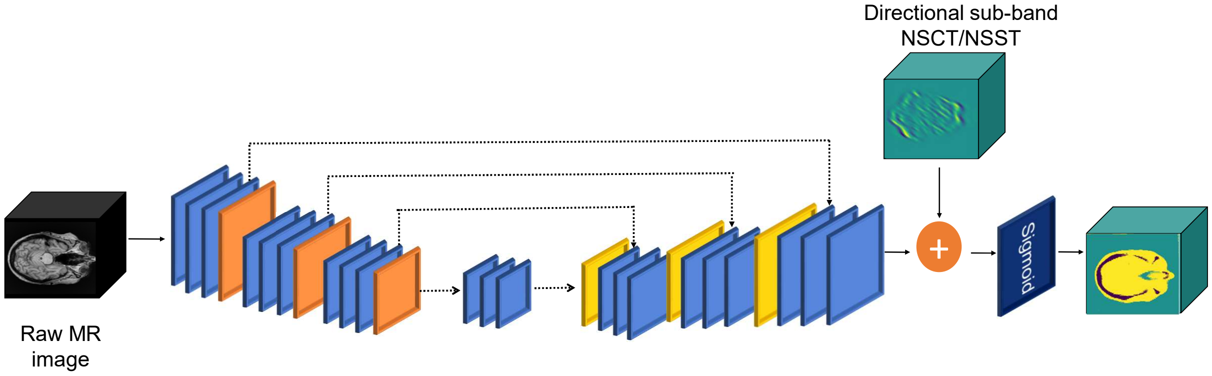


Figure 3.13: The network design of concatenating the NSCT directional coefficients with the CNN features at the end of the network.

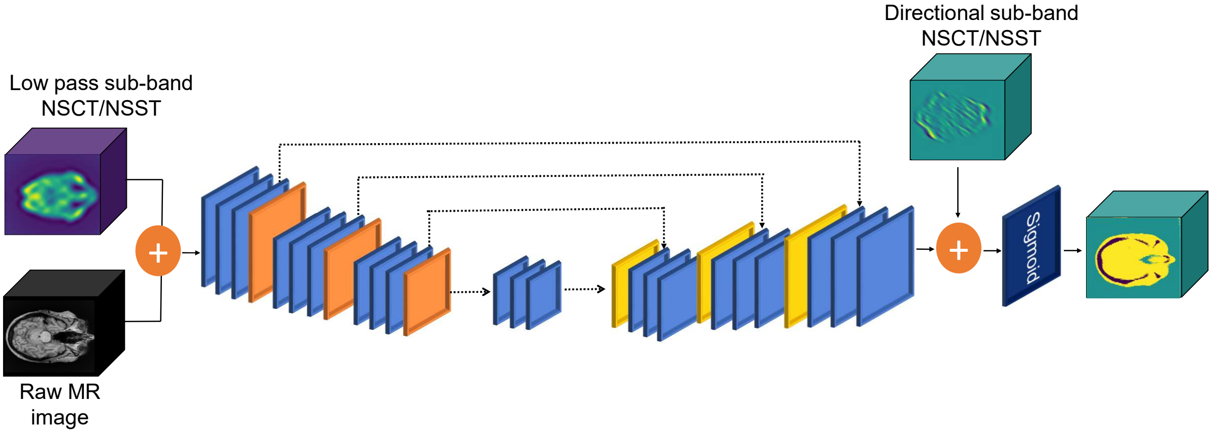


Figure 3.14: The network design of concatenating the NSCT low pass coefficients with the input MR images. The NSCT directional coefficients are also concatenated with the CNN features at the end of the network.

3.5 Non-Subsampled Shearlet Transform Features with CNN based Features for MR segmentation

3.5.1 Non-subsampled Shearlet Transform

Shearlet transform is a multivariate expansion of the wavelet transform in multidimensional space. Despite Contourlet transform utilises rotation and translation mechanisms, Shearlet utilises an additional shearing mechanism to provide directional selectivity and localisation.

The non-subsampled Shearlet (NSST) is the shift invariant version of the Shearlet transform with high capability of directional sensitivity. It consists of multiscale decomposition using non-subsampled Laplacian pyramid transform and multidirectional decomposition using Shear filter with translation invariance. The Shear filter overcomes the limitation of the number of directions which exists in non-subsampled Contourlet [272, 273].

In general, the NSST transform outperforms NSCT transform thanks to their well-localisation at different scales and directions which enables the fine detection of edges and boundaries. These advantages make the NSST a suitable tool to identify the different brain tissue classes [183]. In this work, the NSST coefficients are generated using the ShearLab 3D tool [274]. Each 2D slice is decomposed using the NSST up to four scales using different shearing levels defined by the vector L:

$$L = [1, 1, 2, 2] \tag{3.9}$$

Each level k generates S Shearlets coefficients which are calculated by:

$$S_k = 2 \times (2 \times 2^{L(k)} + 1) \tag{3.10}$$

The total number of generated NSST coefficients is fifty seven (fifty six directional high frequency sub-bands + one low frequency sub-band). The size of each sub-band is the same as the size of the input image. Figure 3.15 shows an example of some directional high frequency sub-bands from different scales and one low frequency sub-band for a single slice of one randomly selected patient from the training dataset.

3.5.2 Non-Subsampled Shearlet Coefficients Selection

The dimensions of the resultant sub-bands of each volume is 128 X 128 X 48 X 57. In order to reduce the dimensionality and the redundancy, the entropy for each NSST coefficient per slice is calculated to select only the coefficients that contain dominant information. The mathematical formula of the entropy is defined in Equation 3.8 in Section 3.4. Figure 3.16 illustrates the

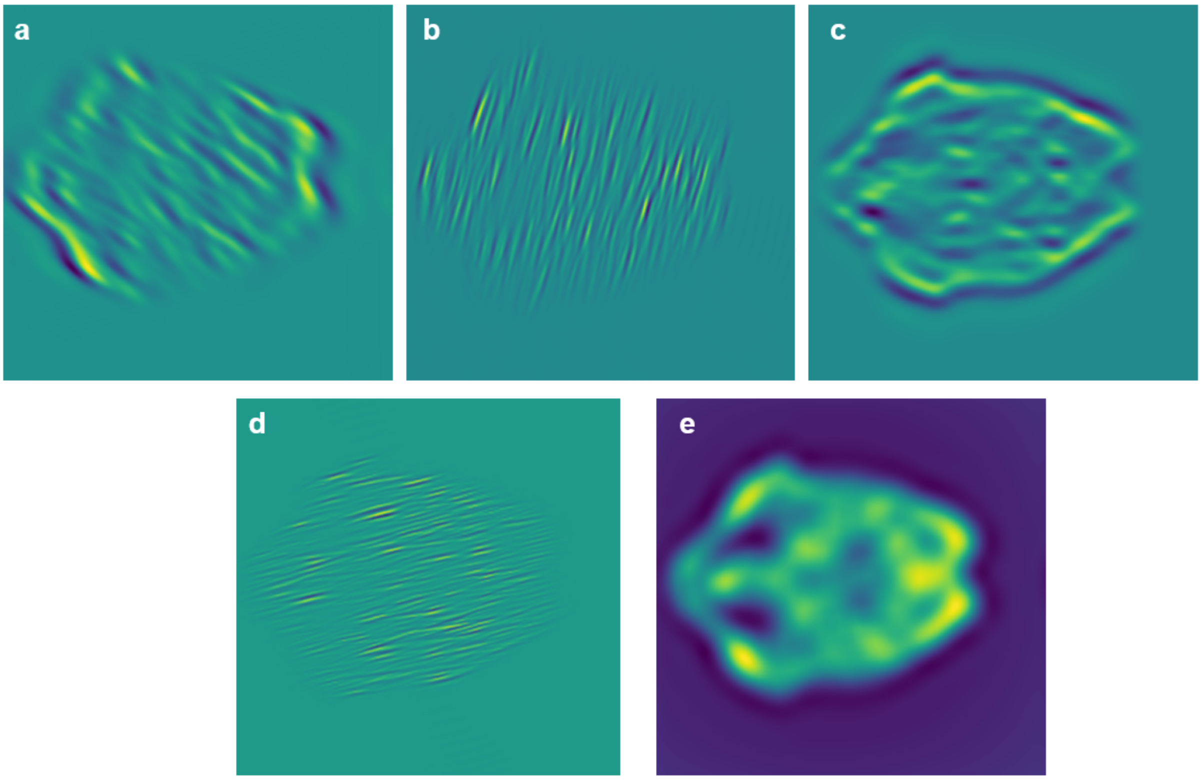


Figure 3.15: Four NSST directional sub-bands from four scales (a-d) and one low frequency sub-band (e) for a single slice of one patient.

entropy values of the top ten selected NSST directional sub-bands coefficients of one randomly selected slice of one patient with a minimum value of 4.04 and a maximum value of 4.61. Moreover, the entropy values of ten randomly selected slices are illustrated in Figure 3.17 and all slices of one volume are shown in Figure 3.18. Similar to NSCT, the figures demonstrate that there is variation on the entropy values within and among each slice but with higher entropy values.

Out of fifty six directional sub-bands, the top ten coefficients with the highest entropy values are selected as handcrafted features which will be combined with CNN's features. Figure 3.19 represents the histogram of the index of the ten selected NSST coefficients for each slice of the fifty patients used in this study. It is observed from this figure that the directional sub-bands which are located between indices 28 and 34 represent the highest entropy values which correspond to the third scale. It is also clear from the same figure that the directional sub-bands of the first level are not representing truly dominant features for most of the slices of each patient.

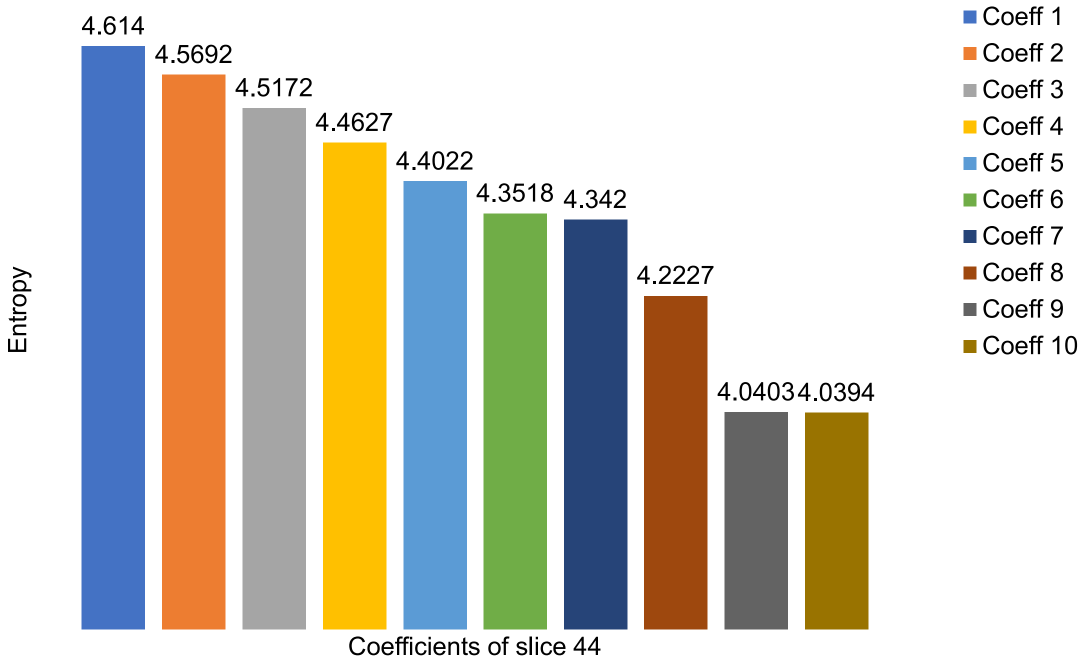


Figure 3.16: The entropy values of the top ten selected NSST directional sub-bands coefficients of one randomly selected slice (slice 44) of patient 1.

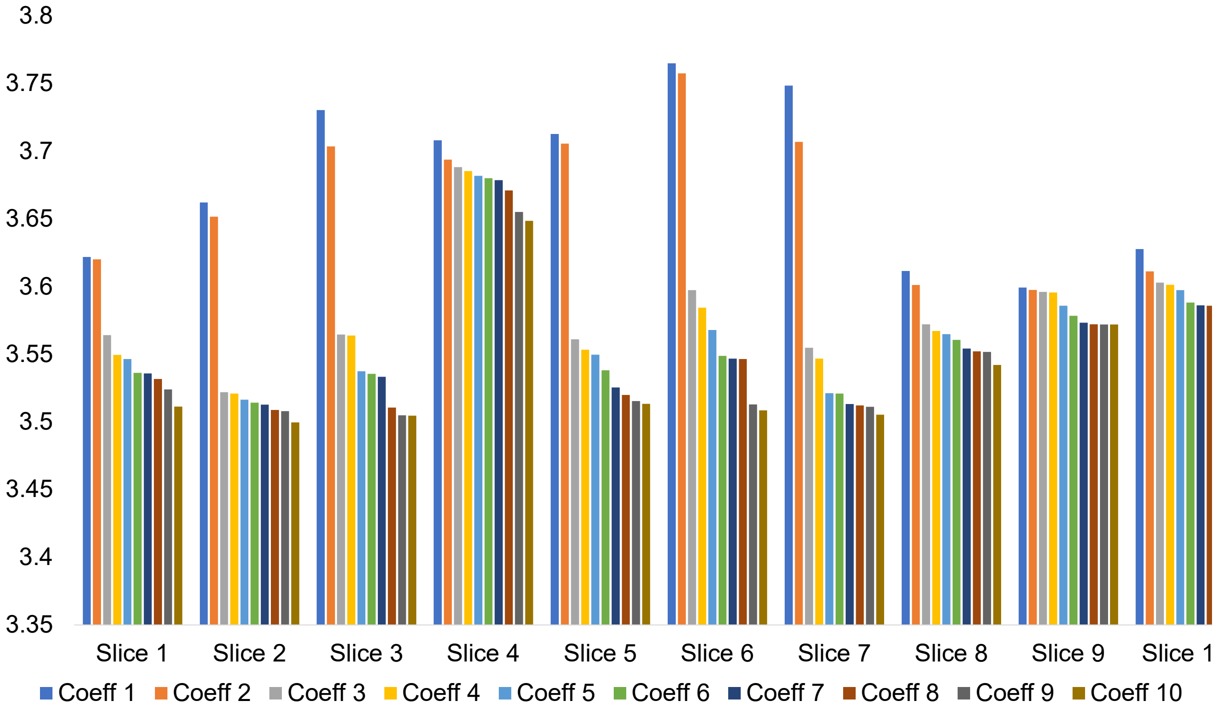


Figure 3.17: The entropy values of the top ten selected NSST directional sub-bands coefficients of ten randomly selected slices of one patient.

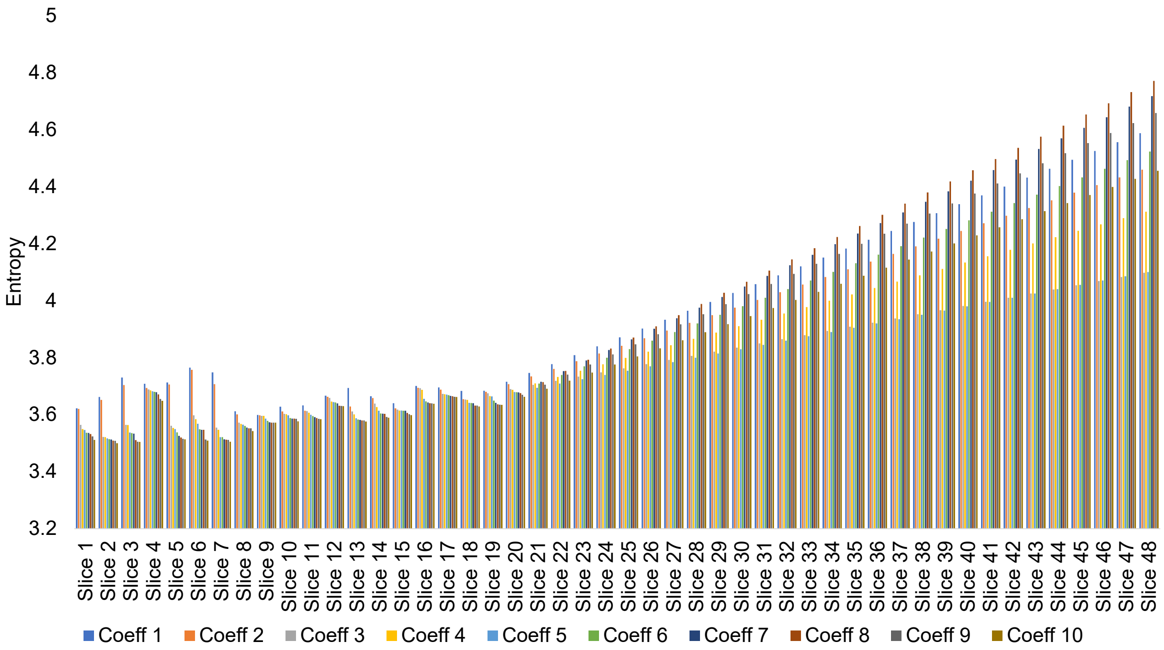


Figure 3.18: The entropy values of the top ten selected NSST directional sub-bands coefficients of all slices of one patient (forty eight slices).

3.5.3 Fusion of Non-Subsampled Shearlet and CNN based Features

The fusion mechanisms of the NSST features with CNN based features are similar to the applied mechanisms for NSCT features which are described in Section 3.4. The main objective is to explore if NSST coefficients are able to add more dominant features than NSCT coefficients.

3.6 Combining Non-Subsampled Contourlet and Non-Subsampled Shearlet Features with CNN based Features for MR Segmentation

The NSCT and NSST transforms generate different set of directional features using different decomposition filters which operate with different complexity manners. In order to achieve the full potential of the directional features as well as the low frequency sub-band, the NSCT and

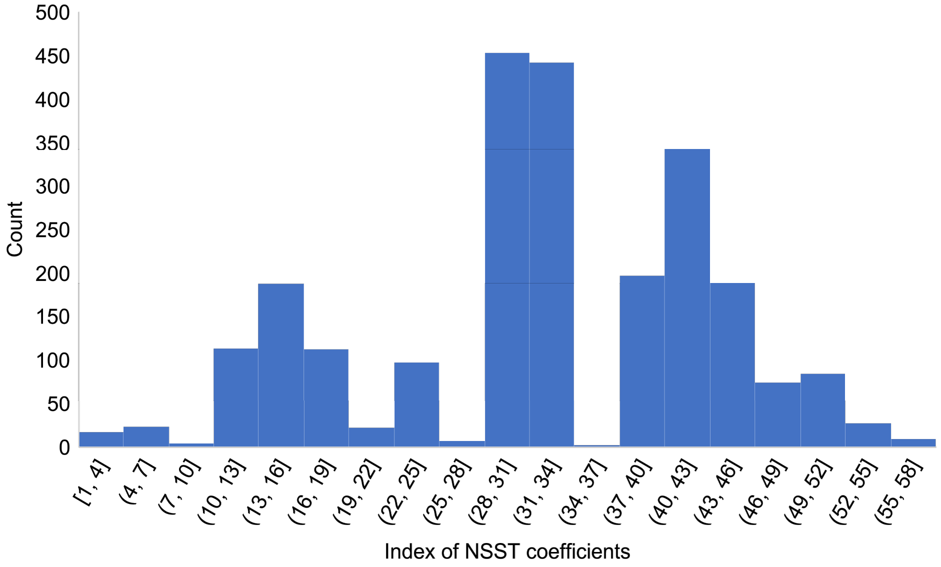


Figure 3.19: The histogram of the index of the ten selected NSST coefficients for each slice of fifty patients.

NSST coefficients are combined then fused with CNN's features.

The first fusion approach considers the directional features from the NSCT bandpass and the NSST high-pass sub-bands. The NSCT and NSST coefficients are concatenated with the CNN's features at the end of the features extractor part of the network and just before the last activation function (sigmoid) as depicted in Figure 3.20.

The second approach aims at exploring the benefit of concatenating the low frequency sub-band of NSCT or NSST with the MR image to increase the input dimensionality and add more data for training. Additionally, the directional sub-bands of both NSCT and NSST are concatenated at the end of the network with CNN's features as shown in Figure 3.21.

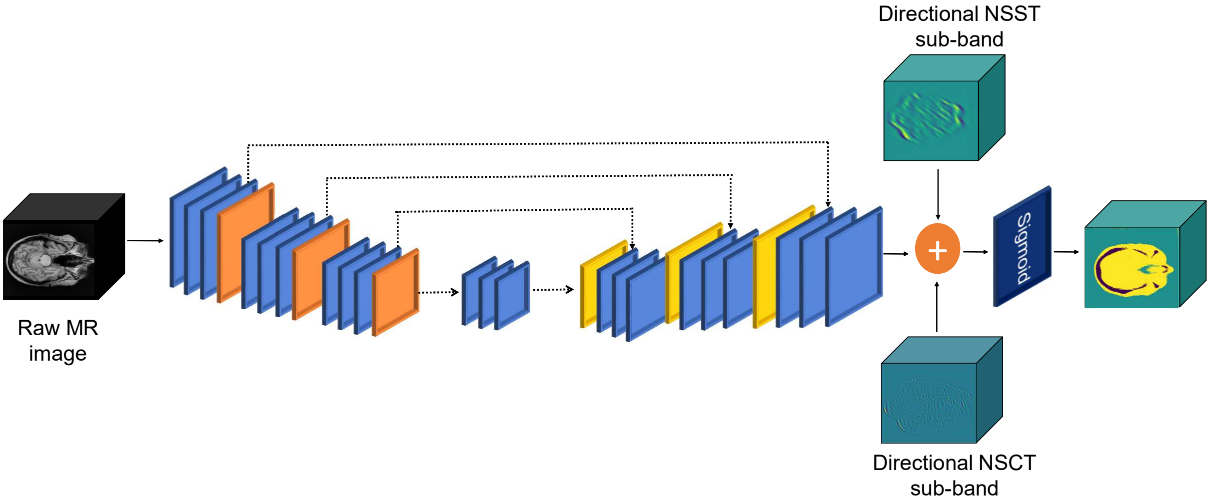


Figure 3.20: The network design of concatenating the NSCT and NSST directional coefficients with the CNN's features at the end of the network.

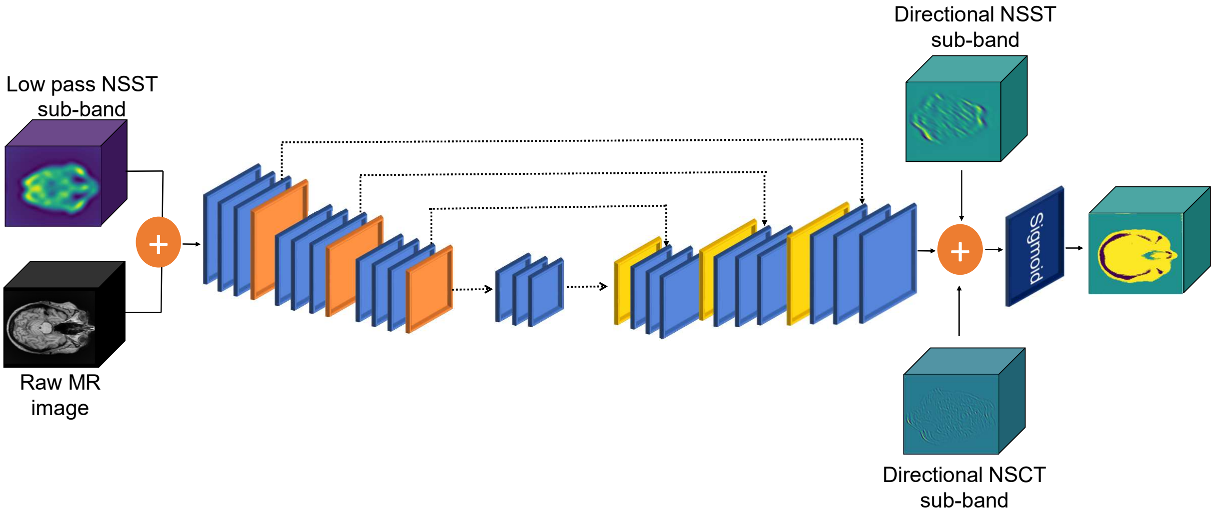


Figure 3.21: The network design of concatenating the NSCT low pass coefficient with the input MR images. The NSCT and NSST directional coefficients are also concatenated with the CNN's features at the end of the network.

3.7 Experimental Results for Brain Segmentation

3.7.1 Model Training and Evaluation

The proposed CNN's architecture which is introduced in Section 3.3 is adopted herein to study the impact of adding handcrafted features with CNN based features. The network is trained using the same parameters, hyperparameters and computing resources. The training dataset consists of thirty patients and each of the validation and testing datasets consist of ten patients.

The evaluation of the proposed method is conducted by calculating the same segmentation metrics which are introduced and defined in Section 3.3 The segmentation metrics for each brain tissue class are precision, recall, DSC, and JSC.

3.7.2 Evaluating the Effectiveness of Adding Handcrafted Directional NSCT Coefficients

The impact of adding the directional NSCT coefficients with CNN's features to perform MR image segmentation is firstly evaluated by concatenating only one NSCT coefficient per slice which has the highest entropy value. The NSCT coefficient is concatenated at the end of the network with the CNN's extracted features. Secondly, the top ten selected NSCT coefficients with the highest entropy values are concatenated with CNN's features in order to increase the directional features. The addition of ten NSCT coefficients is also evaluated by concatenating the directional features with the MR input images to pass through the convolutional layers and perform another features extraction process to maintain only the most useful and dominant features.

Tables 3.3, 3.4, and 3.5 show the results of the segmentation of the three brain classes using different mechanisms of concatenating the NSCT directional features with the CNN based features. The segmentation evaluation metrics show that the aggregation of only one directional NSCT coefficient with the input image has slight improvement in the segmentation results of the bone class. The increase of the dimensionality of the NSCT coefficients which are concatenated with the input image leads to include more directional features; Yet, it has not shown any improvement in the segmentation performance. The conservation of the raw NSCT coefficients is able to improve the segmentation of the three brain classes by concatenating these features with the resultant CNN's features at the end of the network. The reduction of the NSCT coefficients dimensionality by selecting only the coefficient with the highest entropy value decreases the segmentation performance. Overall, the concatenation of NSCT coefficients at the end of the features extraction process generates the best segmentation results even with the use of less coefficients. The inclusion of more NSCT coefficients and concatenating them with CNN's features without passing them through the convolutional layers is the winning approach to improve the segmentation performance of the three brain classes: air, soft tissue, and bone. By comparing the DSC of the different methods, it is noticeable that the performance of the bone class is improved with a significant amount while the segmentation performance of soft tissue and air classes is slightly improved.

Table 3.3: The evaluation metrics of the bone tissue segmentation using different mechanisms of concatenating the NSCT directional features with the CNN's features.

Bone tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.6890	0.5601	0.6179	0.4471
+ 1 directional NSCT (input)	0.6905	0.5637	0.6207	0.4500
+ 1 directional NSCT (end)	0.6821	0.5717	0.6221	0.4515
+ 10 directional NSCT (input)	**0.6910**	0.5539	0.6149	0.4439
+ 10 directional NSCT (end)	0.6883	**0.5962**	**0.6390**	**0.4695**

Table 3.4: The evaluation metrics of the soft tissue segmentation using different mechanisms of concatenating the NSCT directional features with the CNN's features.

Soft tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.8913	**0.9232**	0.9070	0.8298
+ 1 directional NSCT (input)	0.8936	0.9178	0.9056	0.8274
+ 1 directional NSCT (end)	0.8918	0.9194	0.9054	0.8271
+ 10 directional NSCT (input)	0.8906	0.9214	0.9058	0.8277
+ 10 directional NSCT (end)	**0.8981**	0.9171	**0.9075**	**0.8306**

Table 3.5: The evaluation metrics of the air tissue segmentation using different mechanisms of concatenating the NSCT directional features with the CNN features.

Air tissue				
Model	PRE	REC	DSC	JSC
Baseline	**0.9632**	0.9614	0.9623	0.9274
+ 1 directional NSCT (end)	0.9629	0.9614	0.9621	0.9270
+ 1 directional NSCT (input)	0.9600	**0.9642**	0.9621	0.9270
+ 10 directional NSCT (input)	0.9615	0.9622	0.9619	0.9265
+ 10 directional NSCT (end)	0.9615	0.9634	**0.9625**	**0.9276**

The visual segmentation results of the brain classes using the different proposed mechanisms of combining NSCT coefficients with CNN's features are depicted in Figure 3.22. The observations from the three randomly selected slices reveal that the baseline model which is represented in column c generates many false positive pixels of air class in the nasal cavities area as well as misclassified air and soft tissue pixels. Moreover, the quality of the segmentation of the bone in some challenging locations (e.g., nasal cavities area) is not good. However, the segmentation of the bone in other parts of the head is promising. The slices in column d show that the addition of only one NSCT coefficient with the input MR image removes some discontinuities in the bone class compared to the baseline model. Additionally, the segmentation of the air cavities pixels is enhanced. The increase of the dimensionality of the input by adding ten NSCT coefficients does not shown any improvement in the segmentation where the false positive pixels of bone in nasal cavities area are increased with the appearance of more bone discontinuities. The results of concatenating the NSCT coefficients at the end of the network with CNN's features by adding one or ten NSCT coefficients show the expansion of false positive pixels of air class in most of the selected slices. However, the slice in the first row (column e) shows that the segmentation of some challenging areas in the head is improved. Overall, the visual comparison indicates that the increase of the NSCT coefficients dimensionality which in turns increases the input size is an efficient way to improve the segmentation results. The capability of the model is improved when more directional features are added.

3.7.3 Evaluating the Effectiveness of adding Handcrafted Directional NSST Coefficients

The impact of combining the NSST directional coefficients is firstly evaluated by concatenating only one selected NSST coefficient per slice which has the highest entropy value with the input MR images. The second experiment tends to study the impact of adding the raw directional NSST coefficients without passing them through convolutional layers by concatenating the handcrafted features at the end of the network and just before the sigmoid activation function. The third experiment aims at increasing the number of the NSST coefficients to add more

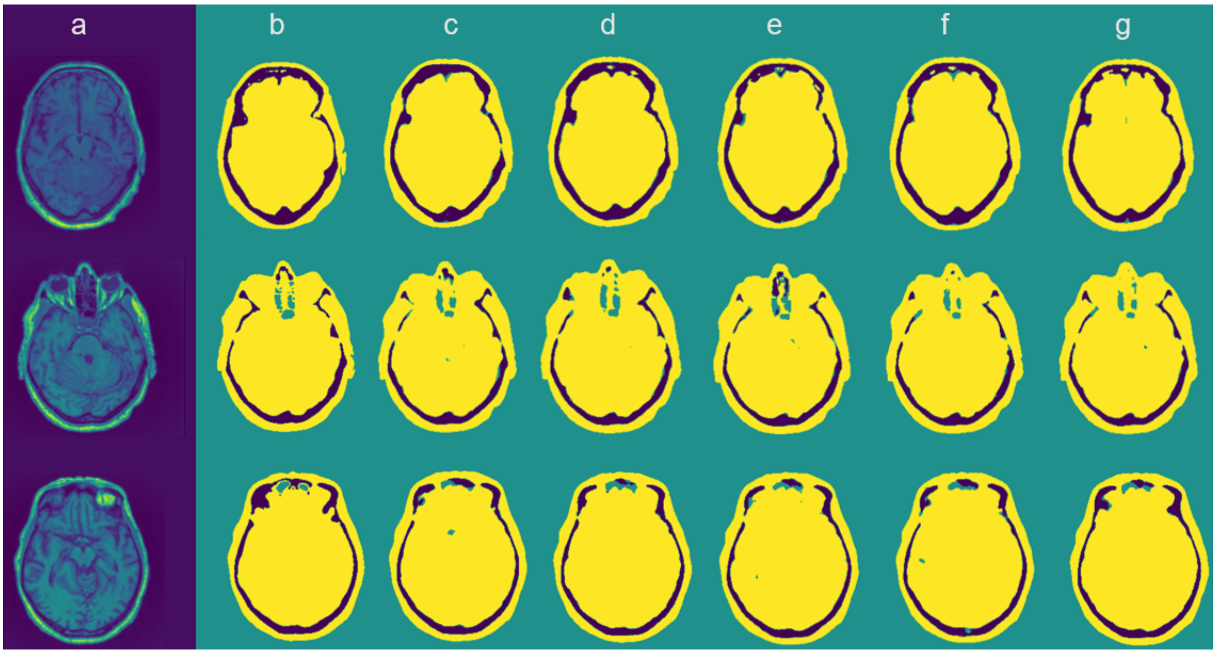


Figure 3.22: The segmentation results of some randomly selected slices from the testing datasets. (a) MR images, (b) the CT images which used as ground truth, (c) the segmentation results with the baseline model, (d) the segmentation results with the addition of one directional NSCT coefficient with the input, (e) the segmentation results with the addition of ten directional NSCT coefficients with the input, (f) the segmentation results with the addition of one directional NSCT coefficient with CNN's features at the end of the network, (g) the segmentation results with the addition of ten directional NSCT coefficients with the CNN's features at the end of the network.

directional information which may help to improve the segmentation results. Similar to the second experiment, the top ten selected NSST directional coefficients are concatenated with the resultant features of the CNN at the end of the network.

Tables 3.6, 3.7, and 3.8 show the quantitative results of the segmentation of the brain using different mechanisms of concatenating the NSST directional features with CNN based features. According to the results, the aggregation of the directional NSST coefficients is able to improve the segmentation results when it is concatenated at the end of the network. The concatenation of these handcrafted features with the generated CNN based features at the end of the network shows good improvements in terms of most of the evaluation metrics and specially for the bone and soft tissue classes. For the air class, only the recall metric is slightly improved. The increase of the dimensionality of the NSST coefficients which is experimented by selecting the top ten directional coefficients and concatenating them at the end of the network shows a slight improvement in the bone class. Thus, only one directional NSST coefficient is able to add sufficient directional information and enriches. However, the concatenation of one directional NSST features with the input MR image does not have any impact due to injection of few examples of a new type of data to the network. The increase of the dimensionality of the network's input by concatenating ten NSST directional coefficients with the input image increases the capability of the network to learn enough features from directional NSST coefficients and shows the best performance.

Table 3.6: The evaluation metrics of the bone tissue segmentation using different mechanisms of concatenating the NSST directional features with the CNN features.

Bone tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.6890	0.5601	0.6179	0.4471
+ 1 directional NSST (input)	0.6882	0.5582	0.6164	0.4456
+ 1 directional NSST (end)	0.7027	0.5690	0.6288	0.4586
+ 10 directional NSST (input)	**0.7211**	0.5595	**0.6301**	**0.4600**
+ 10 directional NSST (end)	0.6737	**0.5779**	0.6221	0.4515

Table 3.7: The evaluation metrics of the soft tissue segmentation using different mechanisms of concatenating the NSST directional features with the CNN features.

Soft tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.8913	0.9232	0.9070	0.8298
+ 1 directional NSST (input)	0.8908	0.9218	0.9060	0.8282
+ 1 directional NSST (end)	**0.8947**	0.9237	0.9090	0.8332
+ 10 directional NSST (input)	0.8934	**0.9286**	**0.9107**	**0.8360**
+ 10 directional NSST (end)	0.8933	0.9155	0.9043	0.8252

Table 3.8: The evaluation metrics of the air tissue segmentation using different mechanisms of concatenating the NSST directional features with the CNN features.

Air tissue				
Model	PRE	REC	DSC	JSC
Baseline	**0.9632**	0.9614	0.9623	0.9274
+ 1 directional NSST (input)	0.9628	0.9620	0.9624	0.9276
+ 1 directional NSST (end)	0.9614	0.9625	0.9620	0.9267
+ 10 directional NSST (input)	0.9631	**0.9636**	**0.9633**	**0.9293**
+ 10 directional NSST (end)	0.9615	0.9619	0.9617	0.9262

The visual segmentation results of some randomly selected slices of the testing datasets are illustrated in Figure 3.23. The slices shown in column d are the output of the model which concatenates only one NSST directional coefficient with the input MR image are not revealing any enhancement to the baseline model. For instance, the false negative pixels of the bone are decreased which in turns decreases the model sensitivity. On the other hand, the addition of more NSST coefficients enhances the segmentation of the challenging areas of the bone class. The slices shown in column f and g are the output of the models which concatenates one and ten NSST directional coefficients with the CNN's features at the end of the network. The results show some misclassified soft tissue pixels as air and other misclassified bone pixels as soft tissue. Moreover, the discontinuities in the bone class are increased. By comparing the slices in column f and g, it is clear that the segmentation of the model which uses ten NSST directional coefficients is better than the model that takes only one NSST coefficient especially in the nasal cavity area. The concatenation of the NSST coefficients with the CNN's features as raw features without passing them through the convolutional layers shows its ability to generate more accurate results by using only one NSST coefficients rather then ten NSST coefficients. One directional coefficient per slice is able to add the useful directional features to improve the segmentation results.

3.7.4 Evaluating the Effectiveness of Adding Handcrafted Low Pass NSCT Coefficients

The decomposition of NSCT generates only one low pass (LP) sub-band which represents the approximate information which can be used as an additional data input to the MR images. The inclusion of low pass sub-band increases the size of the training datasets and tends to add more visual representations. Moreover, the directional features of one or more NSCT coefficients are concatenated with CNN's features at the end of the network.

Tables 3.9, 3.10, and 3.11 show the evaluation of the brain segmentation using both low pass and band pass NSCT sub-bands concatenated with the CNN's features. The results reveal that the addition of the low pass NSCT sub-band with only one directional NSCT coefficient at the end of the network is able to attain better segmentation performance for the

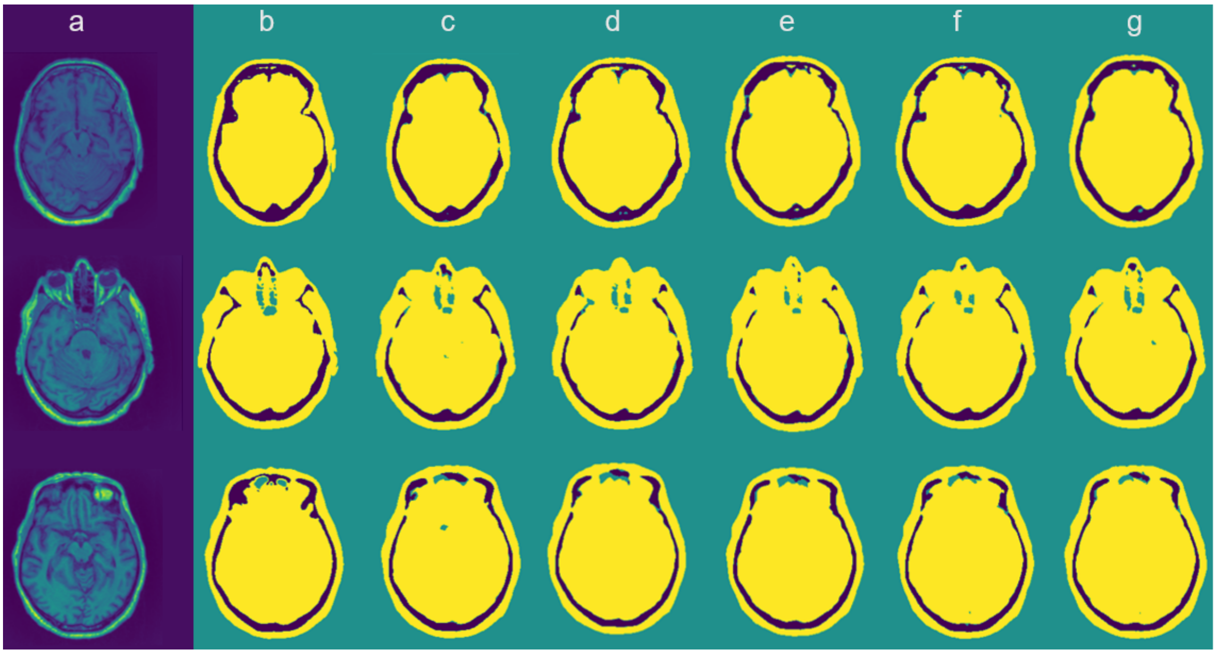


Figure 3.23: The segmentation results of some randomly selected slices from the testing datasets. (a) the MR images, (b) the CT images which used as ground truth, (c) the segmentation results with the baseline model, (d) the segmentation results with the addition of one directional NSST coefficient with the input, (e) the segmentation results with the addition of ten directional NSST coefficients with the input, (f) the segmentation results with the addition of one directional NSST coefficient with CNN's features at the end of the network, (g) the segmentation results with the addition of ten directional NSST coefficients with the CNN's features at the end of the network.

bone class.

Table 3.9: The evaluation metrics of the bone tissue segmentation using different mechanisms of concatenating the NSCT low frequency coefficient with CNN based features.

Bone tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.6890	0.5601	0.6179	0.4471
+ 1 LP NSCT (input) + 1 directional NSCT (end)	0.6931	**0.5744**	**0.6282**	**0.4580**
+ 1 LP NSCT (input) + 10 directional NSCT (end)	**0.7072**	0.5446	0.6153	0.4444

Table 3.10: The evaluation metrics of the soft tissue segmentation using different mechanisms of concatenating the NSCT low frequency coefficient with CNN based features.

Soft tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.8913	0.9232	0.9070	0.8298
+ 1 LP NSCT (input) + 1 directional NSCT (end)	**0.8950**	0.9194	0.9070	0.8299
+ 1 LP NSCT (input) + 10 directional NSCT (end)	0.8893	**0.9274**	**0.9079**	**0.8314**

The visual segmentation results of the brain using the proposed mechanisms are depicted

Table 3.11: The evaluation metrics of the air tissue segmentation using different mechanisms of concatenating the NSCT low frequency coefficient with CNN based features.

Air tissue				
Model	PRE	REC	DSC	JSC
Baseline	**0.9632**	0.9614	0.9623	0.9274
+ 1 LP NSCT (input) + 1 directional NSCT (end)	0.9615	**0.9639**	**0.9627**	**0.9281**
+ 1 LP NSCT (input) + 10 directional NSCT (end)	**0.9632**	0.9619	0.9625	0.9278

in Figure 3.24. The figure shows that the addition of the low pass sub-band of NSCT is improving slightly the segmentation performance of the bone class in the nasal cavities area. However, the model that concatenates the low pass sub-band with the input along with ten directional NSCT coefficients with CNN's features is not improving the segmentation results of the bone class and is even performing worst than the baseline model.

3.7.5 Evaluating the Effectiveness of Adding Handcrafted Low Pass NSST Coefficients

Similar to the NSCT decomposition, the NSST also generates only one low pass sub-band. Both NSCT and NSCT apply the Laplacian pyramid filter which generates an identical low pass sub-band. Hence, the low pass sub-band of NSCT or NSST is concatenated to the input MR image while the NSST directional coefficients are concatenated at the end of the network to study the effect of combining low pass sub-band and NSST high pass sub-bands.

Tables 3.12, 3.13, and 3.14 show the evaluation measurements of the the three classes segmentation by concatenating the low pass sub-band of the NSST with the input image and the high pass sub-bands of NSST coefficients with CNN's features. The results reveal that the addition of the low pass sub-band with the inclusion of only one directional NSST sub-band is enhancing the segmentation performance. The second model that takes ten NSST coefficient while the low pass is concatenated with the input shows better performance compared to the model that only concatenates ten NSST coefficients without low pass sub-band inclusion. The inclusion of low pass coefficient of NSST is able to reduce the false positive of bone class and the false negative of air class. Collectively, the model that takes less data and produces better results

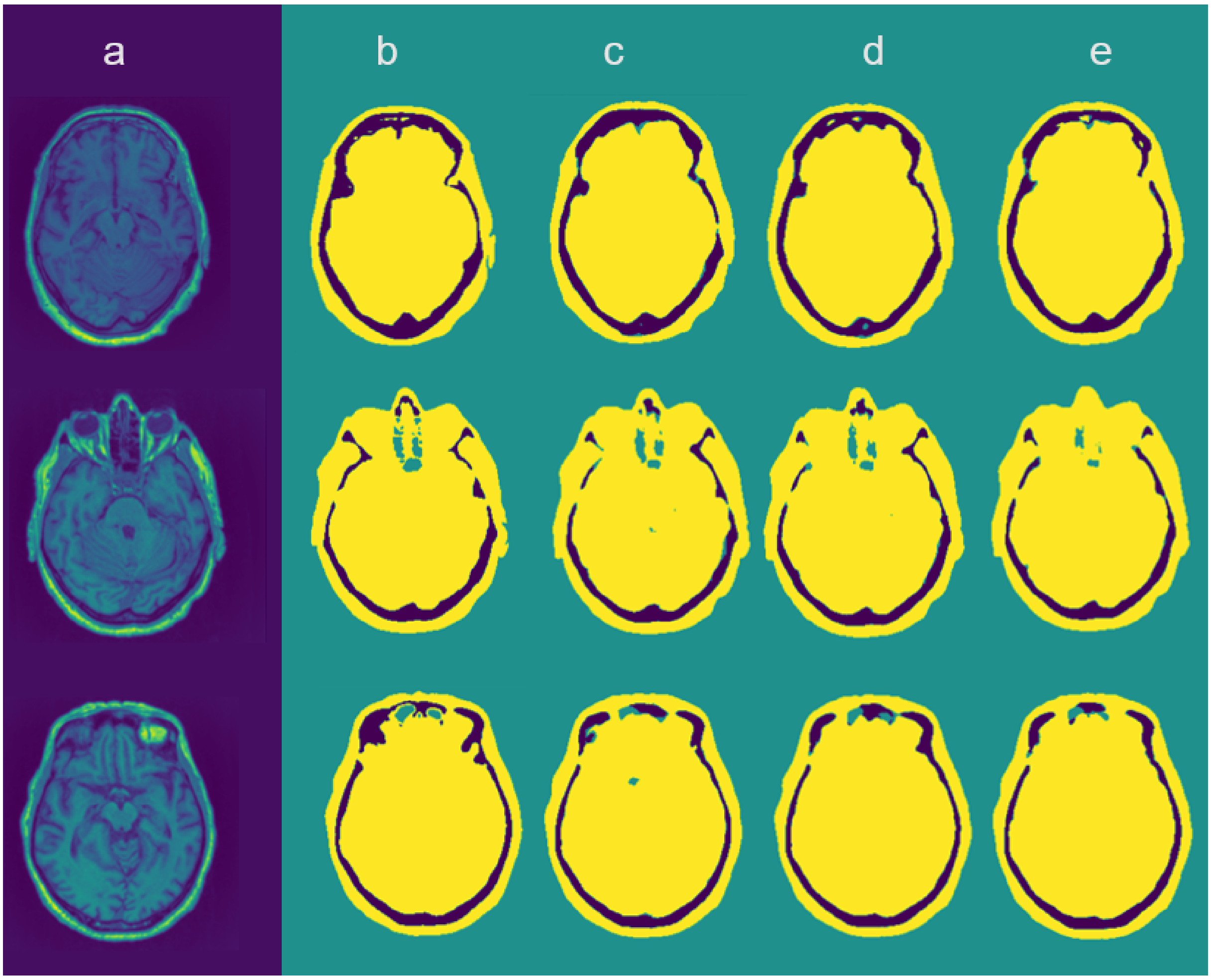


Figure 3.24: The segmentation results of some randomly selected slices from the testing datasets. (a) the input MR images, (b) the CT images which used as ground truth, (c) the segmentation results with the baseline model, (d) the segmentation results with the addition of the NSCT low pass sub-band (input) and one directional NSCT coefficient (end), (e) the segmentation results with the addition the NSCT low pass sub-band (input) and ten directional NSCT coefficients (end).

is the model of choice. Hence, the inclusion of low pass sub-band with one NSST directional coefficient is the winning approach. Moreover, the results prove that the model that uses the low pass sub-band concatenated with the input MR image is able to extract useful features from two inputs with different nature and resolutions. Therefore, this model is robust and reliable to be used for other transfer learning applications.

The visual segmentation results of the brain using the proposed mechanisms are depicted in Figure 3.25. The slices show that the inclusion of low pass sub-band of NSST coefficient is improving the bone segmentation significantly compared to the baseline model segmentation. It is also able to segment the bone in the challenging areas of the head such as the nasal cavity

Table 3.12: The evaluation metrics of the bone segmentation using different mechanisms of concatenating the NSST directional features with the CNN features.

Bone tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.6890	0.5601	0.6179	0.4471
+1 LP NSST (input) + 1 directional NSST (end)	0.7029	**0.5730**	**0.6313**	**0.4612**
+1 LP NSST (input) +10 directional NSST (end)	**0.7031**	0.5703	0.6298	0.4596

Table 3.13: The evaluation metrics of the soft tissue segmentation using different mechanisms of concatenating the NSST directional features with the convolutional network based features.

Soft tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.8913	0.9232	0.9070	0.8298
+1 LP NSST (input) + 1 directional NSST (end)	0.8940	**0.9235**	**0.9085**	**0.9085**
+1 LP NSST (input) +10 directional NSST (end)	**0.8946**	0.9221	0.9081	0.8317

Table 3.14: The evaluation metrics of the air segmentation using different mechanisms of concatenating the NSST directional features with the convolutional network based features.

Air tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.9632	0.9614	0.9623	0.9274
+1 LP NSST (input) + 1 directional NSST (end)	**0.9633**	0.9633	**0.9633**	**0.9297**
+1 LP NSST (input) +10 directional NSST (end)	0.9613	**0.9635**	0.9624	0.9276

and reduces the discontinuities. However, there are some misclassified soft tissue pixels as air. The concatenation of the low pass sub-band with the input and adding one NSST directional coefficient with the CNN's features at the end of the network is producing more accurate segmentation than using ten NSST directional coefficients since the bone discontinuities are reduced as well as the false positive of the air and bone classes. Moreover, the model with ten directional NSST coefficients is able to segment the challenging bone area in the head.

3.7.6 Evaluating the Effectiveness of Adding Combined Handcrafted NSCT and NSST Coefficients

Aiming at achieving the full potential of directional handcrafted features and adding more useful features to the CNN model, the coefficients of both NSCT and NSST are combined with

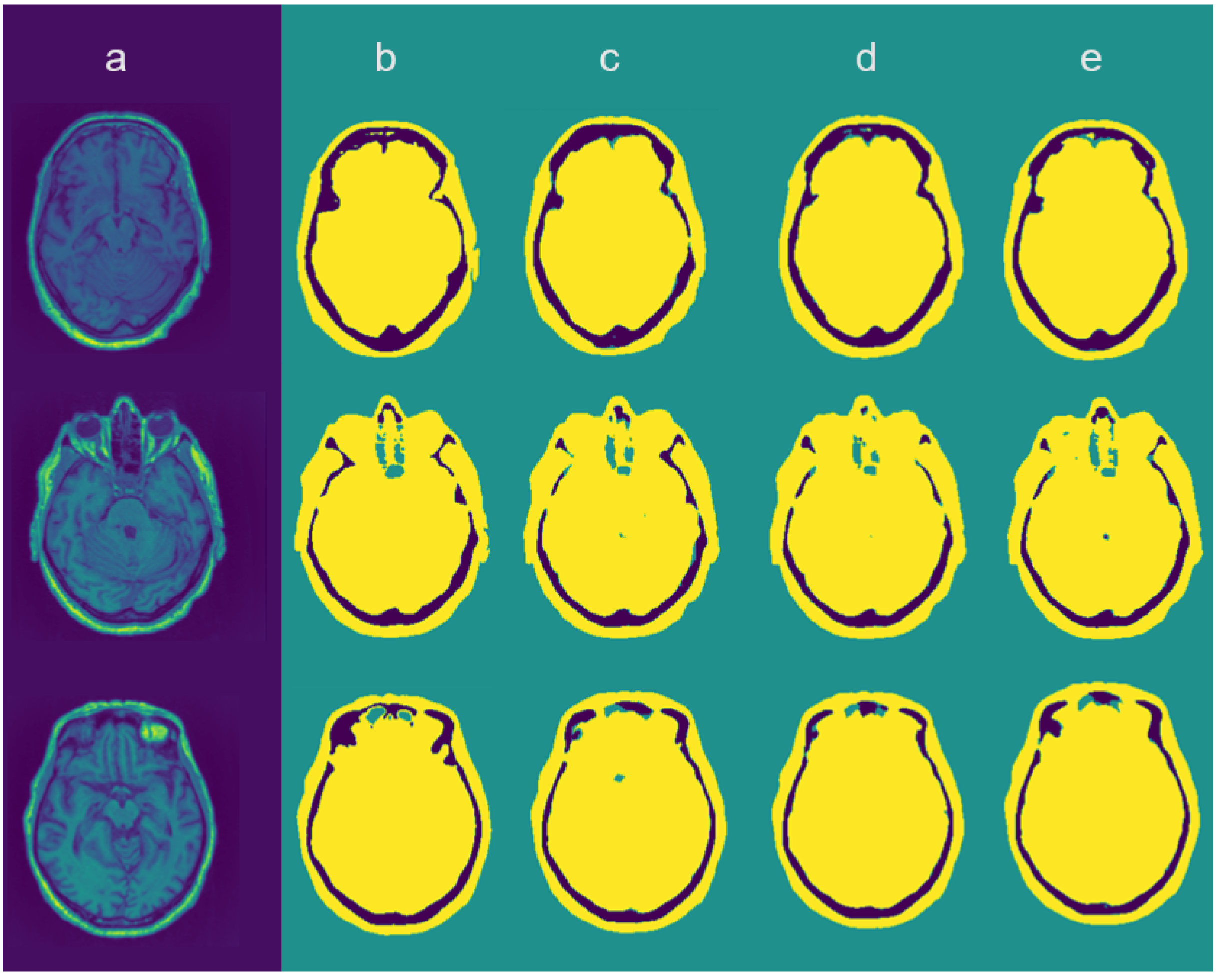


Figure 3.25: The segmentation results of some randomly selected slices from the testing datasets. (a) the MR images, (b) the CT images which used as ground truth, (c) the segmentation results with the baseline model, (d) the segmentation results with the addition of the NSST low pass sub-band (input) and one directional NSST coefficient (end), (e) the segmentation results with the addition the NSST low pass sub-band (input) and ten directional NSST coefficients (end).

different designs. Firstly, only one directional coefficient with the highest entropy value of each transform is selected and concatenated with the CNN's features. Secondly, the dimensionality of the directional features is increased by selecting the top ten directional coefficients of both transforms. Additionally, the low pass sub-band of NSST is concatenated with the input MR image where the directional coefficients of both NSCT and NSST are combined with CNN's features at the end of the network.

The quantitative evaluation of the conducted experiments per brain tissue class are shown in Tables 3.15, 3.16, and 3.17. The results show that the combination of only one directional coefficient of each transform is not improving the segmentation accuracy significantly.

For instance, it is slightly improving the soft tissue and air classes as well as reducing the false positive of the bone class. Moreover, the results reveal that increasing the dimensionality of the directional coefficients of both transforms is improving the segmentation results by comparing the DSC of the three classes. This combination is outperforming all other combination mechanisms and shows that adding more directional features using different multiresolution transforms is adding value to the segmentation model. However, the results of adding the low pass sub-band with the input image while concatenating one or many directional NSCT/NSST coefficients with CNN's features are not revealing any significant segmentation improvement.

Table 3.15: The evaluation metrics of the bone segmentation results using different mechanisms of concatenating the NSCT and NSST coefficients with the CNN features.

Bone tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.6890	0.5601	0.6179	0.4471
+ 1 directional NSCT and 1 directional NSST (end)	**0.7218**	0.5485	0.6233	0.4528
+ 10 directional NSCT and 10 directional NSST (end)	0.7113	**0.5831**	**0.6416**	**0.4723**
+ 1 NSST LP (input) and 1 directional NSCT and NSST (end)	0.6941	0.5751	0.6290	0.4588
+ 1 NSST LP (input) and 10 directional NSCT and NSST (end)	0.6980	0.5645	0.6242	0.4537

Table 3.16: The evaluation metrics of the soft tissue segmentation results using different mechanisms of concatenating the NSCT and NSST coefficients with the CNN features.

Soft tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.8913	0.9232	0.9070	0.8298
+ 1 directional NSCT and 1 directional NSST (end)	0.8905	**0.9296**	0.9096	0.8342
+ 10 directional NSCT and 10 directional NSST (end)	**0.8943**	0.9273	**0.9105**	**0.8357**
+ 1 NSST LP (input) and 1 directional NSCT and NSST (end)	0.8940	0.9219	0.9078	0.8311
+ 1 NSST LP (input) and 10 directional NSCT and NSST (end)	0.8920	0.9241	0.9078	0.8311

Tables 3.18, 3.19, 3.20, and Figure 3.26 show the evaluation metrics per patient of the testing datasets by using the best performing model which combines ten directional coefficients of NSCT and ten directional coefficients of NSST sub-bands with CNN's features at the end of the network.

Table 3.17: The evaluation metrics of the air segmentation results using different mechanisms of concatenating the NSCT and NSST coefficients with the CNN features.

Air tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.9632	0.9614	0.9623	0.9274
+ 1 directional NSCT and 1 directional NSST (end)	0.9632	0.9624	0.9628	0.9283
+ 10 directional NSCT and 10 directional NSST (end)	**0.9647**	0.9616	**0.9631**	**0.9289**
+ 1 NSST LP (input) and 1 directional NSCT and NSST (end)	0.9629	**0.9625**	0.9627	0.9281
+ 1 NSST LP (input) and 10 directional NSCT and NSST (end)	0.9636	0.9622	0.9629	0.9285

Table 3.18: The evaluation metrics of the bone class segmentation of each patient in the testing dataset using the combination of ten directional NSCT and ten directional NSST coefficients with CNN's features at the end of the network.

Bone tissue				
Patient	PRE	REC	DSC	JSC
Patient 1	0.8129	0.5180	0.6328	0.4628
Patient 2	0.8478	0.6804	0.7549	0.6063
Patient 3	0.6361	0.3906	0.4840	0.3192
Patient 4	0.6816	0.3604	0.4715	0.3084
Patient 5	0.8975	0.7856	0.8378	0.7209
Patient 6	0.4683	0.5309	0.4976	0.3312
Patient 7	0.1960	0.0909	0.1242	0.0662
Patient 8	0.5477	0.7918	0.6475	0.4788
Patient 9	0.9123	0.8238	0.8658	0.7633
Patient 10	0.6400	0.4761	0.5460	0.3755

Table 3.19: The evaluation metrics of the soft tissue class segmentation of each patient in the testing dataset using the combination of ten directional NSCT and ten directional NSST coefficients with CNN's features at the end of the network.

Soft tissue				
Patient	PRE	REC	DSC	JSC
Patient 1	0.9197	0.9586	0.9387	0.8845
Patient 2	0.9269	0.9424	0.9346	0.8773
Patient 3	0.9079	0.9255	0.9166	0.8460
Patient 4	0.8585	0.9373	0.8962	0.8119
Patient 5	0.9416	0.9574	0.9494	0.9037
Patient 6	0.9200	0.8743	0.8965	0.8125
Patient 7	0.9232	0.9185	0.9208	0.8532
Patient 8	0.9588	0.8796	0.9175	0.8476
Patient 9	0.9581	0.9727	0.9653	0.9329
Patient 10	0.6749	0.9037	0.7727	0.6296

Table 3.20: The evaluation metrics of the air class segmentation of each patient in the testing dataset using the combination of ten directional NSCT and ten directional NSST coefficients with CNN's features at the end of the network.

Air tissue				
Patient	PRE	REC	DSC	JSC
Patient 1	0.9716	0.9896	0.9805	0.9617
Patient 2	0.9583	0.9862	0.9720	0.9456
Patient 3	0.9507	0.9707	0.9606	0.9242
Patient 4	0.9731	0.9805	0.9768	0.9547
Patient 5	0.9814	0.9914	0.9864	0.9731
Patient 6	0.9683	0.9909	0.9794	0.9597
Patient 7	0.9543	0.9896	0.9716	0.9448
Patient 8	0.9815	0.9922	0.9868	0.9740
Patient 9	0.9848	0.9889	0.9869	0.9741
Patient 10	0.9136	0.7516	0.8247	0.7017

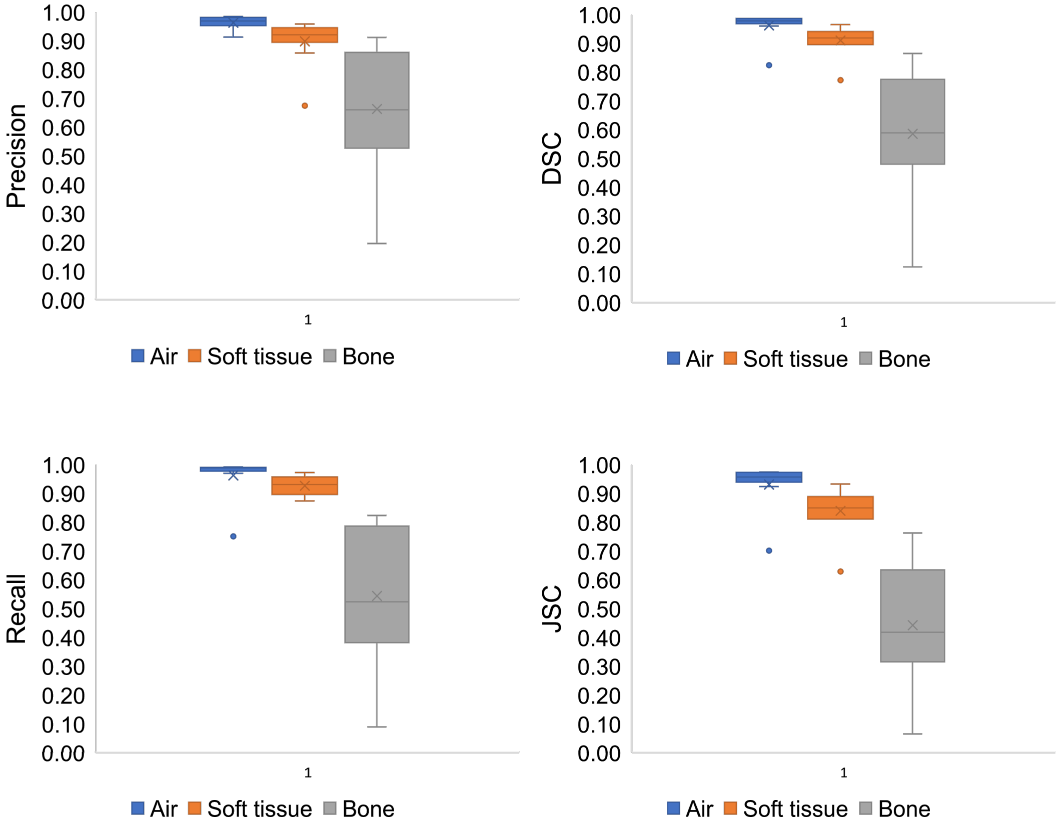


Figure 3.26: The box plots of the evaluation metrics of the three tissue classes of the testing datasets using the method which uses the combined ten directional NSCT and NSST coefficients concatenated with CNN features at the end of the network.

The segmentation results of the proposed mechanisms are depicted in Figure 3.27 where the slices that are located at columns d and e show that the combination of two different directional handcrafted features is improving the segmentation performance compared to the baseline model in column c. On the other hand, the inclusion of one directional coefficient of each transform helps to identify the soft tissue class in very challenging areas in the head. It also decreases the false positive of air class which increases the precision of air class inside the head contour. The increase of the number of directional features reduces the false positive of air and bone classes although some misclassified bone pixels as soft tissue pixels are still exist. Column f depicts that the addition of low pass sub-band with the directional coefficients is improving the bone class segmentation while decreasing the sensitivity and the precision of soft tissue class. The results of the dimensionality increase of the directional coefficients which are shown in column g indicate the appearance of more discontinuities in bone class which do not appear in the slice of column f. Moreover, the false negative pixels of soft tissue class are increased. Overall, the performance of adding the low pass sub-band with one or ten directional coefficients is not adding a substantial improvement in the segmentation performance.

3.7.7 Validation with Other Methods

The proposed methods are compared with other segmentation studies that applied CNN to segment T1-w MR images for PET attenuation correction [103, 104] . These methods are re-implemented then trained and tested using the existing datasets to segment the brain into three tissue classes: air, soft tissue, and bone.

The evaluation metrics of the three brain classes segmentation of the three proposed methods are compared with other studies as depicted in Tables 3.21, 3.22, and 3.23. The results show that the proposed methods which introduce the concatenation of different features from NSCT and NSST coefficients with CNN's features using different aggregation mechanisms are enhancing the segmentation performance of the three brain classes. The baseline model which represents the CNN without the addition of any handcrafted features is outperforming the other proposed methods in the literature. Liu et al. [103] have applied the Segnet architecture to perform the MR images segmentation for PET attenuation correction and Jang et al. [104] have applied the same pretrained model as [103] using UTE MR images by applying the transfer

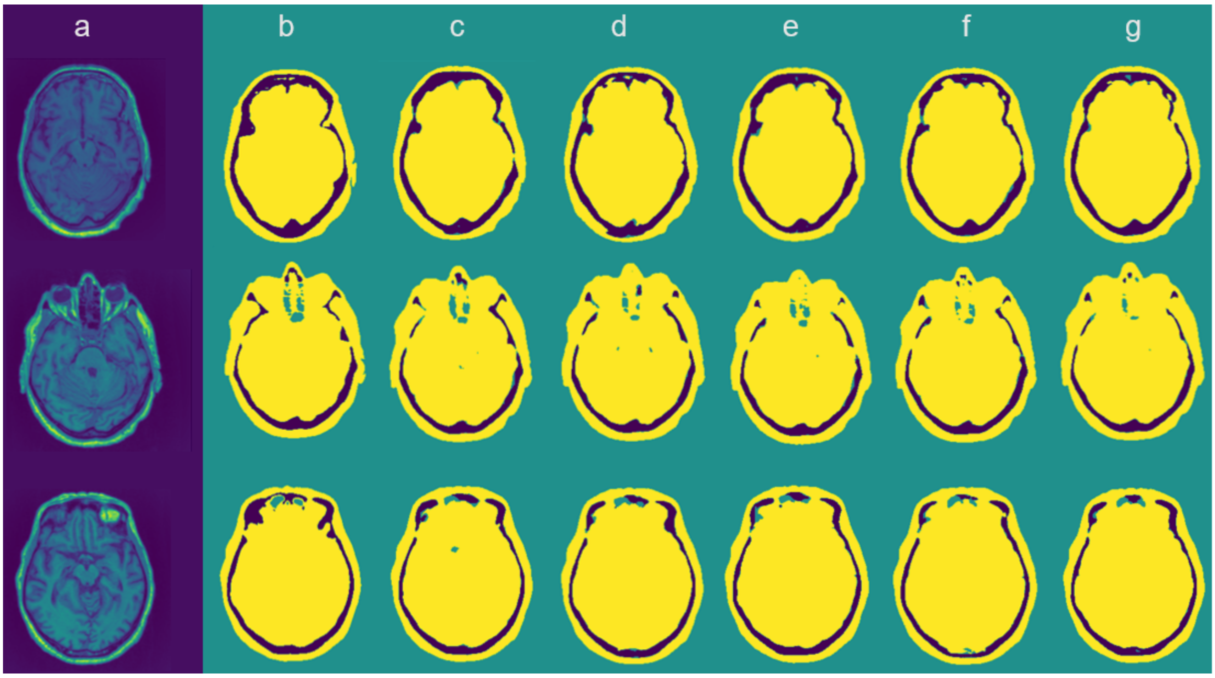


Figure 3.27: The segmentation results of some randomly selected slices from the testing datasets. (a) the MR images, (b) the CT images which used as ground truth, (c) the segmentation results with the baseline model, (d) the segmentation results by combining one directional NSCT and NSST coefficients, (e) the segmentation results by combining ten directional NSCT and NSST coefficients, (f) the segmentation results by adding the NSST low pass sub-band with the input image and combining one directional NSCT and NSST coefficients with CNN's features at the end of the network, (g) the segmentation results by adding the NSST low pass sub-band with the input image and combining ten directional NSCT and NSST coefficients with CNN's features at the end of the network.

learning technique in addition to conditional random field as post processing technique. The model which combines both directional features of NSCT and NSST shows its superiority in performing the segmentation of the bone and soft tissue classes.

Table 3.21: The comparison of the segmentation results of the bone class using the optimal methods with other techniques in the literature.

Bone tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.6890	0.5601	0.6179	0.4471
Baseline + 1 LP NSST (input) + 1 directional NSST (end)	0.7029	0.5730	0.6313	0.4612
Baseline +10 directional NSCT (end)	0.6883	**0.5962**	0.6390	0.4695
Baseline + 10 directional NSCT and NSST (end)	**0.7113**	0.5831	**0.6416**	**0.4723**
Segnet[103]	0.6278	0.3649	0.4616	0.3000
Segnet + CRF [104]	0.6278	0.3649	0.4616	0.3000

The segmentation results of some randomly selected slices are illustrated in Figure

Table 3.22: The comparison of the segmentation results of the soft tissue class using the optimal methods with other techniques in the literature.

Soft tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.8913	0.9232	0.9070	0.8298
Baseline + 1 LP NSST (input) + 1 directional NSST (end)	0.8940	0.9235	0.9085	**0.9085**
Baseline +10 directional NSCT (end)	**0.8981**	0.9171	0.9075	0.8306
Baseline + 10 directional NSCT and NSST (end)	0.8943	**0.9273**	**0.9105**	0.8357
Segnet[103]	0.8696	0.8810	0.8753	0.7783
Segnet + CRF [104]	0.8696	0.8810	0.8753	0.7783

Table 3.23: The comparison of the segmentation results of the air class using the optimal methods with other techniques in the literature.

Air tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.9632	0.9614	0.9623	0.9274
Baseline + 1 LP NSST (input) + 1 directional NSST (end)	0.9633	0.9633	**0.9633**	**0.9297**
Baseline +10 directional NSCT (end)	0.9615	**0.9634**	0.9625	0.9276
Baseline + 10 directional NSCT and NSST (end)	**0.9647**	0.9616	0.9631	0.9289
Segnet[103]	0.9083	0.9510	0.9291	0.8677
Segnet + CRF [104]	0.9083	0.9510	0.9291	0.8677

3.28. The visual observations indicate that the models that include ten directional NSCT and NSST coefficients for segmenting the brain are outperforming the other proposed methods in the literature. The approach of [103] which is shown in column g is not able to identify the bone and air classes in challenging areas. Moreover, the precision and the sensitivity of the bone and air classes are considerably lower than the proposed baseline model. The approach of [104] is also not showing any improvement in the segmentation results since the conditional random field which iis used for post processing is not enhancing the segmentation performance.

The main differences of the three proposed methods (shown in columns d - f) are as follows. The addition of the NSCT coefficients with CNN's features is able to distinguish between the soft tissue and bone classes although more cases of false positive of air class have appeared in different locations. Besides that, the concatenation of NSST coefficients increases the sensitivity of the soft tissue class and improves the performance of the bone segmentation. Moreover, the combination of both transforms coefficients with the CNN's features increases the competency of the model to segment the air class. However, there are more discontinuities that appear in the bone class and more misclassified bone pixels as soft tissue.

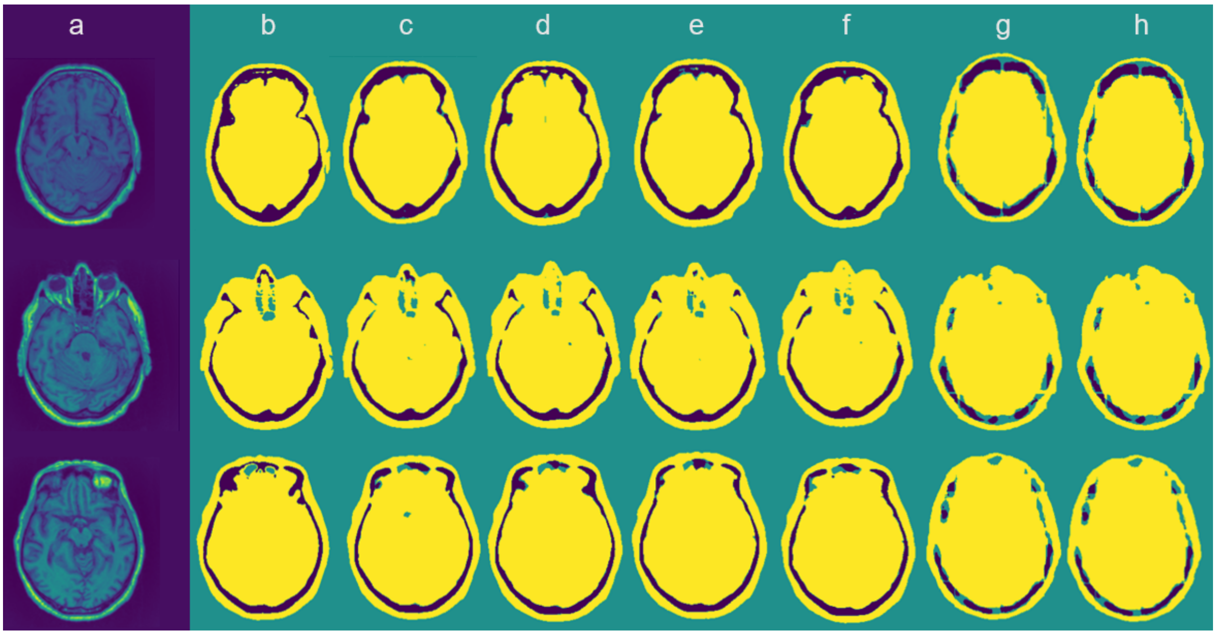


Figure 3.28: The segmentation results of some randomly selected slices from the testing datasets. (a) the input MR images, (b) the CT images which used as ground truth, (c) the segmentation results of the baseline model, (d) the segmentation results with ten directional NSCT coefficients concatenated with CNN's features (end), (e) the segmentation results with the addition of the NSST low pass sub-band (input) and one directional NSST coefficients (end), (f) the segmentation results by combining ten directional NSCT and NSST coefficients (end), (g) the segmentation results of the model proposed by [103], (h) the segmentation results of the model proposed by [104].

3.8 Summary and Limitations

The design of the CNN's architecture has been studied in this chapter by exploring the different network components. Several experiments have been conducted to select the best design of the network which improves the brain segmentation into three tissue classes: air, bone, and soft tissue.

This chapter has also proposed the aggregation of handcrafted features using multiresolution analysis techniques such as NSCT and NSST transforms with the CNN based features to enhance the brain T1-w MR images segmentation. The handcrafted features consist of the low pass sub-band which is concatenated with the input MR image and the directional coefficients of different levels which are concatenated with CNN's features at the end of the network.

A detailed objective evaluation using the brain MR images of ten patients has been carried out. Various experiments have been performed to study the impact of introducing the multiresolution handcrafted features with CNN's features. The addition of these features separately or combined show their efficiency in enhancing the segmentation results. The approach of adding NSCT directional features with CNN's features has shown the potential to improve the segmentation results when using ten directional coefficients. The inclusion of the low pass sub-band with the input MR image enhances the segmentation results but shows lower performance compared to the model that adds ten directional NSCT coefficients. One reason of this is the need to design a deeper model that is able to handle the low pass coefficient of the NSCT transform and extract the dominant features. The addition of only raw directional NSCT features without the need to increase the dimensionality of the input MR images with the low pass sub-band is the model of choice as it requires less complicated modelling and processing while improving the segmentation performance. On the other hand, the model that concatenates the low pass sub-band with the input image and one directional NSST coefficients has shown a good segmentation performance. The Shear information of one NSST directional coefficient are enough to add the directionality features to the model. Moreover, the combined NSCT and NSST coefficients has shown even better performance where different directional information are combined with CNN's features.

One limitation of these proposed methods is the lack of experimenting the ideal decomposition levels of the multiresolution transforms. The more decomposition levels the more useful directional features are produced. However, there is a trade off between the model complexity and the generation of useful directional coefficients. In this work, the NSST and NSCT coefficients are produced by decomposing the system into a randomly selected number of levels (four levels). The decomposition levels value should be carefully chosen empirically to select the optimal number which generates useful features with a reasonable amount of complexity and computation time.

Another limitation of this work is the complexity component that has been added to the proposed methods by endeavouring to find and select useful handcrafted features. The cumbersome tasks of features extraction and selection which have been omitted from the CNN's approach are introducing overheads in the proposed mechanisms. This limitation will be addressed in the next chapter by proposing a new method which aims at enhancing the segmentation performance without relying on handcrafted features.

Chapter 4

Attention based 3D Deep Convolutional Neural Network for MR Images Segmentation

4.1 Introduction

This chapter is proposing a novel excitation-based CNN to recalibrate the network features adaptively and improve the segmentation of the three brain tissue classes: bone, soft tissue, and air. This approach employs two existing mechanisms of features excitation namely: (1) spatial squeeze and channel excitation (cSE) and (2) channel squeeze and spatial excitation (sSE). The main contributions of this chapter are:

- Aggregating the two excitation mechanisms sequentially and integrating them seamlessly into the 3D convolutional encoder decoder network. The combination of the two excitation blocks sequentially intends to improve the segmentation performance and reduce the model complexity.
- Designing an ensemble of three SE based CNNs to boost the bone segmentation results and increase the model reliability.

Using clinical MR and CT images of fifty patients, this approach is evaluated through a comparison with computed tomography (CT) images as ground truth and validated with other methods in the literature that applied different deep CNN architectures to perform segmentation for MR based attenuation correction. The segmentation performance of the three brain classes is calculated objectively using precision, recall, DSC, and JSC as evaluation metrices. Moreover, the resultant segmented slices of some randomly selected patients from the testing data are evaluated subjectively.

This chapter is structured as follows: The applied excitation based CNN with the experimental setup, conducted experiments, and obtained results are detailed in section 4.2. The ensemble of CNNs approach with the conducted experiments and the obtained results are reported in section 4.3. Finally, the summary with the limitations are given in section 4.4.

4.2 Features Recalibration Mechanism

The down-sampling operation of the CNN using the pooling layers reduces the features map size, yet causes loss of useful features by taking the maximum or average values of each patch of the features map. One approach to overcome this pitfall is the features recalibration mechanism that employs global averaging operation to capture the most distinguishable features and suppress the less distinguishable ones. This recalibration technique constitutes of a generic and resilience component that combines few blocks which can be integrated within any CNN architecture. The recalibration mechanism is the resultant of squeezing the features along the spatial or the channel domain then exciting the same set of features spatially or channel wise. This technique identifies a relationship between the number of channels, the spatial locations, and the features maps.

4.2.1 Channel Excitation Mechanism

The channel excitation depends on the channels interdependencies by squeezing the features maps along the spatial domain then exciting the same set of features along the channel domain. The squeeze and excitation process aims to recalibrate and reweight the features. This type of recalibration is called spatial squeeze and channel excitation (cSE). It squeezes the global spatial features into a channel descriptor. It is more applied in classification applications as it enables the extraction of global context information and improves the channel interdependencies [155].

This mechanism is applied by utilising the global average pooling operation to shrink the features maps spatially followed by two fully connected layers to reduce the model complexity. Aftwerwards, the ReLU activation function is added to initiate some nonlinearity and introduce generalisation. The last component is a sigmoid activation function which is used to rescale the features to the rang [0,1]. The output feature vector is then multiplied by the original features maps for excitement and recalibration. Considering a feature map U with c channels:

$$U = [u_1, u_2, ..., u_c], \tag{4.1}$$

the spatial squeezing is calculated by vector z with k elements:

$$z_k = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_k(i, j), \tag{4.2}$$

where H and W are the height and width of the feature map. By adding the two fully connected layers and the activation functions, z transformed to:

$$s_k = \sigma(W_2(\delta(W_1 z_k))) \tag{4.3}$$

where σ and δ refer to sigmoid and ReLU activation functions respectively, $W_1 \in R^{\frac{c}{rxc}}$ and $W_2 \in R^{\frac{c}{r}}$. The parameter r is the reduction ratio that compresses the c channels to $\frac{c}{r}$ channels that can be selected empirically, W_1 and W_2 are the weights of the first and second fully connected layers. The vector s is multiplied by the features map U to excite and recalibrate the features and produce the final vector:

$$X_{cSE} = [u_1.s_1, u_2.s_2, ..., u_c.s_c] \tag{4.4}$$

The architecture of this block is illustrated in Figure 4.1.

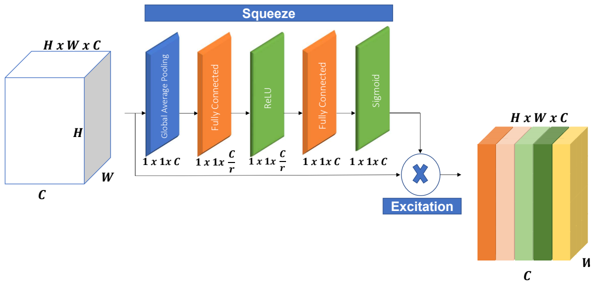


Figure 4.1: The architecture of the spatial squeeze and channel excitation block (cSE).

4.2.2 Spatial Excitation Mechanism

The spatial excitation (sSE) mechanism squeezes the features maps on the channel domain to produce a global scalar value which represents the spatial information which excites the features maps spatially. This approach preserves the global semantic information and enables the representation capabilities of feature maps. It is a useful mechanism for image segmentation [156] as it enhances the salient spatial locations. The architecture of this mechanism is shown in Figure 4.2 where the spatial squeeze part is represented by the convolution operation using a convolutional layer with size $1 \times 1 \times 1$ as follows:

$$q = w \times u_{(i,j)} \tag{4.5}$$

where w is the weight and u is a single spatial location (i,j) at the features map. The output vector q combines all channels for a spatial location then passes through the sigmoid activation function to rescale the features in the range [0,1]. The resultant is multiplied by the feature maps

u to excite and recalibrate the features. The spatially excited feature map is calculated as:

$$X_{sSE} = [\sigma(q_1).u_1, \sigma(q_2).u_2, ..., \sigma(q_{(H,W)}).u_{(H,W)}] \tag{4.6}$$

where H and W are the height and width of the feature map.

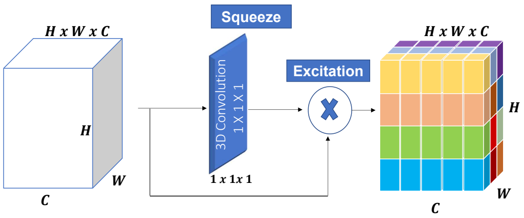


Figure 4.2: The architecture of the channel squeeze and spatial excitation block (sSE).

4.2.3 Combined Spatial and Channel Excitation Mechanisms

The exploitation of the two excitation mechanisms channel and spatial wise (scSE) is achieved by combining them within the CNN architecture. This combination helps to achieve the full potential of recalibrating the features and capturing the dominant ones. The combination has two different configurations: parallel combination and sequential combination.

A Parallel Combination

The parallel combination has been proposed by [152] where the 2D input feature maps pass through the cSE and sSE blocks concurrently. The outputs of both blocks are aggregated using different methods such as max-out, addition, multiplication, and concatenation. This parallel combination is applied herein on the 3D domain using the concatenation aggregation method which provides the best segmentation performance. Figure 4.3 depicts the architecture of the parallel combination.

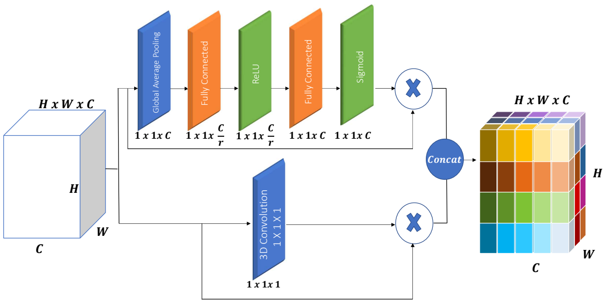


Figure 4.3: The architecture of the parallel combination of channel and spatial excitation blocks.

B Sequential Combination

The sequential combination of the cSE and sSE blocks is introduced in this chapter to decrease the computation complexity and time. Additionally, the rationale behind this combination structure is to create an emphasis on the recalibration mechanism by passing only the excited useful features to the next excitation mechanism. The order of the blocks matters where each SE block receives different excited features either channel or spatial wise. The order of placing the cSE and sSE blocks is tested experimentally. Figure 4.4 illustrates the architecture of the sequential combination of SE blocks which starts with channel then spatial excitation. Figure 4.5 shows the excitation based network architecture where one SE block is placed after each three convolutional layers.

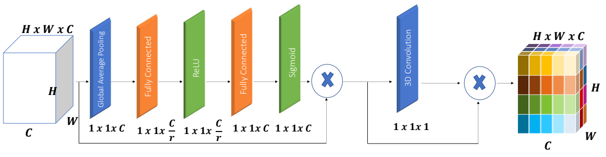


Figure 4.4: The architecture of the sequential combination of channel then spatial excitation blocks (cSE then sSE).

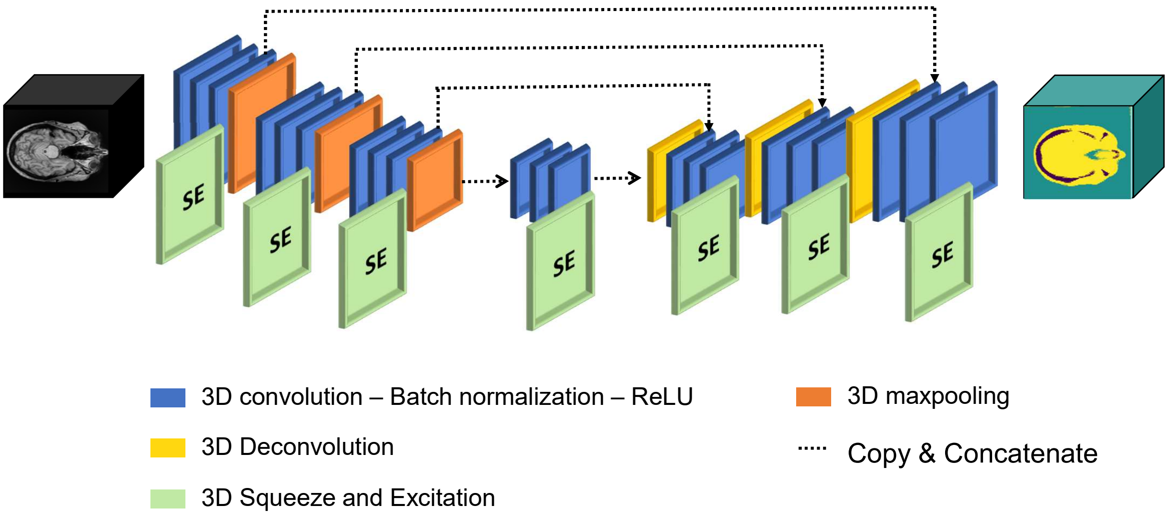


Figure 4.5: The architecture of the convolutional encoder decoder network with multiple squeeze and excitation blocks.

4.2.4 Experimental Setup

A Model Input and Training Parameters

The volume size of each patient is 256 × 256 × 48 where 48 is the number of selected slices per patient. By dividing each patient's volume into overlapped patches to minimize the input size and increase the size of the training datasets, the network input size is 128 × 128 × 48.

The CNN is trained from scratch using thirty patients while ten patients are used for validation, and other ten patients are reserved for testing. The layers weights are initialized using Glorot Uniform initialisation scheme and updated using Adam optimizer with a moving learning rate that starts with 0.0001 then reduces by a factor of 0.75 on plateau mode (i.e., when the training accuracy is not increasing for five continuous epochs). The batch size is two and the maximum number of epochs is two hundred; Yet the training can be stopped before reaching 100 epochs when the validation accuracy stops increasing. The experiments are conducted on Tesla V100 GPU with 16 GB RAM.

B Evaluation Metrics and Validation Methods

The segmentation results are evaluated using a set of metrics by comparing the segmented MR image with the ground truth which are generated from CT images. The evaluation metrics are: precision, recall, DSC, and JSC. The definition of each metric is represented in Chapter 3. The proposed methods are compared with other segmentation studies that applied different CNN's architecture using T1-w MR images for PET attenuation correction application. The validation methods are also described in Chapter 3.

4.2.5 Experimental Results

A Segmentation Results

Several experiments are conducted to study the impact of each type of the excitation blocks using the same CNN architecture. The proposed methods are channel excitation based CNN, spatial excitation based CNN, concurrent channel and spatial excitation based CNN, channel excitation followed by spatial excitation based CNN, and spatial excitation followed by channel excitation based CNN. The segmentation results for bone, soft tissue, and air classes of each method are evaluated for all testing datasets as shown in in Tables 4.1, 4.2, and 4.3, respectively. The segmentation results of the optimal approach (cSE then sSE (sequential)) for each patient are depicted in Tables 4.4, 4.5, and 4.6 for bone, soft tissue, and air classes, respectively and their corresponding box plots are illustrated in Figure 4.6.

The results reveal that the addition of any type of the SE block to the CNN architecture improves the segmentation results and especially the bone class segmentation. The spatial excitation outperforms the channel excitation as it enables the global contextual information on the spatial domain. By observing the segmentation results of bone and soft tissue classes, it is clear the two types of the sequential combination (cSE - sSE and sSE - cSE) outperform the parallel combination. However, the segmentation of the air class using the parallel combination of SE blocks outperforms the other proposed methods. All networks are trained with the same hyperparameters, yet the only hyperparameter that has to be changed while training the parallel combination is the batch size that is set to one to fit the GPU memory size. The batch size of training other networks is two.

Another important observation is that positioning the cSE block before the sSE block in the sequential configuration shows superiority in the segmentation results. Starting with recalibrating the features channel wise helps to emphasise the most useful inter-channel features that will be excited spatially in the next stage to identify the spatial locations of the most useful channels.

The box plots of different evaluation metrics which are shown in Figure 4.6 indicate a high variability on the segmentation results of the bone class with high standard deviation range. The reason of this variability among the patients is the nonhomogeneity of the datasets which consist of patients with different neurological diseases. In fact, different neurological diseases can affect the bone health differently [275]. Hence, the segmentation of each neurological condition should be addressed separately where a model should be trained and tested with datasets of patients who were diagnosed with the same disease. For instance, the DSC value of the bone class of patient 7 in Table 4.4 is very low (DSC = 0.0669) compared to patient 9 (DSC = 0.8789). The proposed model has the capability to segment the bone class accurately for some patients while it fails with others. The lack of the detailed information about the clinical disease of each patient prevents us from making a thorough discussion about the high variability of the bone class.

Table 4.1: The evaluation metrics of the bone class segmentation of the testing datasets using different configurations of SE blocks.

Bone tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.6890	0.5601	0.6179	0.4471
+ cSE	0.7066	0.5689	0.6303	0.4601
+ sSE	0.7054	0.5886	0.6417	0.4724
+ cSE and sSE (parallel)	0.6769	0.5500	0.6069	0.4357
+ cSE then sSE (sequential)	**0.7261**	**0.5857**	**0.6484**	**0.4797**
+ sSE then cSE (sequential)	0.7010	0.5678	0.6274	0.4571

The proposed method is compared with other segmentation approaches in the literature

Table 4.2: The evaluation metrics of the soft tissue class segmentation of the testing datasets using different configurations of SE blocks.

Soft tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.8913	0.9232	0.9070	0.8298
+ cSE	0.8953	0.9230	0.9090	0.8331
+ sSE	0.8944	0.9247	0.9093	0.8337
+ cSE and sSE (parallel)	**0.8987**	0.9187	0.9086	0.8325
+ cSE then sSE (sequential)	0.8962	**0.9290**	**0.9123**	**0.8388**
+ sSE then cSE (sequential)	0.8910	0.9247	0.9076	0.8308

Table 4.3: The evaluation metrics of the air class segmentation of the testing datasets using different configurations of SE blocks.

Air tissue				
Model	PRE	REC	DSC	JSC
Baseline	0.9632	0.9614	0.9623	0.9274
+ cSE	0.9590	0.9619	0.9604	0.9239
+ sSE	**0.9650**	0.9619	0.9635	0.9296
+ cSE and sSE (parallel)	0.9599	**0.9682**	**0.9640**	**0.9306**
+ cSE then sSE (sequential)	0.9641	0.9626	0.9633	0.9293
+ sSE then cSE (sequential)	0.9638	0.9609	0.9624	0.9274

Table 4.4: The evaluation metrics of the bone tissue segmentation of each patient in the testing datasets using the sequential configuration of cSE then sSE blocks.

Bone tissue				
Patients	PRE	REC	DSC	JSC
Patient 1	0.8063	0.6439	0.7160	0.5576
Patient 2	0.8471	0.6916	0.7615	0.6149
Patient 3	0.6570	0.4007	0.4978	0.3314
Patient 4	0.6995	0.2849	0.4049	0.2538
Patient 5	0.8950	0.7765	0.8315	0.7116
Patient 6	0.5155	0.5087	0.5121	0.3442
Patient 7	0.1090	0.0483	0.0669	0.0346
Patient 8	0.5572	0.7028	0.6216	0.4509
Patient 9	0.8993	0.8593	0.8789	0.7839
Patient 10	0.6290	0.5036	0.5594	0.3883

that applied deep CNN to segment MR images for PET attenuation correction [103, 104]. These methods are re-implemented then trained and tested using the brain MR and CT images. The results of these methods are compared with the optimal proposed approach as shown in Table 4.7, 4.8, and 4.9 for bone, soft tissue, and air classes, respectively.

Table 4.5: The evaluation metrics of the soft tissue segmentation of each patient in the testing datasets using the sequential configuration of cSE then sSE blocks.

Soft tissue				
Patients	PRE	REC	DSC	JSC
Patient 1	0.9397	0.9530	0.9463	0.8981
Patient 2	0.9278	0.9425	0.9351	0.8781
Patient 3	0.9136	0.9282	0.9208	0.8532
Patient 4	0.8417	0.9532	0.8940	0.8083
Patient 5	0.9400	0.9578	0.9488	0.9026
Patient 6	0.9226	0.8920	0.9070	0.8299
Patient 7	0.9235	0.9070	0.9151	0.8436
Patient 8	0.9521	0.8910	0.9206	0.8528
Patient 9	0.9680	0.9671	0.9676	0.9372
Patient 10	0.6766	0.8979	0.7717	0.6282

Table 4.6: The evaluation metrics of the air tissue segmentation of each patient in the testing datasets using the sequential configuration of cSE then sSE blocks.

Air tissue				
Patients	PRE	REC	DSC	JSC
Patient 1	0.9727	0.9908	0.9817	0.9640
Patient 2	0.9593	0.9849	0.9719	0.9454
Patient 3	0.9531	0.9762	0.9645	0.9315
Patient 4	0.9801	0.9780	0.9790	0.9589
Patient 5	0.9808	0.9908	0.9858	0.9720
Patient 6	0.9651	0.9917	0.9782	0.9574
Patient 7	0.9483	0.9947	0.9709	0.9435
Patient 8	0.9753	0.9915	0.9833	0.9672
Patient 9	0.9846	0.9923	0.9884	0.9771
Patient 10	0.9133	0.7508	0.8241	0.7008

Table 4.7: The comparison of the bone segmentation results using the optimal proposed method with other techniques reported in the literature using four different evaluation metrics: precision, recall, DSC, and JSC.

Bone tissue				
Model	PRE	REC	DSC	JSC
SegNet[103]	0.6278	0.3649	0.4616	0.3000
SegNet + CRF[104]	0.6278	0.3649	0.4616	0.3000
Proposed method	**0.7261**	**0.5857**	**0.6484**	**0.4797**

The segmentation results of some randomly selected slices from the testing datasets are shown in Figure 4.7. The first column represents the input MR images, the second column shows their corresponding CT images which used as ground truth, and the last three columns

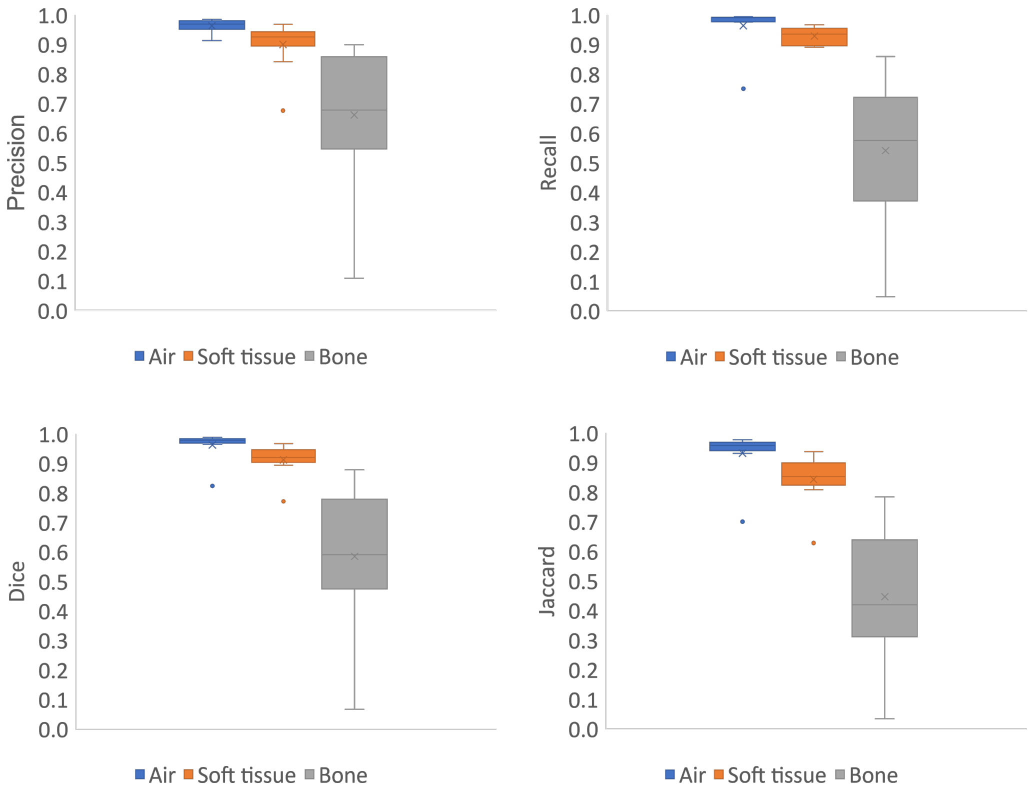


Figure 4.6: The box plot of the evaluation metrics of the three tissue classes of the testing datasets using the sequential configuration of cSE then sSE blocks.

Table 4.8: The comparison of the soft tissue segmentation results using the optimal proposed method with other techniques reported in the literature using four different evaluation metrics: precision, recall, DSC, and JSC.

Soft tissue				
Model	PRE	REC	DSC	JSC
SegNet[103]	0.8696	0.8810	0.8753	0.7783
SegNet + CRF[104]	0.8696	0.8810	0.8753	0.7783
Proposed method	**0.8962**	**0.9290**	**0.9123**	**0.8388**

Table 4.9: The comparison of the air segmentation results using the optimal proposed method with other techniques reported in the literature using four different evaluation metrics: precision, recall, DSC, and JSC.

Air tissue				
Model	PRE	REC	DSC	JSC
SegNet[103]	0.9083	0.9510	0.9291	0.8677
SegNet + CRF[104]	0.9083	0.9510	0.9291	0.8677
Proposed method	**0.9641**	**0.9626**	**0.9633**	**0.9293**

illustrate the segmentation results using the proposed approach (sequential cSE - sSE) , Segnet network [103], and Segnet with CRF technique [104].

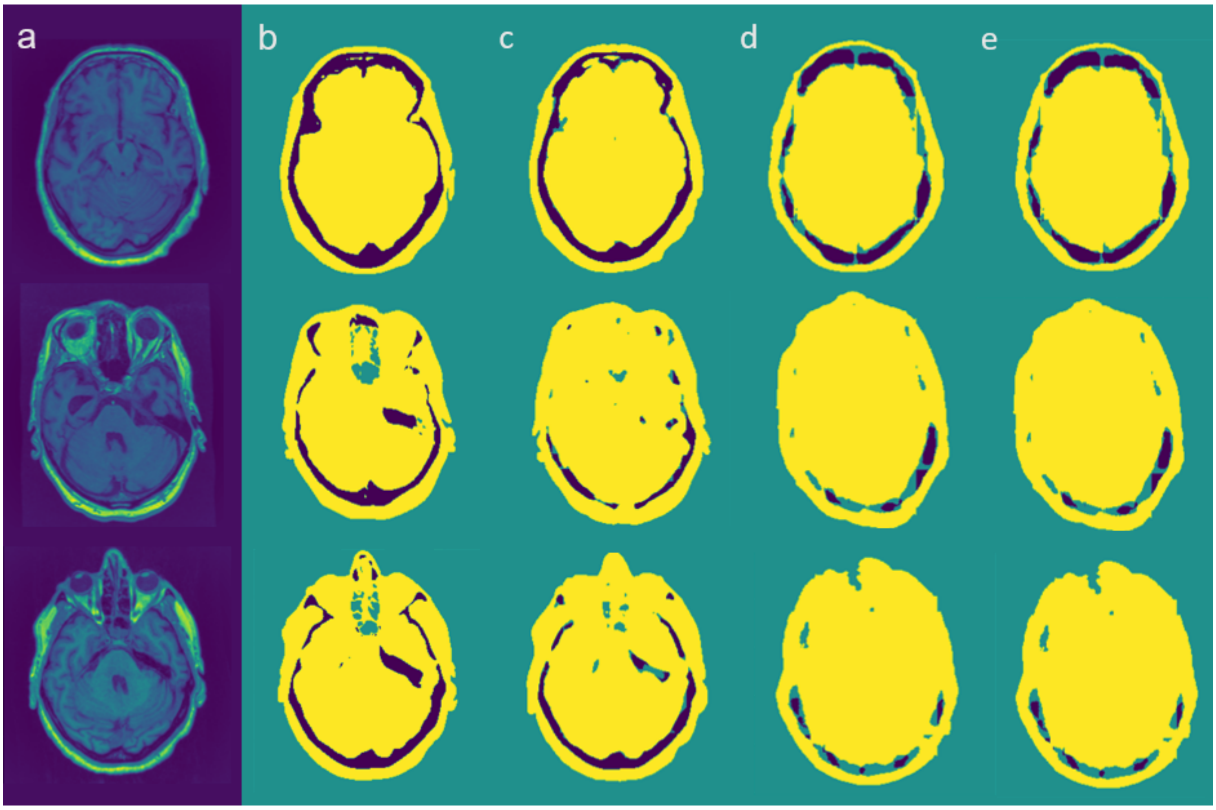


Figure 4.7: Some slices of the testing data. (a) MR image, (b) CT or ground truth, (c) segmented MR image with the proposed deep network, (d) Segmented MR image generated by [103], (e) segmented MR generated by [104]. The colors in columns b - e refer to the following classes: green is the air, yellow is the soft tissue, and the purple is the bone.

B Position of SE Block

The placement of the squeeze and excitation block within the CNN is studied by experimenting different configurations shown in Figure 4.8 to find the optimal place as . The comparison between the different network architectures is carried out by calculating the DSC. According to the comparison results which are recorded in Table 4.10, the chosen architecture consists of one sequential combination of the spatial and channel excitation block after each three convolutional layers at the encoder part, the bottleneck part, and the decoder part. The enhancement in the bone segmentation is more significant compared to soft tissue and air classes with percentage increase of 4.93%, 0.75%, and 0.17% for bone, soft tissue, and air classes, respectively.

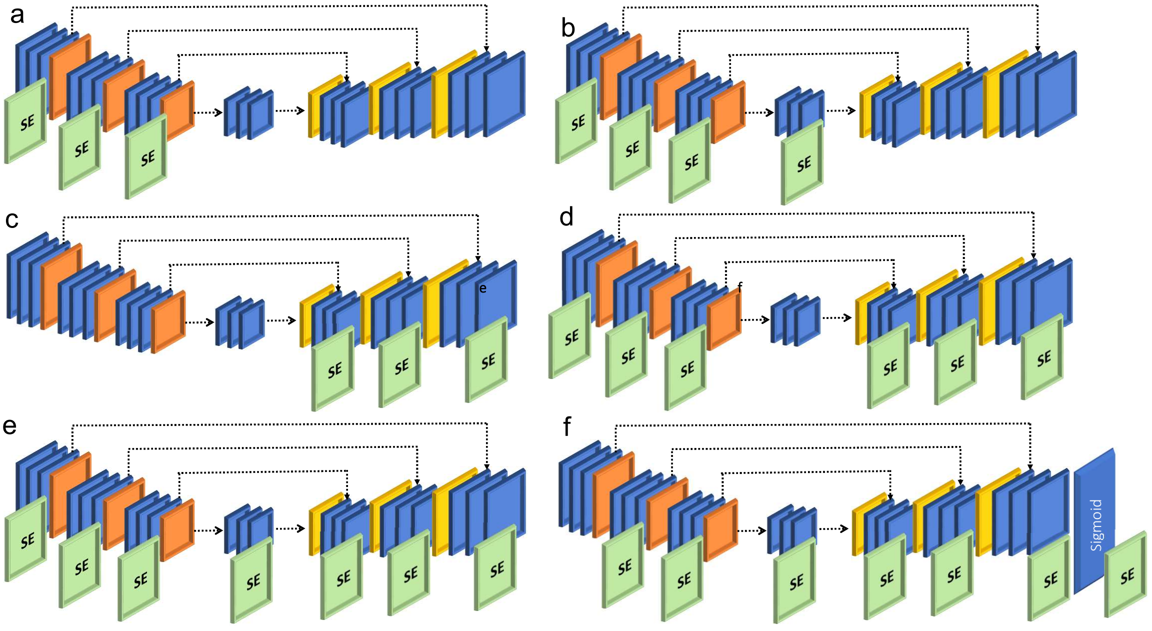


Figure 4.8: Different network architectures (a-f) of placing the SE block at different locations.

Table 4.10: The DSC of the three tissue classes which are segmented using different models with different positions of the csSE block.

Model	Bone	Soft tissue	Air
without csSE block	0.6179	0.9070	0.9623
csSE at the encoder	0.6258	0.9078	0.9629
csSE at the encoder + bottleneck layer	0.6393	0.9089	0.9624
csSE at the decoder	0.6273	0.9074	0.9621
csSE at the encoder + decoder	0.6393	0.9055	0.9624
csSE at the encoder + decoder + bottleneck layer	**0.6484**	**0.9123**	**0.9633**
csSE at the encoder + decoder + bottleneck layer + classifier layer	0.6429	0.9087	0.9616

C Combining Mechanism

The comparison between the sequential and the parallel combination of channel and spatial excitation components is conducted by measuring the DSC of each configuration. The main drawback of the parallel combination is the large number of parameters which requires a large GPU memory and longer computation time. Although the memory size of the utilised GPU in this study is 16 GB, the parameters number of CNN with parallel combination of SE blocks could not be handled by the available memory. Due to this limitation, two different architectures are proposed to minimize the number of parameters that can be trained with the available GPU capacity. The first experiment is conducted by placing the cSE and sSE blocks on parallel into a network with only ten convolutional layers and two downsampling layers using the batch size of two. The second experiment is conducted by placing the cSE and sSE blocks on parallel using the baseline architecture illustrated in Figure 5.1 with the decrease of the batch size to one. The segmentation results are compared using the DSC as shown in Table 4.11.

Table 4.11: The comparison of the segmentation results (DSC) of the three tissue classes which are generated by applying different combination mechanisms of the sSE and cSE blocks.

Combination mechanism	Convolution	Pooling	Batch size	DSC (bone)	DSC (soft tissue)	DSC (air)
Parallel	21	3	1	0.6069	0.9086	0.9640
Sequential	21	3	1	**0.6236**	**0.9094**	**0.9660**
Parallel	10	2	2	0.6368	0.9104	0.9635
Sequential	10	2	2	**0.6402**	**0.9111**	**0.9634**

The sequential combination of cSE then sSE blocks not only reduces the number of training parameters but also shows superiority of the segmentation results in terms of DSC by testing two different CNN architectures. The DSC of the bone class shows substantial improvement compared to other classes which reveal slight enhancement. The sequential combination mechanism is capable of recalibrating the features in channel wise then spatial wise significantly by helping the network to recalibrate the useful spatial regions of the most dominant channels.

D Relationship Between SE block and Convolution Layers

The design of placing the SE block after n convolutional layers is studied and evaluated by conducting several experiments. In theory, the addition of one SE block after each convolutional layer is the optimal design especially with shallow convolutional networks that have few trainable parameters. However, due to GPU memory limitation, it is not possible to add one SE block after each convolutional layer in deep networks such as the proposed network which consists of twenty one convolutional layers. The trade off between the number of convolutional layers and the number of SE blocks leads to include one SE after each three convolutional layers which generates experimentally the optimal results. The DSC values of the segmentation results using different network designs are shown in Table 4.12.

The inclusion of more SE blocks can substitute the addition of more convolutional layers to the model. The experiments show that using less convolutional layers with the addition of one SE block after each convolutional layer is able to generate better features rather than using more convolutional layers with less SE blocks.

Table 4.12: The study of the relation between the number of SE blocks and the convolutional layers by comparing the DSC of each tissue class.

SE type	Conv	SE	DSC (bone)	DSC (soft tissue)	DSC (air)
cSE	7	7	0.6249	**0.9097**	0.9610
cSE	14	7	0.6108	0.9077	0.9601
cSE	21	7	**0.6302**	0.9090	**0.9614**
cSE-sSE	7	7	0.6394	0.9095	0.9620
cSE-sSE	14	7	0.6192	0.9080	0.9629
cSE-sSE	21	7	**0.6484**	**0.9123**	**0.9633**

E Model Complexity

The addition of any SE block to the CNN increases the number of trainable parameters, the model complexity, and the computation time. However, each excitation type adds different amount of complexity. The additional number of parameters by adding the cSE block is calculated by:

$$I = \frac{2}{r} \sum_{s=1}^{S} N_s C_s^2 \tag{4.7}$$

where r is the reduction ratio that is set to 16 in all experiments, S is the number of stages/blocks in the network, N is the number of repetition of the cSE block which is 1 in all experiments, and C is the number of channels in the feature map.

On the other hand, the sSE block only introduces C_s additional parameters. The overall model complexity is given by:

$$X = \frac{2}{r} \sum_{s=1}^{S} N_s C_s^2 + C_s \tag{4.8}$$

The number of parameters of different architectures and configurations along with the increase rate of the model complexity are illustrated in Table 4.13. The results show that the sequential combination of SE components can be added to any architecture with almost no cost.

Table 4.13: The number of parameters and the complexity increase rate of the model using different excitation mechanisms.

Model	Parameters	Complexity increase rate
Baseline	37, 834, 115	
cSE	37, 942, 659	0.3%
sSE	37, 835, 530	0.004%
Parallel (cSE and sSE)	55, 438, 154	47%
Sequential (cSE then sSE)	37, 889, 802	0.15%

4.3 Ensemble of 3D CNN

Ensemble learning is a technique used to improve the stability of the designed framework and enhance the segmentation performance. The ensemble method also aims at combining the strength of different learning models to improve the segmentation performance [194]. The main objective of proposing an ensemble approach is to train multiple deep CNNs that complement each other's deficiencies for segmenting the brain into three tissue classes. The outputs of all models are aggregated to provide more robust and reliable results.

The diversity of the network's components generates complementary segmentation results. It has been shown that using base models with similar learning capabilities helps to build an effective model. At this end, the main contribution herein is the employment of the same CNN's architecture to build a base model with modifications in the activation function and the objective function (loss). The resultant probability maps of all base models are averaged to produce the final segmentation probability map.

4.3.1 ReLU vs Leaky ReLU Activation Function

ReLu is a non-linear activation function which has been widely applied in the hidden layer of the CNN. This simple function solves the issue of vanishing gradient that is caused by sigmoid and tanh activation functions. The ReLu is a simple and light computation activation function which returns the positive input value or zero if the input value is greater or equal to zero. Otherwise, it returns zero [276]. This function can be described mathematically as follows:

$$f(x_i) = \begin{cases} x_i, & \text{if } x_i \geq 0 \\ 0, & \text{if } x_i < 0 \end{cases}$$

On the other hand, ReLU has some limitations such as "dying ReLU" problem. This problem is caused by turning off the neurons which have negative values. Over time the network ends up with loosing large number of neurons which are not contributing to the process of learning and discriminating their inputs. This issue has been solved by introducing the leaky ReLU (LReLU) activation function [277] which produces a small slope for negative values instead of zero. The degree of slope is a predefined coefficient which has to be tuned during the training process. The mathematical definition of LReLU is:

$$f(x) = \begin{cases} x_i, & \text{if } x \geq 0 \\ \dfrac{x_i}{\alpha_i}, & \text{if } x_i < 0 \end{cases}$$

where α_i is a fixed hyperparameter in the range of $(1, +\infty)$ [278].

4.3.2 Dice vs Focal Tversky Loss

The dice loss is a commonly used objective function to train CNN with imbalanced datasets. The mathematical definition of the dice is defined in Chapter 3. Beside the dice loss, the focal tversky (FT) loss which is proposed recently by [279]. This objective function shows its ability to achieve better trade off between precision and recall on small regions of the image. The focal loss function which is based on Tversky index for binary classification takes the following form:

$$L_{FT} = (1 - \frac{\sum_{i=1}^{N} p_{ic} t_{ic} + \epsilon}{\sum_{i=1}^{N} p_{ic} t_{ic} + \alpha \sum_{i=1}^{N} p_{ic'} t_{ic} + \beta \sum_{i=1}^{N} p_{ic} t_{ic'} + \epsilon})^{\gamma} \tag{4.9}$$

where p_{ic} and $p_{ic'}$ are the predicted probabilities that pixel i belongs to class c and class c', respectively. t_{ic} and $t_{ic'}$ are the true value that pixel i belongs to class c and class c', respectively. α and β are hyperparameters which should be tuned to improve the precision and recall with the occurrence of class imbalacning issue. ϵ is a small value which prevents the division by zero. The parameter γ can be tuned in the range of [1,3].

The focal tversky loss formula can be rewritten in the following format to use it for mulitclass segmentation:

$$L_{FT} = (1 - \frac{TP + \epsilon}{TP + \alpha FN + \beta FP + \epsilon})^{\gamma} \tag{4.10}$$

where TP, FN, FP stand for true positive, false negative, and false positive respectively.

In order to create diversity, an ensemble of three CNNs models which employ different combinations of ReLU and LReLU activation functions and dice and focal tversky losses. Figure 4.9 illustrates an example of an ensemble of three CNNs.

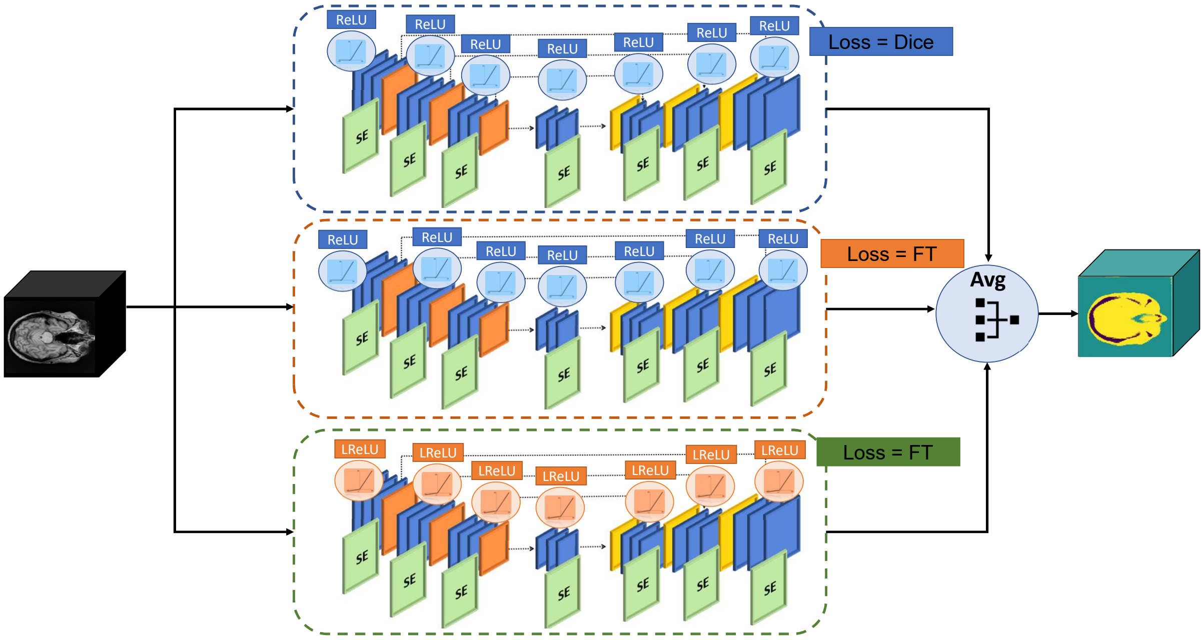


Figure 4.9: The design of the proposed ensemble of three CNN models.

4.3.3 Experimental Results

The ensemble model consists of different base CNN's models which employ different activation and objective functions. The selected functions show superior performance when tested with the baseline model which is presented in Chapter 3 (section 3.3). By using two activation functions (ReLU and LReLU) and two objective functions (dice and focal tversky loss), four different CNN architectures are tested using the DSC. The conducted experiments show that CNN with focal tversky loss outperforms CNN with dice loss and CNN with LReLu function generates better bone segmentation results than CNN with ReLU function as shown in Figure 4.10. The variation and the diversity of the obtained segmentation results for the bone class enables the aggregation of these models to create an ensemble of CNNs models which aims at enhancing the segmentation performance for the three tissue classes generally and the bone class specifically.

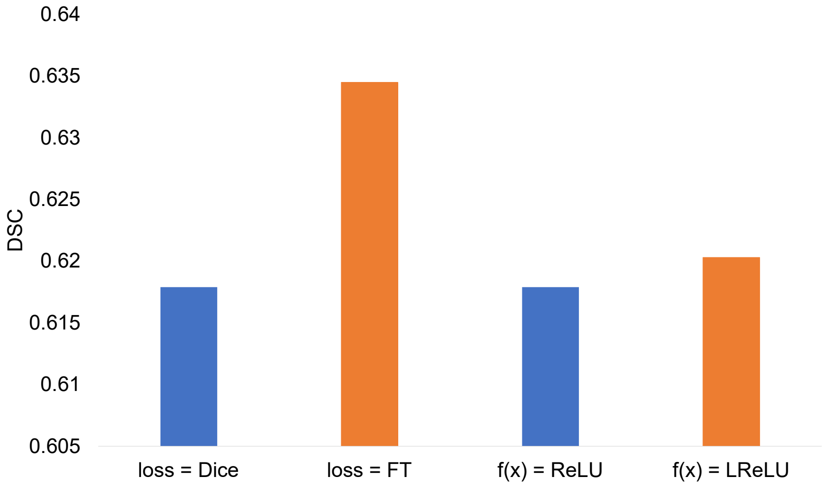


Figure 4.10: The dice of the bone class using different loss and activation functions using the baseline model.

Tables 4.14, 4.15, and 4.16 illustrate the evaluation of the segmentation results of each base CNN's model which can be used to build the ensemble model. The obtained segmentation results are calculated using the precision, recall, DSC, and JSC of each tissue class (air, bone, and soft tissue). The results demonstrate that the base model which employs dice loss and ReLU activation function achieves better segmentation results for the soft tissue and air classes. On the other hand, the model which applies the focal tversky loss and ReLU activation function obtains higher DSC and JSC for the bone class. The LReLU activation function with dice loss generates the highest sensitivity score (less false negative).

Table 4.14: The evaluation of the segmentation results of the bone class by applying four base CNN models.

Bone tissue				
Model	PRE	REC	DSC	JSC
Base 1: Dice-ReLU	**0.7261**	0.5857	0.6484	0.4797
Base 2: FT-ReLU	0.7203	0.5908	**0.6491**	**0.4805**
Base 3: Dice-LReLU	0.6786	**0.5971**	0.6352	0.4655
Base 4: FT-LReLU	0.6997	0.5891	0.6397	0.4702

Table 4.15: The evaluation of the segmentation results of the sof tissue class by applying four base CNN models.

Soft tissue				
Model	PRE	REC	DSC	JSC
Base 1: Dice-ReLU	0.8962	**0.9290**	**0.9123**	**0.8388**
Base 2: FT-ReLU	**0.8984**	0.9236	0.9108	0.8363
Base 3: Dice-LReLU	0.8953	0.9182	0.9066	0.8292
Base 4: FT-LReLU	0.8969	0.9192	0.9080	0.8314

Table 4.16: The evaluation of the segmentation results of the air class by applying four base CNN models.

Air tissue				
Model	PRE	REC	DSC	JSC
Base 1: Dice-ReLU	**0.9641**	0.9626	**0.9633**	**0.9293**
Base 2: FT-ReLU	0.9605	**0.9634**	0.9620	0.9267
Base 3: Dice-LReLU	0.9633	0.9603	0.9618	0.9264
Base 4: FT-LReLU	0.9596	0.9621	0.9608	0.9246

Different combinations with different ensemble sizes ranging from two to four have been designed and tested using the same evaluation metrics: precision, recall, DSC, and JSC as shown in Tables 4.17, 4.18, and 4.19 . Some ensembles of size two show better precision and sensitivity for the bone class while the ensemble of size three outperforms other ensembles considering the different evaluation metrics for the three tissue classes. Figures 4.11, 4.12, 4.13, and 4.14 depict the precision, recall, DSC, and JSC of the bone class, respectively using different ensembles combinations. The approach that combines all four base models shows lower performance compared to three base models due to the increase of results variability.

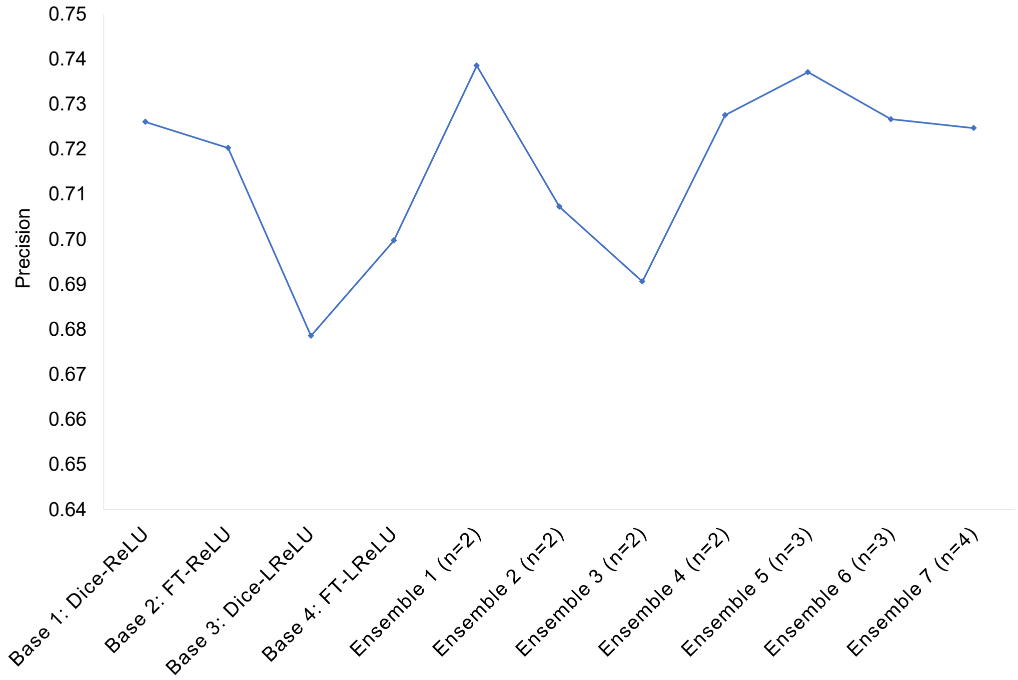


Figure 4.11: The precision of the bone class using different base and ensemble models.

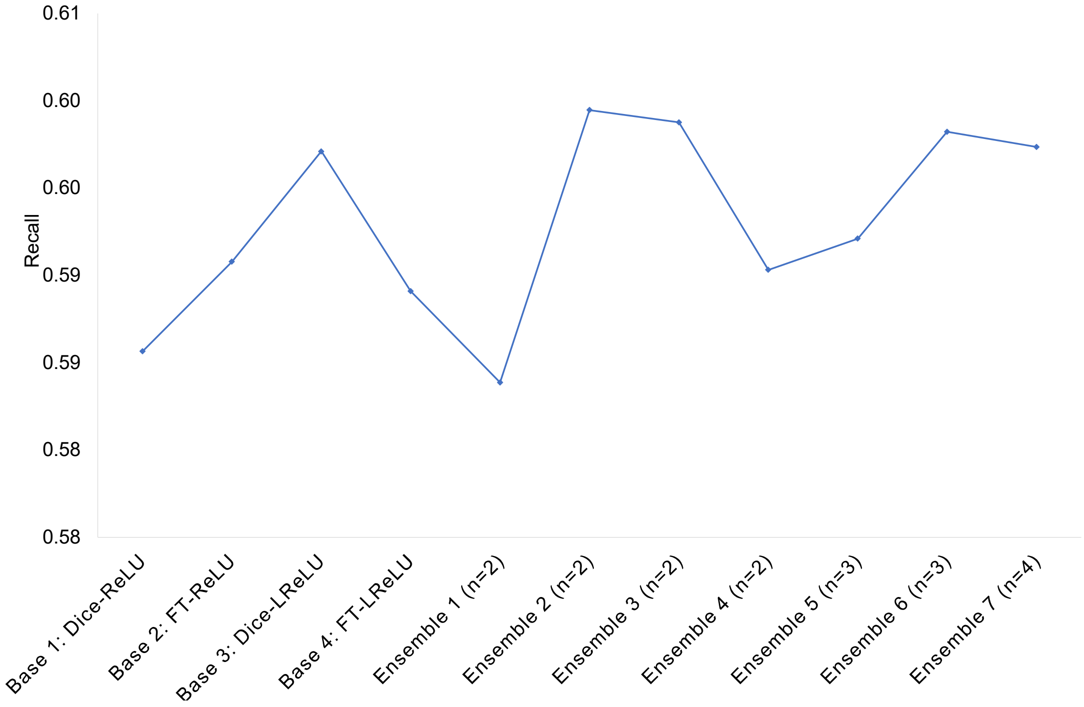


Figure 4.12: The recall of the bone class using different base and ensemble models.

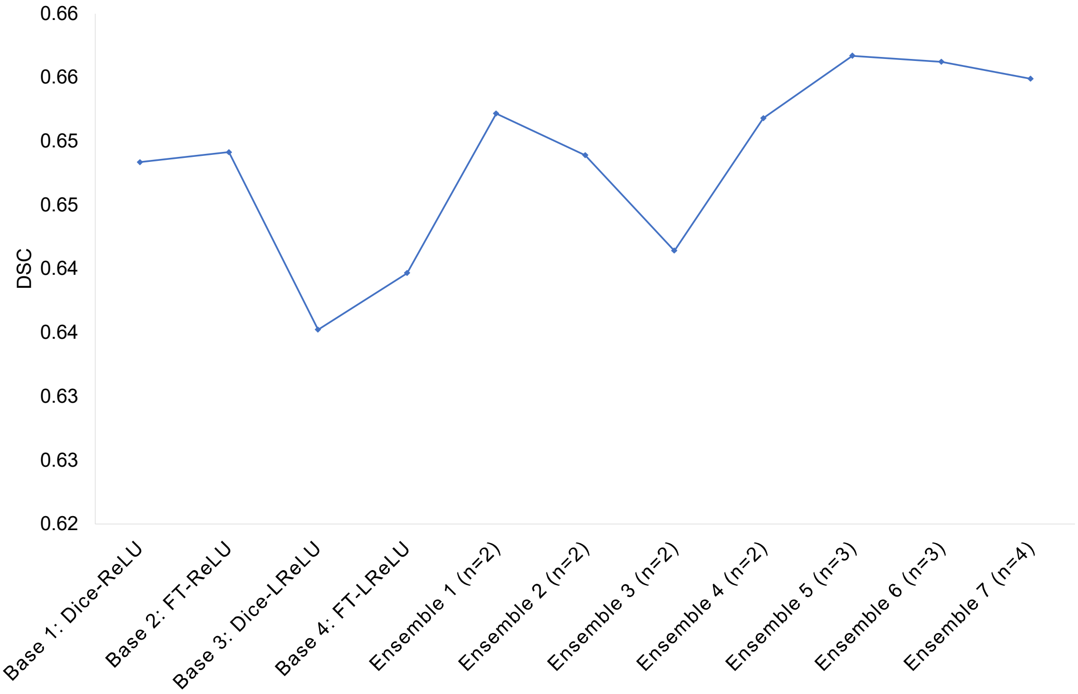


Figure 4.13: The DSC of the bone class using different base and ensemble models.

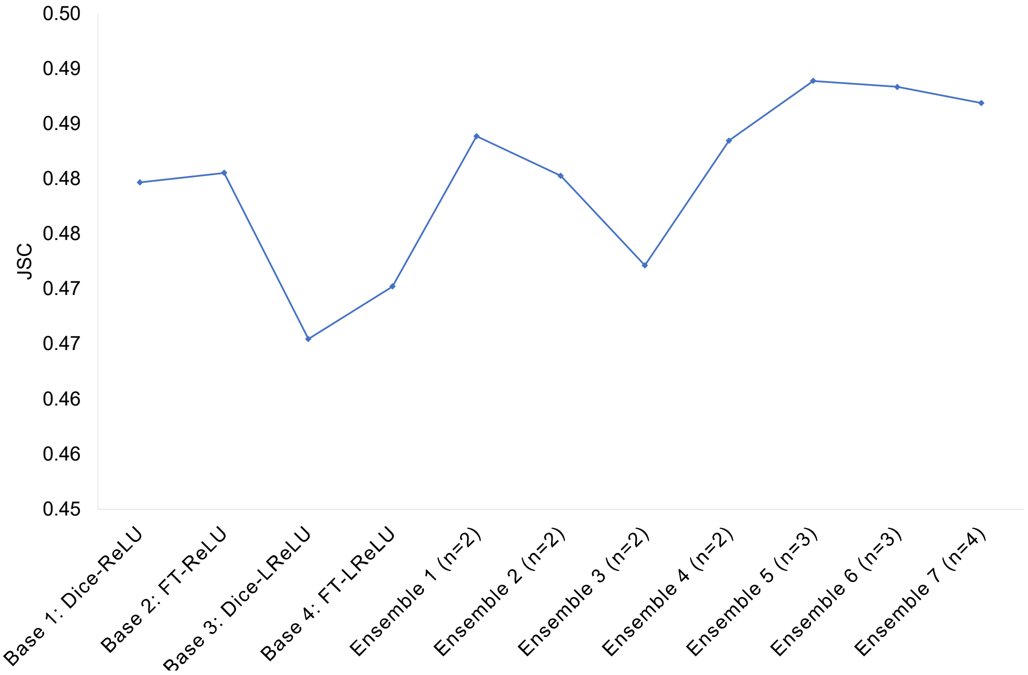


Figure 4.14: The JSC of the bone class using different base and ensemble models.

Table 4.17: The evaluation of the bone segmentation results of different ensemble models with different sizes.

Bone tissue					
Ensemble size	Models	PRE	REC	DSC	JSC
1	Baseline	0.6890	0.5601	0.6179	0.4471
2	ReLU-Dice + ReLU-FT	**0.7386**	0.5839	0.6522	0.4839
2	Dice-ReLU + Dice-LReLU	0.7072	**0.5995**	0.6489	0.4803
2	LReLU-Dice + LReLU-FT	0.6907	0.5988	0.6414	0.4721
2	FT-ReLu + FT-LReLu	0.7276	0.5903	0.6518	0.4835
3	Dice-ReLU + FT-ReLU + FT-LReLU	0.7371	0.5921	**0.6567**	**0.4889**
3	Dice-ReLU + FT-ReLU + Dice-LReLU	0.7267	0.5982	0.6562	0.4883
4	All base models	0.7247	0.5974	0.6549	0.4869

Table 4.18: The evaluation of the soft tissue segmentation results of different ensemble models with different sizes.

Soft tissue					
Ensemble size	Models	PRE	REC	DSC	JSC
1	Baseline	0.8913	0.9232	0.9070	0.8298
2	ReLU-Dice + ReLU-FT	0.8964	**0.9327**	0.9142	0.8420
2	Dice-ReLU + Dice-LReLU	0.8972	0.9251	0.9109	0.8364
2	LReLU-Dice + LReLU-FT	0.8960	0.9205	0.9081	0.8316
2	FT-ReLu + FT-LReLu	0.8975	0.9277	0.9124	0.8389
3	Dice-ReLU + FT-ReLU + FT-LReLU	0.8973	0.9319	**0.9143**	**0.8421**
3	Dice-ReLU + FT-ReLU + Dice-LReLU	**0.8976**	0.9302	0.9136	0.8409
4	All base models	0.8973	0.9299	0.9133	0.8404

Table 4.19: The evaluation of the air segmentation results of different ensemble models with different sizes.

Air tissue					
Ensemble size	Models	PRE	REC	DSC	JSC
1	Baseline	0.9632	0.9614	0.9623	0.9274
2	ReLU-Dice + ReLU-FT	0.9655	0.9632	0.9643	0.9311
2	Dice-ReLU + Dice-LReLU	0.9651	0.9622	0.9636	0.9298
2	LReLU-Dice + LReLU-FT	0.9637	0.9611	0.9624	0.9275
2	FT-ReLu + FT-LReLu	0.9631	0.9631	0.9631	0.9289
3	Dice-ReLU + FT-ReLU + FT-LReLU	0.9658	**0.9633**	**0.9645**	**0.9315**
3	Dice-ReLU + FT-ReLU + Dice-LReLU	0.9662	0.9627	**0.9645**	0.9313
4	All base models	**0.9663**	0.9627	**0.9645**	**0.9315**

The ensemble with three base CNNs using a combination of dice-ReLU, FT-ReLU, and FT-LReLU shows superior segmentation results considering the three tissue classes. Figures

4.15, 4.16, and 4.17 illustrate the effectiveness of the ensemble approach by comparing the segmentation results of a single base CNN's model with the ensemble model with size three using different evaluation metrics of the three tissue classes. The ensemble effectively improves the precision, the DSC, and the JSC of the bone class while the recall (sensitivity) is better while using one base model (focal tversky with LReLU). The soft tissue segmentation performance is also improved in term of recall, DSC, and JSC. However, the precision of the base model (dice - ReLU) is better. The air class segmentation is also improved while using the ensemble model considering the four evaluation metrics.

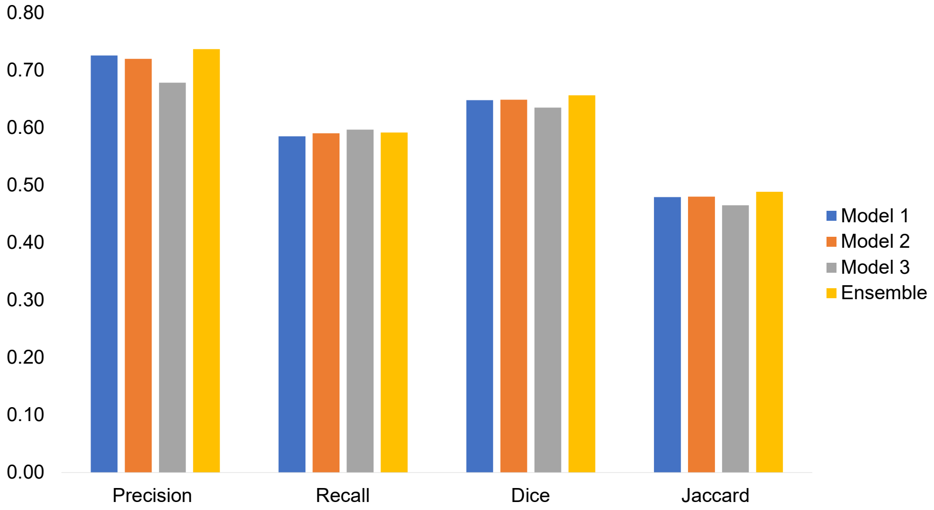


Figure 4.15: The evaluation metrics of the bone class using the optimal ensemble model which consists of three base CNN models (dice-ReLU, FT-ReLU, and FT-LReLU).

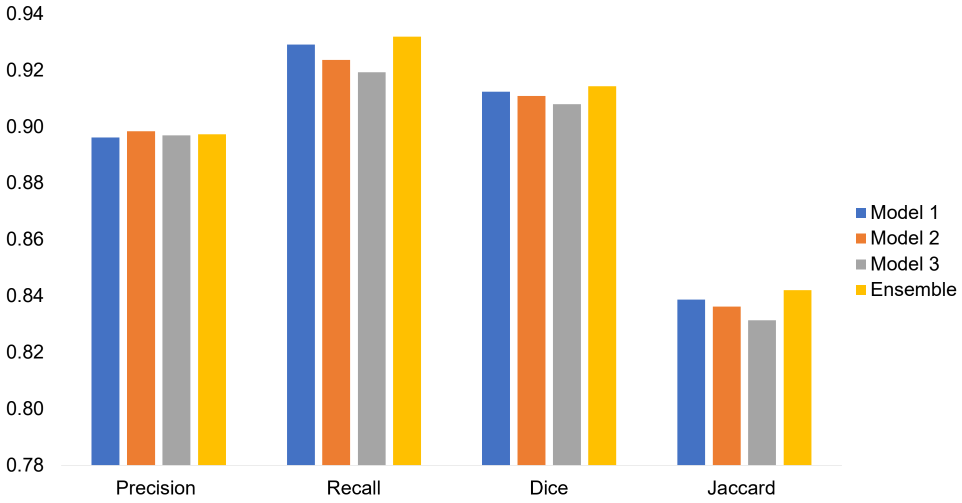


Figure 4.16: The evaluation metrics of soft tissue class using the optimal ensemble model which consists of three base CNN models (dice-ReLU, FT-ReLU, and FT-LReLU)

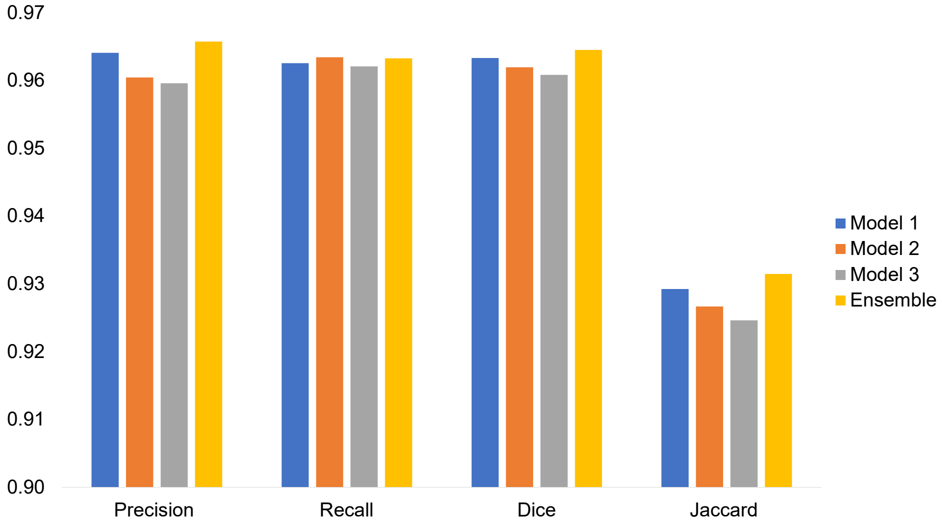


Figure 4.17: The evaluation metrics of air class using the optimal ensemble model which consists of three base CNN models (dice-ReLU, FT-ReLU, and FT-LReLU)

The segmentation maps of one randomly selected slice from patient 5 in the testing dataset are shown in Figure 4.18 where it depicts the capability of the ensemble model to segment the highlighted area in the brain more accurately. Compared to the ground truth in column b, the first and third base models in columns (c) and (e) generate more false positive of air tissue. The second base model in column d produces more false positive of bone tissue. The segmentation map of the ensemble model in column f is the closest one to the ground truth with less false positive pixels.

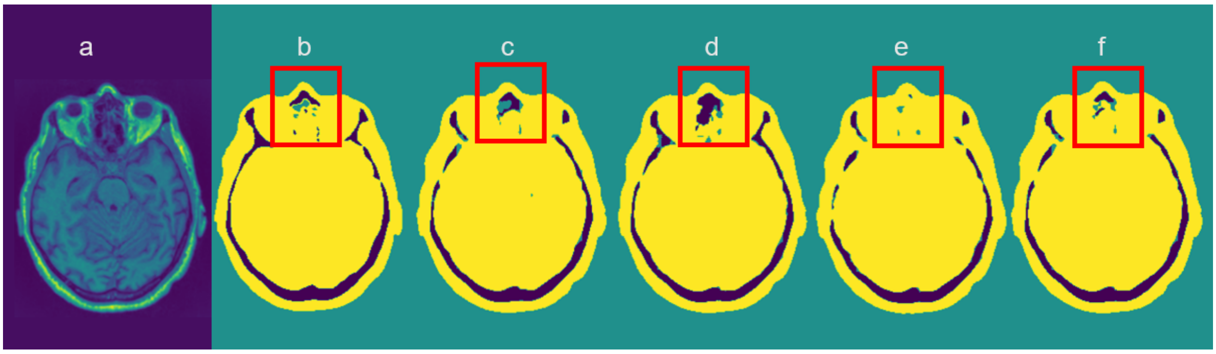


Figure 4.18: The segmentation maps of (a) one randomly selected MR slice with its corresponding (b) CT image using the (c-e) three base models and (f) the proposed ensemble model.

4.4 Summary and Limitations

This chapter has presented a novel way to enhance the T1-w MR images segmentation and reduce the model computation complexity. The excited-based fully CNN model has shown its capability to improve the segmentation of the brain into three tissue classes: bone, soft tissue, and air. This approach embeds two different excitation blocks into the CNN to recalibrate the features by capturing the most useful and distinguishable ones in channel and spatial domain using the sequential aggregation approach.

Another proposed approach is the employment of the ensemble of multiple CNNs to improve the brain segmentation and generate reliable results. Three different CNNs architectures have been aggregated using the averaging operation. Each base CNN applies different activation functions such as ReLU and LReLU and different objective functions such as dice loss and focal tversky loss.

Several experiments have been conducted using clinical 3D MR brain images of fifty patients to explore the impact of each type of the SE blocks on the segmentation results to find the optimal CNN architecture that improves the segmentation performance and reduces the model complexity. The selected excitation based network firstly deploys the channel excitation block to excite the useful channels then applies the spatial attention on certain regions with informative details using the spatial excitation block. By comparing the segmented MR images with the ground truth CT images, the excitation based CNN approach has shown its efficiency to improve the segmentation performance with minimal increase of model complexity and computation time. The results has also shown that the features recalibration mechanism outperforms other proposed methods in the literature. Additionally, the ensemble model has revealed its potential to boost the segmentation results compared to the segmentation results obtained by each base CNN's model.

The datasets consist of patients with different neurological diseases which explains the variability of the bone segmentation results. However, the diversity of individual models with different activation and objective functions has shown its efficiency in building a more accurate ensemble model. This approach needs further evaluation by using the generated segmented images to reconstruct the PET images to evaluate the attenuation correction quality.

Despite the superior performance of the excitation based CNN model with minimal increase of the number of parameters and computation time, this approach needs to be explored with the transfer learning technique to efficiently reduce the complexity of the model learning process. The transfer learning helps to employ the learned knowledge from a specific dataset to another dataset where the model can use the transferred knowledge as a starting point of the learning process to perform a new task or work with data from new domain.

Moreover, the design of the ensemble model requires diverse base CNNs to create a high level of variation which helps to complement the information of each individual model. The diversity is induced by designing models which generate uncorrelated segmentation results. The combination of different CNN's architectures helps to build such diversity and variation.

Chapter 5

Transfer Learning and Multi MR Sequences for MR Imaging Segmentation

5.1 Introduction

This chapter represents the application of transfer learning using multi MR sequences for brain segmentation. The transfer of knowledge from T1-w to T2-w MR sequences is studied by investigating the optimal mechanism of fine tuning the pretrained CNN and the required size of the target datasets (T2-w MR) to fine tune the network. Four mechanisms of transfer learning have been studied which are: (1) transfer learning without fine tuning, (2) transfer learning with open fine tuning where the CNN's weights are initialized using the pretrained network's weights, (3) transfer learning with conservative fine tuning by retraining some pretrained layers and freezing others, and (4) hybrid transfer learning which is the proposed contribution in this chapter aiming at solving the issues of other types of fine tuning and enhances the transfer learning results. The hybrid transfer learning approach consists of two parallel encoders which are fused using an aggregation component. The output of both encoders is fed to a single decoder. Moreover, an augmentation approach is proposed where the T2-w MR images are considered as an intensity based augmentation technique. Both T1-w and T2-w MR images are used to train a CNN from scratch to increase the size of the training dataset and build a multi

MR sequences segmentation model. The adopted CNN architecture follows the excitation based CNN which is proposed in Chapter 4.

The impact of transfer learning is tested on segmenting T1-w and T2-w MR images using four evaluation metrics namely: precision, recall, DSC, and JSC. The visual comparison of the segmentation results of some randomly selected slices from the testing data is also conducted.

This chapter is structured as follows: The different applied methods of transfer learning are presented in section 5.2. The experimental setup is described in section 5.3. The conducted experiments and the obtained segmentation results with the discussion are detailed in section 5.4. Finally, the summary with the limitations are given in section 5.5.

5.2 Transfer Learning Mechanisms

The proposed transfer learning methodology consists of four different strategies which are described in sections 5.2.1, 5.2.2, 5.2.3, and 5.2.4. The backbone network which is used to design and experiment the transfer learning methods is an excitation based CNN that consists of an encoder, a decoder, and multiple squeeze and excitation blocks embedded at different levels of the network. The network architecture which is similar to the network proposed in chapter 4 is illustrated in Figure 5.1.

5.2.1 Transfer Learning using Open Fine Tuning

Fine tuning is the conventional way to conduct transfer learning task. The fine tuning process requires to adjust the weights of pretrained CNN layers. The most common way to fine tune networks is to initialise the new network with the weights of the pretrained network and fine tune all or some number of layers. The required size of the data to fine tune the model is much less than training a network from scratch. The open fine tuning strategy requires to initialise the network with the weights of the pretrained network with T1-w MR images and adjust them by unfreezing all pretrained layers using small dataset of T2-w MR images. This mechanism uses

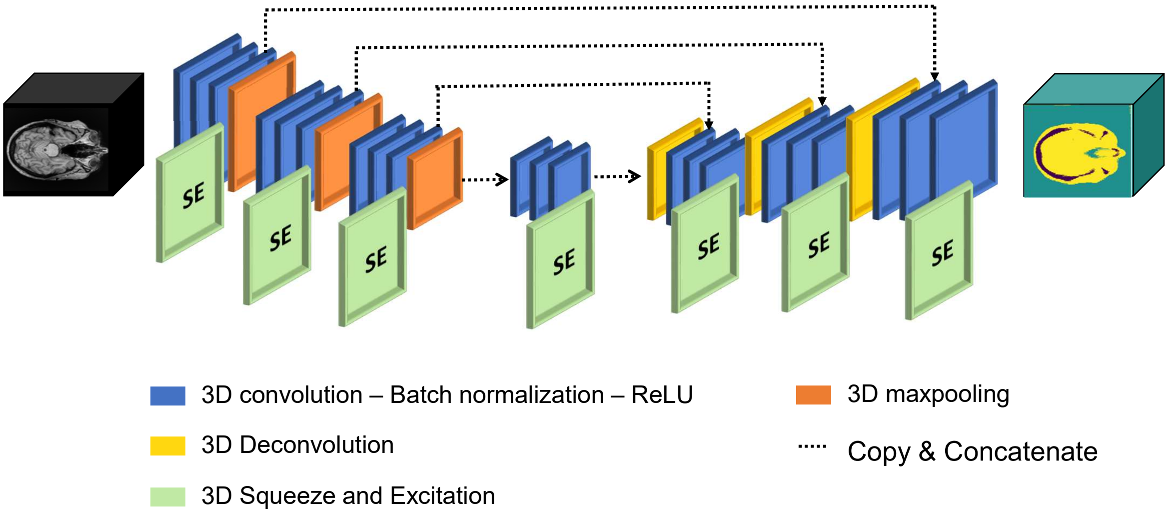


Figure 5.1: The architecture of the convolutional encoder decoder network with multiple squeeze and excitation blocks.

the pretrained network's weights as initialisation values and updates the weights of the whole network using the target datasets. The main drawback of this network is the necessity of big enough target dataset to retrain the network properly. This approach also requires long training time until the model reaches the convergence. Figure 5.2 illustrates the concept of transfer learning with open fine tuning.

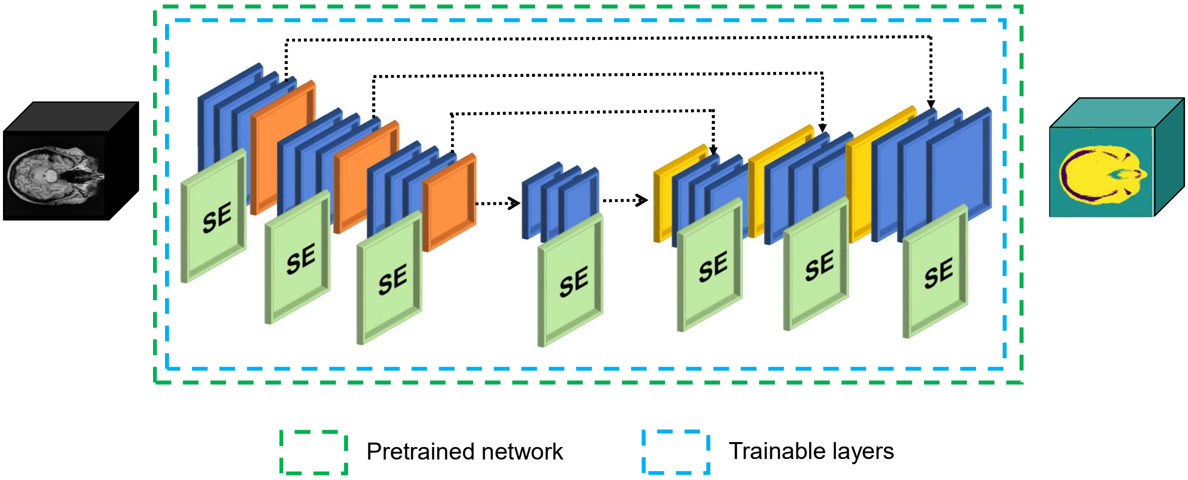


Figure 5.2: An illustration of the transfer learning concept using open fine tuning to segment T2-w MR images.

5.2.2 Transfer Learning using Conservative Fine Tuning

The conservative fine tuning is another type of transfer learning which entails using the weights of pretrained network and tuning them on specific number of the network's layers. For instance, the first few layers of the CNN extract very generic features which can be common for all types of images. These layers should be frozen during the conservative fine tuning process and only the last network's layers can be retrained to adapt the weights to the new domain images. The optimal number of network's layers that should be retrained and the size of the dataset which can be used to conduct the conservative fine tuning process are experimentally determined. Figure 5.3 shows a general illustration of the conservative fine tuning concept.

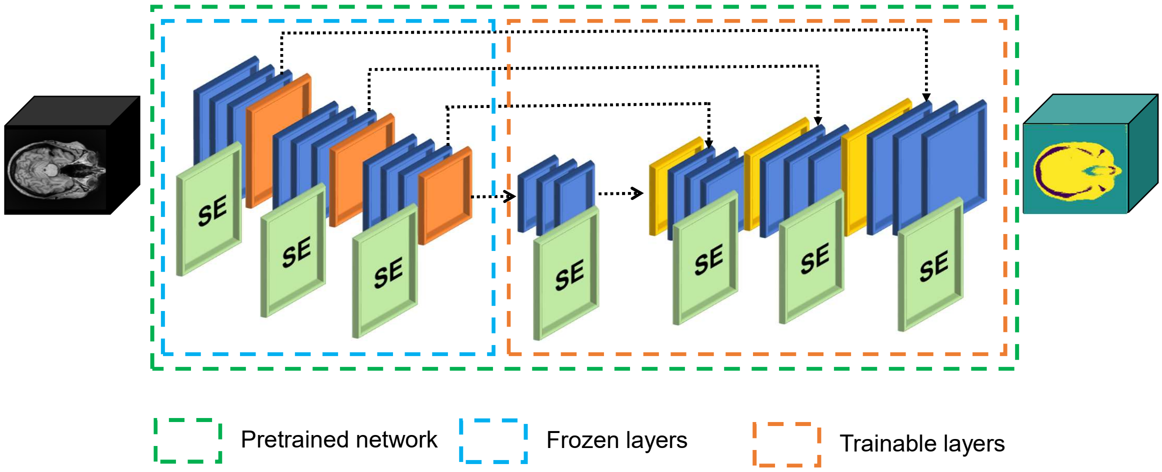


Figure 5.3: An illustration of the transfer learning concept using conservative fine tuning to segment T2-w MR images.

5.2.3 Hybrid Transfer Learning

The hybrid transfer learning consists of two parallel encoders (dedicated and adapted) which carry the features of both T1-w and T2-w MR images. The two set of features are aggregated before passing to a single decoder. The main architecture of the hybrid transfer learning network is demonstrated in Figure 5.4.

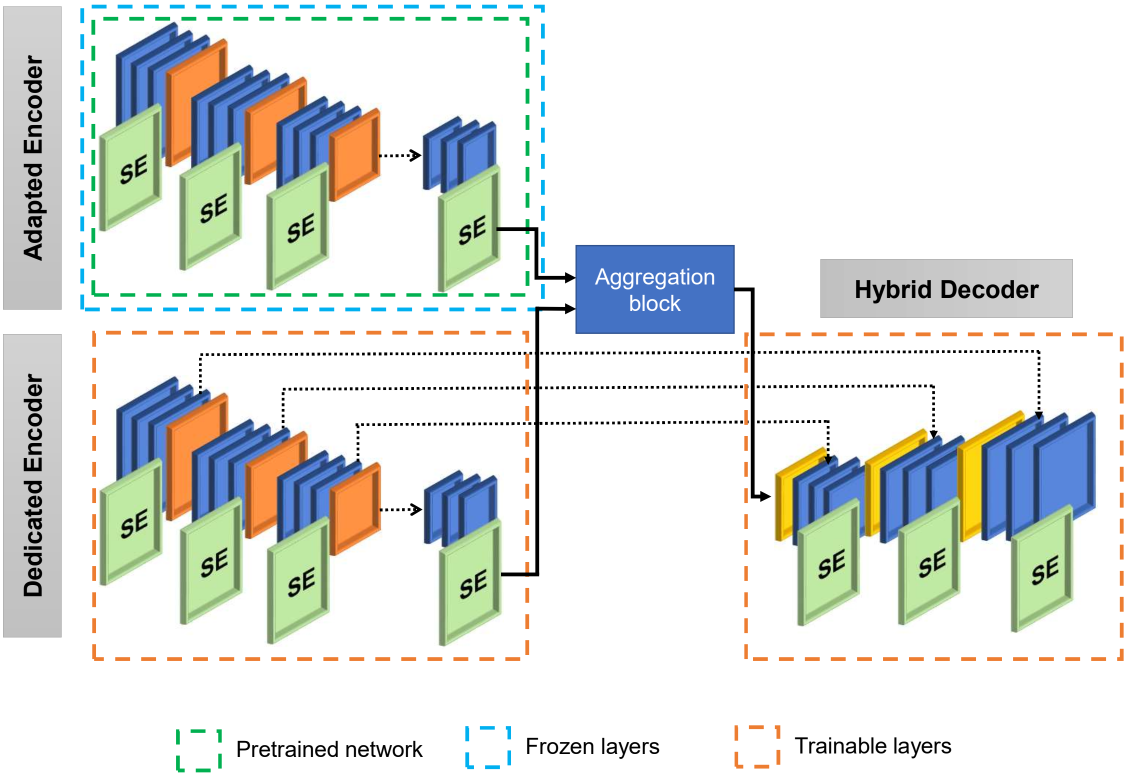


Figure 5.4: An illustration of the hybrid transfer learning to segment T2-w MR images.

A Dedicated Encoder

The dedicated encoder is specifically designed to extract T2-w features by training the encoder from scratch using only T2-w MR images as training datasets. The design of the dedicated encoder follows the design of the network's encoder shown in Figure 5.1 which consists of nine convolutional layers in the encoder path and three convolutional layers in the bottleneck path with four attention blocks placed after each three convolutional layers.

B Adapted Encoder

The adapted encoder is a pretrained encoder with T1-w images which generate features maps with smaller dimensions than the input image. All layers of this pretrained encoder are frozen to preserve the learnt features from T1-w images. The initial conducted experiments have shown that the segmentation accuracy of soft tissue and air classes of T2-w images using transfer learning without any fine tuning of the parameters is promising. This observation was inspiring to adapt the pretrained T1-w encoder without any fine tuning.

C Aggregation Module

The output of the dedicated encoder (T2-w features) and the adapted encoder (T1-w features) are aggregated using a fusion block. There are different possible mechanisms of aggregation such as concatenation, multiplication, maximum, and averaging which are tested experimentally. The segmentation results using different aggregation blocks are recorded in the experiments section. The aggregated features are passed to a decoder which retrieves the original size of the input images using nine convolutional layers, three unpooling layers, and three attention blocks.

5.2.4 Multi-Modality based Augmentation

The main difference between the most conventional MR sequences (T1-w and T2-w images). The main difference between these two sequences is the appearance of the tissue which is called the tissue intensity as illustrated in Figure 5.5.

This variation in the intensity values between two MR sequences can be considered as an intensity based augmentation technique. This idea is implemented by training the CNN with both sequences of T1-w and T2-w images. This augmentation technique firstly helps to increase the size of the training datasets and secondly aims at building a CNN which is able to segment multi MR sequences.

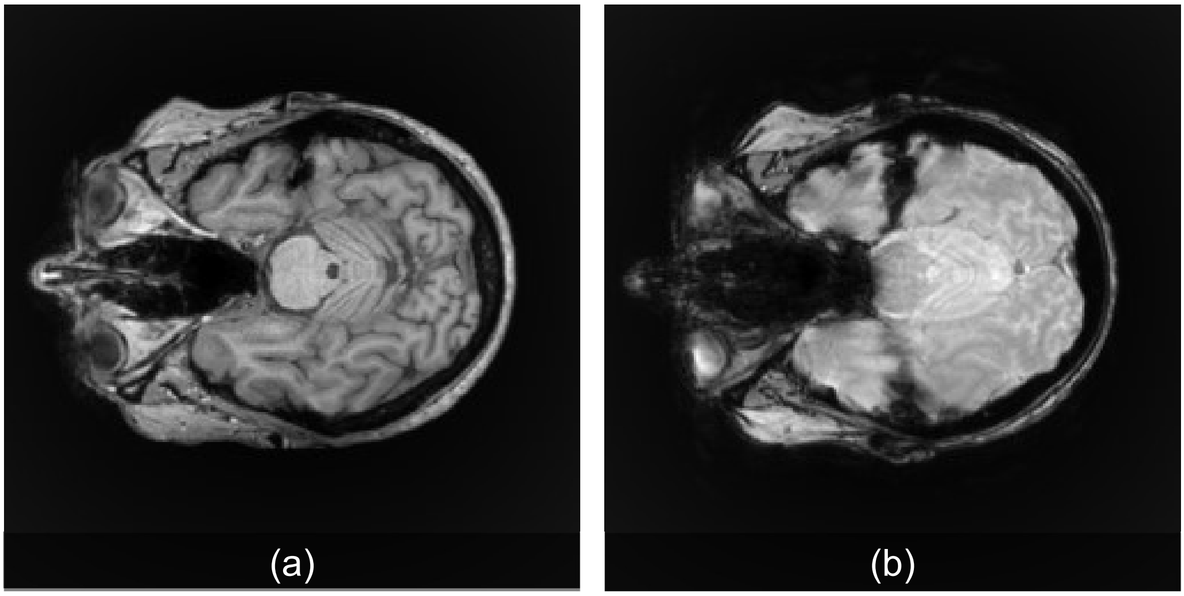


Figure 5.5: (a) T1-w MR image, (b) T2-w MR image.

5.3 Experimental Setup

5.3.1 Datasets

The experiments of this chapter have been conducted using the same fifty T1-w MR and CT images which are described in chapter 3. In addition to that, another dataset of T2-w MR images which consists of fourteen patients is used to evaluate the different transfer learning approaches. T1-w and T2-w datasets consist of fourteen common patients.

5.3.2 Model Input

The CNN's input images are generated using the same preprocessing steps as described in Chapters 3 and 4. Each volume's dimensions are $256 \times 256 \times 48$ where 48 is the number of selected MR slices. Each volume is divided into overlapped patches with size $128 \times 128 \times 48$ to reduce the input dimensions and increase the size of the training datasets.

5.3.3 Training Setup

The fine tuning process is conducted using three folds cross validation with grid search mechanism to find the optimal learning rate. For the hybrid transfer learning where the dedicated decoder is trained from scratch using T2-w images, the datasets are split as follows: training: 60%, validation: 20%, and testing: 20%. The layers weights are initialised using Glorot Uniform initialisation scheme and updated using Adam optimizer with a batch size of two and a moving learning rate that starts with 0.0001 then reduces by a factor of 0.75 on plateau mode when the training accuracy is not increasing for five continuous epochs. The maximum number of epochs is two hundred; Yet the training can be stopped before reaching two hundred epochs when the validation loss stops improving. The same training setup is applied to conduct the T2-w based augmentation experiments where the network is trained with both T1-w and T2-w MR sequences using the size of datasets as described in Table 5.1. All experiments are conducted on Tesla V100 GPU with 16 GB RAM.

Table 5.1: The size of training, validation, and testing datasets of T1-w and T2-w MR sequences.

Data group	T1-w	T2-w
Training data	30 patients	8 patients
Validation data	10 patients	3 patients
Testing data	10 patients	3 patients

5.3.4 Evaluation Metrics

The segmentation results are evaluated using a set of metrics by comparing the segmented MR image with the ground truth which are generated from CT images. The evaluation metrics are: precision, recall, DSC, and JSC. The definition of each metric is represented in Chapter 3.

5.4 Experimental Results

5.4.1 Transfer Learning using Open Fine Tuning

The transfer learning with open fine tuning experiment shows the effect of transfer learning by initialising the weights of the CNN using the weights of the pretrained CNN with T1-w images. The weights are updated by retraining all layers of the CNN with T2-w images. Tables 5.2, 5.3, and 5.4 show the comparison between the segmentation results of T2-w images without any fine tuning and three folds cross validation open fine tuning for bone, soft tissue, and air classes, respectively. The low performance of bone class segmentation which is generated by using a pretrained model without fine tuning indicates the inability of the model to segment the bone of two different MR sequences (T1-w and T2-w). On the other hand, the segmentation results of soft tissue and air classes using the transfer learning without fine tuning show better performance than open fine tuning. The open fine tuning approach requires the update of CNN's weights using a limited number of training dataset of T2-w MR images. Although the CNN weights are initialised with the pretrained model's weights, the process of open fine tuning needs more T2-w datasets to achieve promising segmentation results of the three brain tissue classes.

Table 5.2: The evaluation of the segmentation results of the bone class by applying transfer learning without fine tuning and open fine tuning using four different evaluation metrics: precision, recall, DSC, and JSC.

Bone tissue				
Model	PRE	REC	DSC	JSC
Transfer learning without fine tuning	0.5223	0.3038	0.3841	0.2377
Open fine tuning (fold=1)	0.6913	0.4494	0.5447	0.3743
Open fine tuning (fold=2)	0.6837	0.4043	0.5081	0.3406
Open fine tuning (fold=3)	0.6980	0.3819	0.4937	0.3277
Open fine tuning (avg)	0.6910	0.4119	0.5155	0.3475

Table 5.3: The evaluation of the segmentation results of the soft tissue class by applying transfer learning without fine tuning and open fine tuning using four different evaluation metrics: precision, recall, DSC, and JSC.

Soft tissue				
Model	PRE	REC	DSC	JSC
Transfer learning without fine tuning	0.8445	0.8484	0.8464	0.7338
Open fine tuning (fold=1)	0.7615	0.8950	0.8229	0.6990
Open fine tuning (fold=2)	0.7597	0.8988	0.8234	0.6999
Open fine tuning (fold=3)	0.7460	0.8986	0.8152	0.6881
Open fine tuning (avg)	0.7558	0.8975	0.8205	0.6957

Table 5.4: The evaluation of the segmentation results of the air class by applying transfer learning without fine tuning and open fine tuning using four different evaluation metrics: precision, recall, DSC, and JSC.

Ai tissue				
Model	PRE	REC	DSC	JSC
Transfer learning without fine tuning	0.9184	0.9481	0.9330	0.8744
Open fine tuning (fold=1)	0.9292	0.8612	0.8939	0.8082
Open fine tuning (fold=2)	0.9295	0.8620	0.8945	0.8091
Open fine tuning (fold=3)	0.9275	0.8518	0.8880	0.7986
Open fine tuning (avg)	0.9287	0.8584	0.8922	0.8053

5.4.2 Transfer Learning using Conservative Fine Tuning

The study of the performance of conservative fine tuning approach requires conducting intensive experiments to find the relationship between the required size of the target dataset and the optimal number of CNN layers which need to be fine tuned. Figure 5.6 illustrates the DSC of the three classes segmentation by conducting several experiments. Each experiment involves fine tuning a different number of convolutional layers starting by retraining the last fully connected layer until the encoder's convolutional layers. In general, the performance of the bone class segmentation is increased by updating the weights of more convolutional layers. Yet, the segmentation of the air and soft tissue classes are slightly improved by retraining more convolutional layers. Different evaluation metrics using three folds cross validation are shown in Tables 5.5, 5.6, and 5.7 for bone, soft tissue, and air classes, respectively.

The results show that the segmentation of bone and soft tissue classes is enhanced as

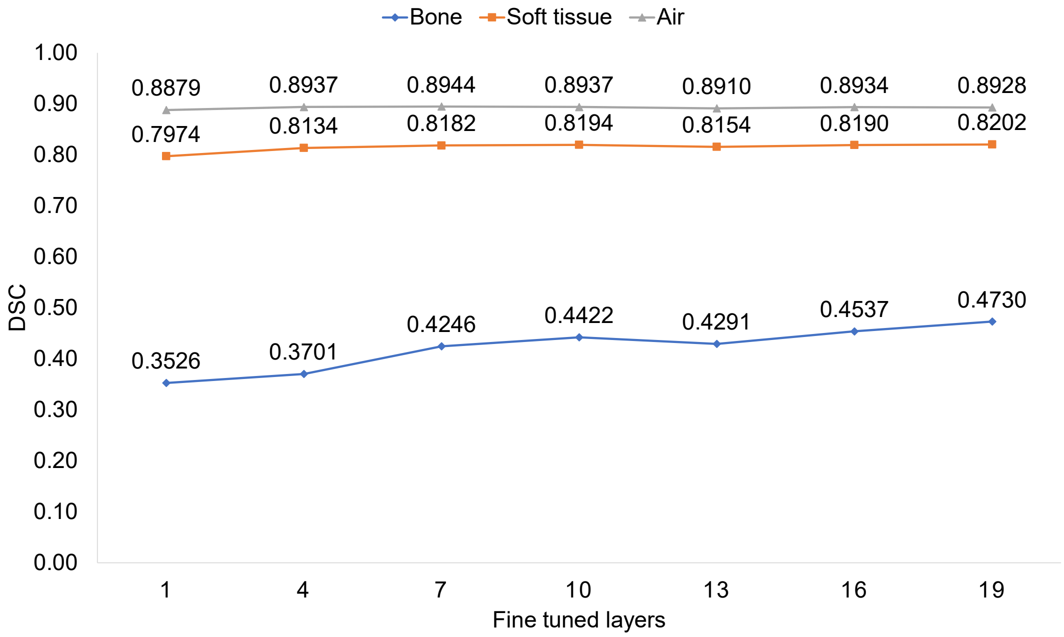


Figure 5.6: The DSC of three tissue classes using conservative fine tuning.

more convolutional layers are retrained and the layers parameters are updated. However, the segmentation of air class requires retraining only six convolutional layers to obtain the highest performance in terms of precision, DSC, and JSC. The highest score of air class sensitivity is achieved by retraining only the last fully connected layer.

The required size of T2-w MR images to train and fine tune the pretrained model is investigated by selecting different sizes of the training dataset and recording the DSC of the segmentation results of each size. The segmentation results are illustrated in Figures 5.7, 5.8, and 5.9 for the bone, soft tissue, and air classes, respectively. The results reveal that the more convolutional layers are fined tuned, the highest DSC is obtained with larger training datasets especially for bone and soft tissue segmentation. In the case of using small training datasets (e.g. two patients only), less convolutional layers are required to be fine tuned to avoid overfitting. The segmentation performance of fine tuning only the last fully connected layer to segment the three tissue classes is almost constant with different sizes of the training datasets. The highest DSC of the air class is achieved by fine tuning only three convolutional layers with a dataset of eight patients. The conducted experiments reveal that fine tuning the pretrained model with T1-w MR images requires the largest target dataset (T2-w MR) in order to transfer the knowledge

Table 5.5: The evaluation of the segmentation results of the bone class by applying conservative fine tuning using eight patients for training.

Bone tissue				
Model	PRE	REC	DSC	JSC
Fine tune FC (fold=1)	0.5105	0.2700	0.3532	0.2145
Fine tune FC (fold=2)	0.5114	0.2687	0.3523	0.2138
Fine tune FC (fold=3)	0.5115	0.2685	0.3522	0.2137
Fine tune FC (avg)	0.5111	0.2691	0.3526	0.2140
Fine tune 3 Conv (fold=1)	0.5642	0.3018	0.3932	0.3932
Fine tune 3 Conv (fold=2)	0.6315	0.2210	0.3274	0.1957
Fine tune 3 Conv (fold=3)	0.6395	0.2801	0.3896	0.2419
Fine tune 3 Conv (avg)	0.6117	0.2676	0.3701	0.2770
Fine tune 6 Conv (fold=1)	0.7063	0.2946	0.4158	0.2624
Fine tune 6 Conv (fold=2)	0.6762	0.2894	0.4053	0.2542
Fine tune 6 Conv (fold=3)	0.6702	0.3417	0.4526	0.2925
Fine tune 6 Conv (avg)	0.6842	0.3086	0.4246	0.2697
Fine tune 9 Conv (fold=1)	0.6687	0.4124	0.5102	0.3424
Fine tune 9 Conv (fold=2)	0.7107	0.2645	0.3855	0.2388
Fine tune 9 Conv (fold=3)	0.6786	0.3157	0.4309	0.2746
Fine tune 9 Conv (avg)	0.6860	0.3309	0.4422	0.2853
Fine tune 12 Conv (fold=1)	0.6135	0.3234	0.4235	0.2686
Fine tune 12 Conv (fold=2)	0.6027	0.3527	0.4450	0.2862
Fine tune 12 Conv (fold=3)	0.6218	0.3156	0.4187	0.2648
Fine tune 12 Conv (avg)	0.6127	0.3306	0.4291	0.2732
Fine tune 15 Conv (fold=1)	0.7147	0.3313	0.4527	0.2926
Fine tune 15 Conv (fold=2)	0.7066	0.3214	0.4418	0.2835
Fine tune 15 Conv (fold=3)	0.6657	0.3592	0.4667	0.3043
Fine tune 15 Conv (avg)	0.6957	0.3373	0.4537	0.2935
Fine tune 18 Conv (fold=1)	0.7343	0.3500	0.4740	0.3106
Fine tune 18 Conv (fold=2)	0.6864	0.3624	0.4743	0.3109
Fine tune 18 Conv (fold=3)	0.6956	0.3558	0.4708	0.3078
Fine tune 18 Conv (avg)	**0.7055**	**0.3560**	**0.4730**	**0.3098**

properly and avoid overfitting.

5.4.3 Hybrid Transfer Learning

The design of the hybrid transfer learning network consists of an aggregation component where several experiments are conducted to choose the best aggregation mechanism. The operations of concatenation, multiplication, averaging, and maximum are applied to integrate the two parallel encoders. The concatenation operation requires more memory allocation to train the network

Table 5.6: The evaluation of the segmentation results of the soft tissue class by applying conservative fine tuning using eight patients for training.

Soft tissue				
Model	PRE	REC	DSC	JSC
Fine tune FC (fold=1)	0.7411	0.8627	0.7973	0.6629
Fine tune FC (fold=2)	0.7409	0.8632	0.7974	0.6631
Fine tune FC (fold=3)	0.7409	0.8633	0.7974	0.6631
Fine tune FC (avg)	0.7410	0.8631	0.7974	0.6630
Fine tune 3 Conv (fold=1)	0.7433	0.8979	0.8133	0.6853
Fine tune 3 Conv (fold=2)	0.7335	0.9117	0.8129	0.6848
Fine tune 3 Conv (fold=3)	0.7424	0.9010	0.8140	0.6864
Fine tune 3 Conv (avg)	0.7397	0.9035	0.8134	0.6855
Fine tune 6 Conv (fold=1)	0.7434	0.9122	0.8192	0.6938
Fine tune 6 Conv (fold=2)	0.7401	0.9113	0.8169	0.6904
Fine tune 6 Conv (fold=3)	0.7478	0.9044	0.8187	0.6930
Fine tune 6 Conv (avg)	0.7438	**0.9093**	0.8182	0.6924
Fine tune 9 Conv (fold=1)	0.7559	0.8986	0.8211	0.6965
Fine tune 9 Conv (fold=2)	0.7393	0.9147	0.8177	0.6916
Fine tune 9 Conv (fold=3)	0.7463	0.9085	0.8195	0.6942
Fine tune 9 Conv (avg)	0.7472	0.9073	0.8194	0.6941
Fine tune 12 Conv (fold=1)	0.7492	0.8963	0.8162	0.6894
Fine tune 12 Conv (fold=2)	0.7516	0.8882	0.8142	0.6866
Fine tune 12 Conv (fold=3)	0.7480	0.8972	0.8158	0.6890
Fine tune 12 Conv (avg)	**0.7496**	0.8939	0.8154	0.6883
Fine tune 15 Conv (fold=1)	0.7488	0.9075	0.8205	0.6957
Fine tune 15 Conv (fold=2)	0.7455	0.9091	0.8192	0.6938
Fine tune 15 Conv (fold=3)	0.7507	0.8971	0.8174	0.6912
Fine tune 15 Conv (avg)	0.7483	0.9046	0.8190	0.6935
Fine tune 18 Conv (fold=1)	0.7504	0.9092	0.8222	0.6981
Fine tune 18 Conv (fold=2)	0.7485	0.9034	0.8187	0.6930
Fine tune 18 Conv (fold=3)	0.7494	0.9049	0.8199	0.6947
Fine tune 18 Conv (avg)	0.7494	0.9059	**0.8202**	**0.6953**

and the batch size of this experiment is set to one instead of two. The experiments of all other aggregation operations are conducted using batch size of two.

Three folds cross validation results of different aggregation mechanisms for the segmentation of bone, soft tissue, and air classes are shown in Tables 5.8, 5.9, and 5.10, respectively. The evaluation metrics of the bone class show that the multiplication of the features maps of each encoder generates higher segmentation performance. However, the concatenation mechanism provides better segmentation results of the soft tissue class. For the air class, all aggregation methods generate very similar performance with slightly higher results with maximum aggre-

Table 5.7: The evaluation of the segmentation results of the air tissue class by applying conservative fine tuning using eight patients for training.

Air tissue				
Model	PRE	REC	DSC	JSC
Fine tune FC (fold=1)	0.9141	0.8630	0.8878	0.7983
Fine tune FC (fold=2)	0.9143	0.8629	0.8879	0.7984
Fine tune FC (fold=3)	0.9144	0.8629	0.8879	0.7984
Fine tune FC (avg)	0.9143	**0.8629**	0.8879	0.7983
Fine tune 3 Conv (fold=1)	0.9331	0.8561	0.8929	0.8066
Fine tune 3 Conv (fold=2)	0.9366	0.8551	0.8940	0.8082
Fine tune 3 Conv (fold=3)	0.9318	0.8596	0.8943	0.8087
Fine tune 3 Conv (avg)	0.9338	0.8569	0.8937	0.8079
Fine tune 6 Conv (fold=1)	0.9370	0.8587	0.8961	0.8118
Fine tune 6 Conv (fold=2)	0.9361	0.8547	0.8936	0.8076
Fine tune 6 Conv (fold=3)	0.9331	0.8574	0.8937	0.8077
Fine tune 6 Conv (avg)	**0.9354**	0.8569	**0.8944**	**0.8090**
Fine tune 9 Conv (fold=1)	0.9316	0.8587	0.8937	0.8078
Fine tune 9 Conv (fold=2)	0.9368	0.8565	0.8949	0.8097
Fine tune 9 Conv (fold=3)	0.9325	0.8561	0.8926	0.8061
Fine tune 9 Conv (avg)	0.9336	0.8571	0.8937	0.8070
Fine tune 12 Conv (fold=1)	0.9278	0.8583	0.8917	0.8045
Fine tune 12 Conv (fold=2)	0.9243	0.8583	0.8901	0.8019
Fine tune 12 Conv (fold=3)	0.9273	0.8577	0.8911	0.8037
Fine tune 12 Conv (avg)	0.9264	0.8581	0.8910	0.8034
Fine tune 15 Conv (fold=1)	0.9333	0.8599	0.8951	0.8101
Fine tune 15 Conv (fold=2)	0.9327	0.8560	0.8927	0.8062
Fine tune 15 Conv (fold=3)	0.9287	0.8588	0.8924	0.8056
Fine tune 15 Conv (avg)	0.9316	0.8582	0.8934	0.8073
Fine tune 18 Conv (fold=1)	0.9339	0.8596	0.8952	0.8103
Fine tune 18 Conv (fold=2)	0.9302	0.8546	0.8908	0.8031
Fine tune 18 Conv (fold=3)	0.9313	0.8567	0.8924	0.8058
Fine tune 18 Conv (avg)	0.9318	0.8569	0.8928	0.8064

gation method. It is also worth mentioning that the precision of the concatenation based results of the bone and air classes is better than the multiplication mechanism.

A clear demonstration of the evaluation metrics of the hybrid transfer learning using different aggregation methods for the bone, soft tissue,and air classes is revealed in Figures 5.10, 5.11, and 5.12, respectively.

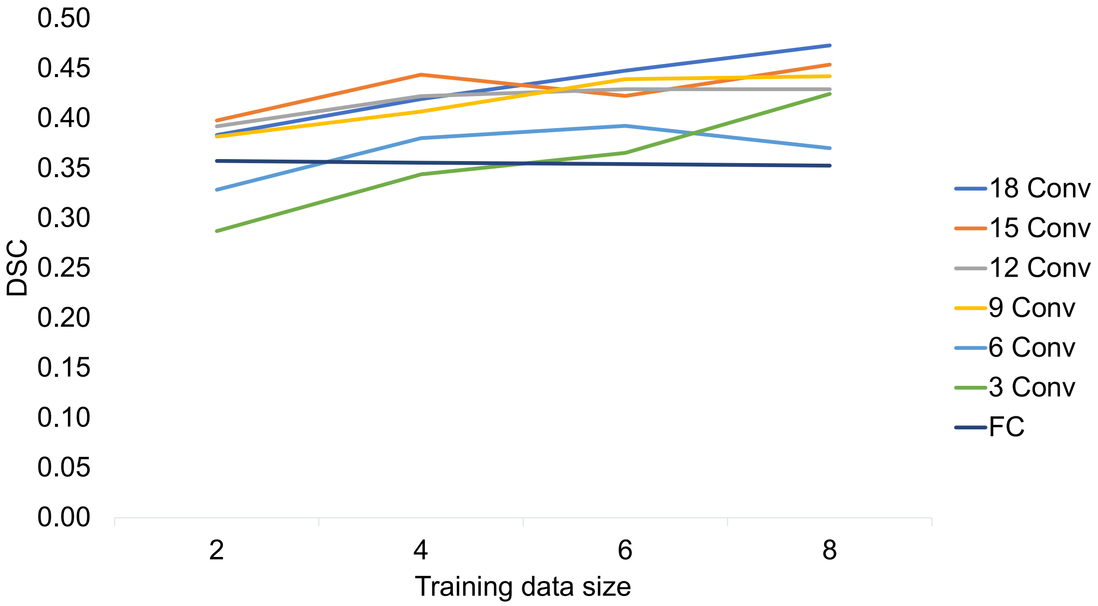


Figure 5.7: The DSC of the bone tissue segmentation using conservative fine tuning for different numbers of convolutional layers with different sizes of training datasets.

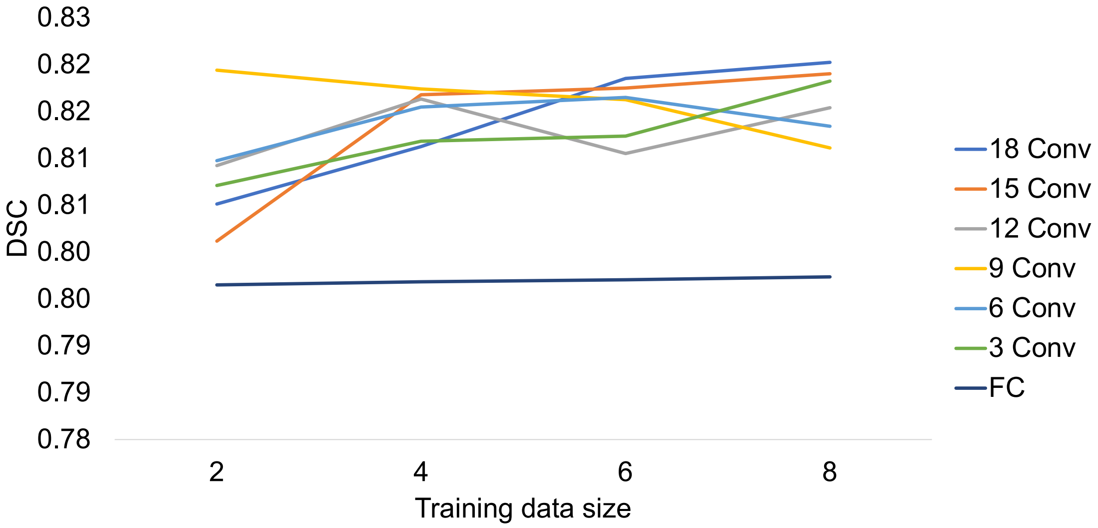


Figure 5.8: The DSC of the soft tissue segmentation using conservative fine tuning for different numbers of convolutional layers with different sizes of training datasets.

5.4.4 Methods Comparison

The comparison between the segmentation results using different transfer learning and fine tuning approaches is illustrated in Tables 5.11, 5.12, and 5.13 for bone, soft tissue, and air

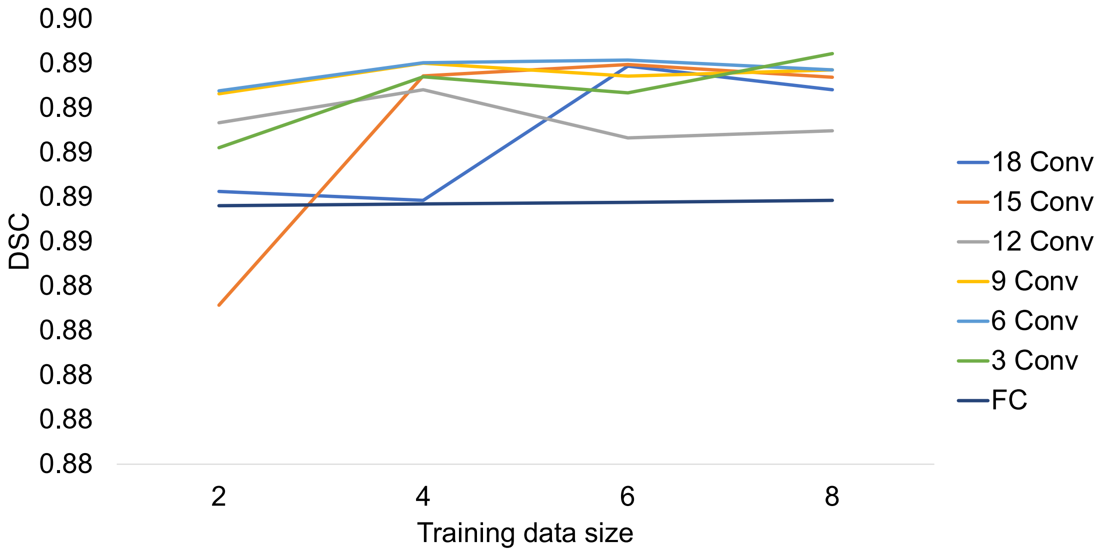


Figure 5.9: The DSC of the air tissue segmentation using conservative fine tuning for different numbers of convolutional layers with different sizes of training datasets.

Table 5.8: The evaluation of the segmentation results of the bone class by applying hybrid transfer learning with different aggregation mechanisms using four different evaluation metrics: precision, recall, DSC, and JSC with three folds cross validation.

Bone tissue				
Model	PRE	REC	DSC	JSC
Aggregation: concatenation (fold=1)	0.7057	0.4308	0.5350	0.3652
Aggregation: concatenation (fold=2)	0.6761	0.2122	0.3230	0.1926
Aggregation: concatenation (fold=3)	0.7026	0.3925	0.5036	0.3366
Aggregation: concatenation (avg)	**0.6948**	0.3451	0.4539	0.2981
Aggregation: multiplication (fold=1)	0.5471	0.6018	0.5731	0.4017
Aggregation: multiplication (fold=2)	0.3990	0.6340	0.4897	0.3243
Aggregation: multiplication(fold=3)	0.5420	0.5685	0.5549	0.3840
Aggregation: multiplication (avg)	0.4960	**0.6014**	**0.5393**	**0.3700**
Aggregation: average (fold=1)	0.6769	0.4945	0.5715	0.4001
Aggregation: average (fold=2)	0.5972	0.4784	0.5312	0.3617
Aggregation: average (fold=3)	0.6804	0.3997	0.5036	0.3366
Aggregation: average (avg)	0.6515	0.4576	0.5355	0.3661
Aggregation: maximum (fold=1)	0.6723	0.4906	0.5673	0.3959
Aggregation: maximum (fold=2)	0.6284	0.4087	0.4953	0.3292
Aggregation: maximum (fold=3)	0.6793	0.4146	0.5149	0.3467
Aggregation: maximum (avg)	0.6600	0.4380	0.5258	0.3573

classes, respectively.

Table 5.9: The evaluation of the segmentation results of the soft tissue class by applying hybrid transfer learning with different aggregation mechanisms using four different evaluation metrics: precision, recall, DSC, and JSC with three folds cross validation.

Soft tissue				
Model	PRE	REC	DSC	JSC
Aggregation: concatenation (fold=1)	0.7423	0.8965	0.8122	0.6837
Aggregation: concatenation (fold=2)	0.7346	0.9042	0.9042	0.6816
Aggregation: concatenation (fold=3)	0.7335	0.8995	0.8081	0.6780
Aggregation: concatenation (avg)	0.7368	**0.9001**	**0.8415**	0.6811
Aggregation: multiplication (fold=1)	0.7850	0.8254	0.8047	0.6732
Aggregation: multiplication (fold=2)	0.7855	0.8006	0.7930	0.6570
Aggregation: multiplication(fold=3)	0.7794	0.8239	0.8010	0.6681
Aggregation: multiplication (avg)	**0.7833**	0.8166	0.7996	0.6661
Aggregation: average (fold=1)	0.7637	0.8912	0.8225	0.6986
Aggregation: average (fold=2)	0.7672	0.8459	0.8047	0.6732
Aggregation: average (fold=3)	0.7434	0.8973	0.8131	0.6851
Aggregation: average (avg)	0.7581	0.8781	0.8134	0.6856
Aggregation: maximum (fold=1)	0.7553	0.8904	0.8173	0.6911
Aggregation: maximum (fold=2)	0.7550	0.8773	0.8116	0.6829
Aggregation: maximum (fold=3)	0.7461	0.8998	0.8158	0.6889
Aggregation: maximum (avg)	0.7521	0.8892	0.8149	**0.6876**

Table 5.10: The evaluation of the segmentation results of the air class by applying hybrid transfer learning with different aggregation mechanisms using four different evaluation metrics: precision, recall, DSC, and JSC with three folds cross validation.

Air tissue				
Model	PRE	REC	DSC	JSC
Aggregation: concatenation (fold=1)	0.9259	0.8436	0.8828	0.7902
Aggregation: concatenation (fold=2)	0.9301	0.8587	0.8930	0.8067
Aggregation: concatenation (fold=3)	0.9268	0.8384	0.8804	0.7863
Aggregation: concatenation (avg)	**0.9276**	0.8469	0.8854	0.7944
Aggregation: multiplication (fold=1)	0.9065	0.8710	0.8884	0.7992
Aggregation: multiplication (fold=2)	0.9058	0.8489	0.8764	0.7800
Aggregation: multiplication(fold=3)	0.9035	0.8691	0.8860	0.7953
Aggregation: multiplication (avg)	0.9053	**0.8630**	0.8836	0.7915
Aggregation: average (fold=1)	0.9293	0.8594	0.8930	0.8066
Aggregation: average (fold=2)	0.9071	0.8678	0.8870	0.7969
Aggregation: average (fold=3)	0.9271	0.8469	0.8852	0.7941
Aggregation: average (avg)	0.9212	0.8580	0.8884	0.7992
Aggregation: maximum (fold=1)	0.9275	0.8512	0.8877	0.7981
Aggregation: maximum (fold=2)	0.9234	0.8630	0.8922	0.8053
Aggregation: maximum (fold=3)	0.9293	0.8475	0.8865	0.7962
Aggregation: maximum (avg)	0.9267	0.8539	**0.8888**	**0.7999**

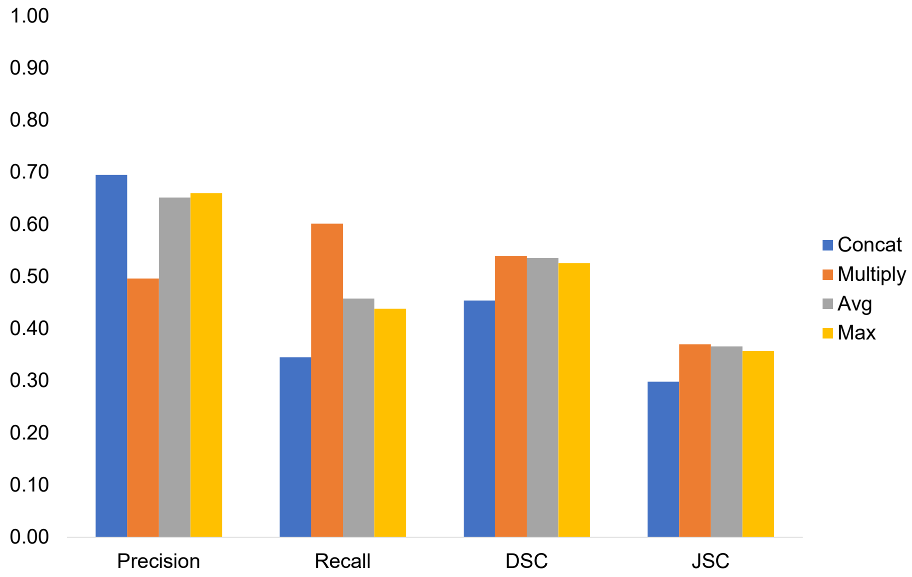


Figure 5.10: The evaluation metrics of the segmentation of bone class of T2-w MR images using different aggregation mechanisms.

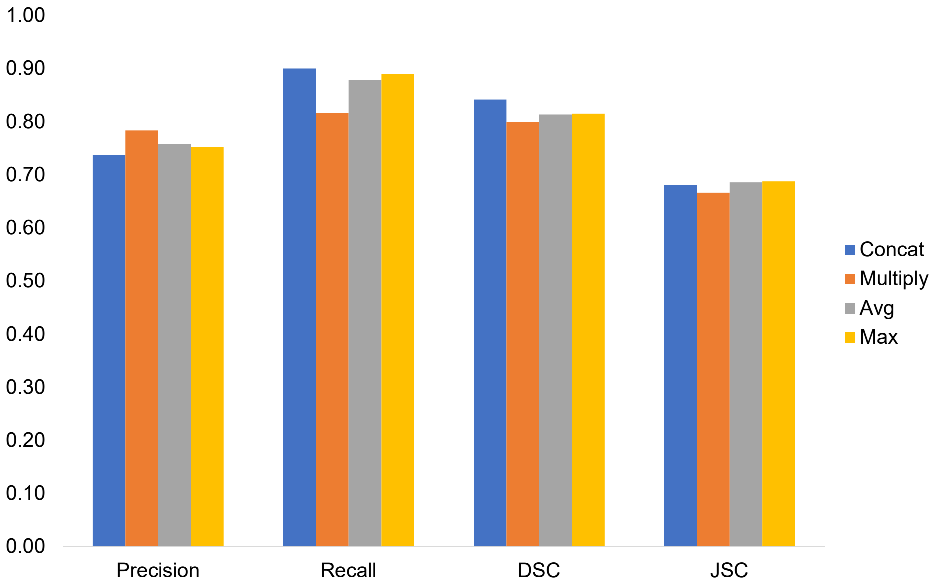


Figure 5.11: The evaluation metrics of the segmentation of soft tissue class of T2-w MR images using different aggregation mechanisms.

The conservative fine tuning approach which requires retraining eighteen convolutional layers using a dataset of eight patients generates the most precise bone segmentation compared to other approaches. Yet, the hybrid fine tuning with the application of multiplication aggregation operation outperforms the other approaches in terms of sensitivity (recall), DSC, and JSC of

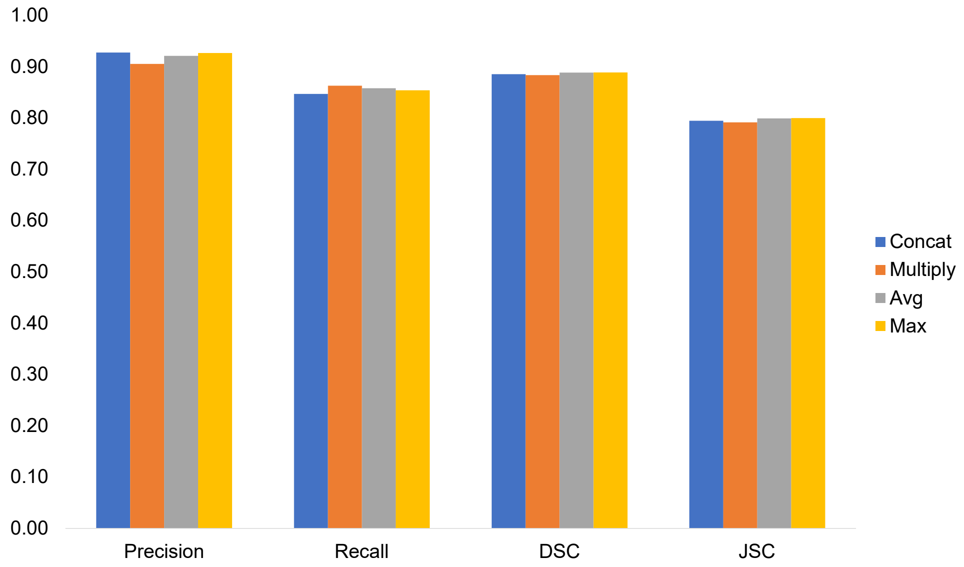


Figure 5.12: The evaluation metrics of the segmentation of air class of T2-w MR images using different aggregation mechanisms.

the bone class. On the other hand, the segmentation of the air and soft tissue classes using the pretrained model of T1-w MR images without fine tuning is performing better than any fine tuning approaches. The main advantage of the application of hybrid transfer learning and conservative fine tuning to segment the soft tissue and air classes is the significant enhancement of the soft tissue sensitivity and the precision of the air class.

Figure 5.13 illustrates the segmentation results of three randomly selected 2D slices from the testing datasets of T2-w MR images. The segmentation results of the slice shown in the first row indicates the ability of the pretrained model without any fine tuning (column c) to segment the air cavities more accurately than the hybrid transfer learning (column h). Yet, the other approaches (d - g) are not able to segment these regions of air. The visual comparison of the different segmentation approaches also shows that the transfer learning without fine tuning or with open or conservative fine tuning are generating many false positive pixels of the bone class as shown in the second row. The hybrid transfer learning using the multiplication aggregation mechanism is able to segment the bone class more precisely and accurately as illustrated on the slices of the second and third rows.

Table 5.11: The comparison between the segmentation results of the bone class by applying different approaches of transfer learning using four different evaluation metrics: precision, recall, DSC, and JSC.

Bone tissue				
Model	PRE	REC	DSC	JSC
Transfer learning without fine tuning	0.5223	0.3038	0.3841	0.2377
Open fine tuning	0.6910	0.4119	0.5155	0.3475
Conservative fine tuning (18 Conv)	**0.7055**	0.3560	0.4730	0.3098
Hybrid transfer learning (mult)	0.4960	**0.6014**	**0.5393**	**0.3700**

Table 5.12: The comparison between the segmentation results of the soft tissue class by applying different approaches of transfer learning using four different evaluation metrics: precision, recall, DSC, and JSC.

Soft tissue				
Model	PRE	REC	DSC	JSC
Transfer learning without fine tuning	**0.8445**	0.8484	**0.8464**	**0.7338**
Open fine tuning	0.7558	0.8975	0.8205	0.6957
Conservative fine tuning (18 Conv)	0.7494	**0.9059**	0.8202	0.6953
Hybrid transfer learning (concat)	0.7368	0.9001	0.8415	0.6811

Table 5.13: The comparison between the segmentation results of the air class by applying different approaches of transfer learning using four different evaluation metrics: precision, recall, DSC, and JSC.

Air tissue				
Model	PRE	REC	DSC	JSC
Transfer learning without fine tuning	0.9184	**0.9481**	**0.9330**	**0.8744**
Open fine tuning	0.9287	0.8584	0.8922	0.8053
Conservative fine tuning (6 Conv)	**0.9354**	0.8569	0.8944	0.8090
Hybrid transfer learning (max)	0.9267	0.8599	0.8888	0.7999

5.4.5 Computation Time

The computation time is an important factor while applying the transfer learning. Figure 5.14 shows the required computation time per epoch to retrain different number of convolutional layers. It is observed that fine tuning twelve or nine convolutional layers requires the same amount of computation time. Figure 5.15 represents the required number of epochs until the model reaches convergence. Although the fine tuning of nine and twelve convolutional layers requires the same amount of time to train one epoch, they differ in the number of required epochs to achieve the model convergence. The lowest number of epochs is twenty one which is recorded

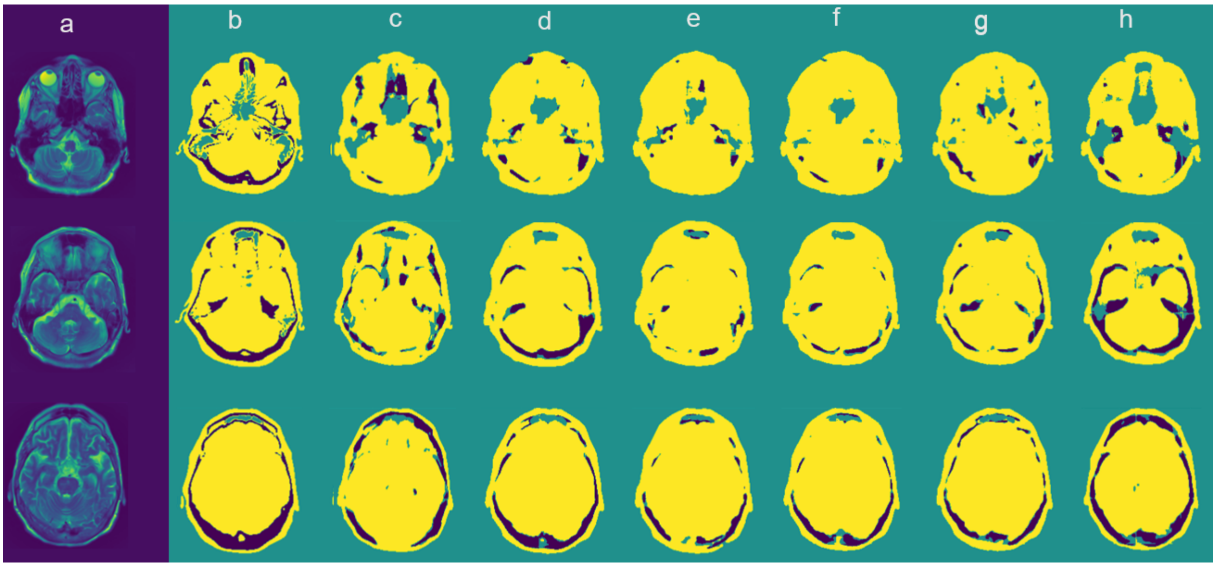


Figure 5.13: The segmentation results of three randomly selected slices from the testing datasets of T2-w MR images. (a) MR images, (b) CT images which used as ground truth, (c) the segmentation results using transfer learning without fine tuning, (d) the segmentation results using open fine tuning, (e) the segmentation results using conservative fine tuning (6 Conv), (f) the segmentation results using conservative fine tuning (18 Conv), (g) the segmentation results using hybrid transfer learning (multiplication aggregation), and (h) the segmentation results using hybrid transfer learning (concatenation aggregation).

by fine tuning nine convolutional layers. Figure 5.16 shows the total computation time of fine tuning different number of convolutional layers. The total computation time is the computation time per epoch multiplied by the required number of epochs to train the model until the model's convergence. The values shown in the figures are the average of the three folds cross validation results.

Figures 5.17, 5.18, and 5.19 illustrate the computation time per epoch, the number of required epochs to train and fine tune the model, and the total computation time for different fine tuning approaches, respectively. Although, the hybrid transfer learning requires longer time to execute one epoch, it requires few epochs to converge and complete the training process. Overall, the hybrid transfer learning requires less amount of time to complete the training followed by the conservative fine tuning by retraining eighteen convolutional layers then the open fine tuning that requires to update the weights of all layers of the CNN. The recorded computation time refers to using eight patients for training the model.

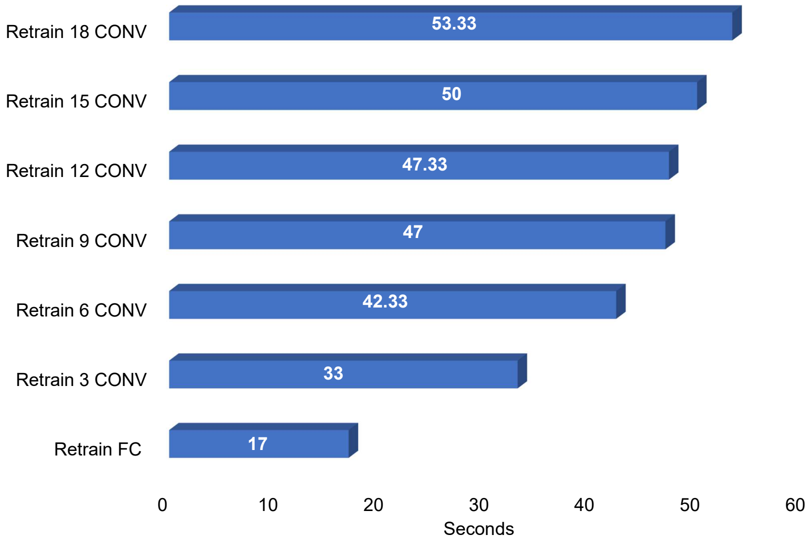


Figure 5.14: The computation time of a single epoch of the conservative fine tuning approach by retraining different number of CNN's layers.

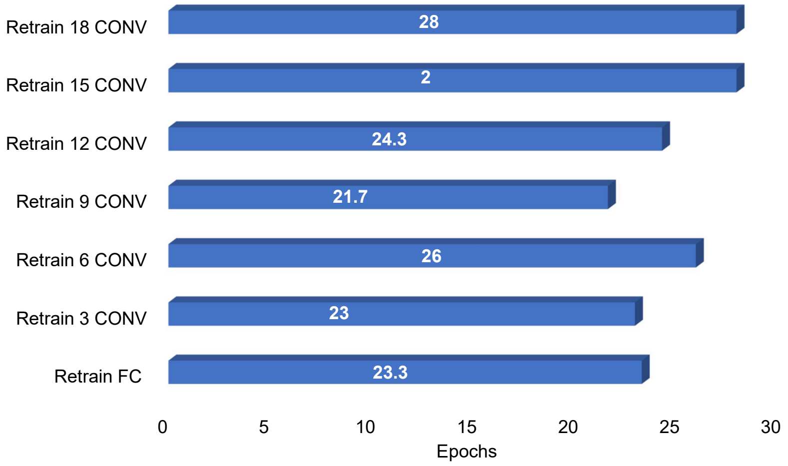


Figure 5.15: The required number of epochs to apply the conservative fine tuning approach by retraining different number of CNN's layers.

5.4.6 T2-w based Augmentation

The segmentation results of the application of intensity based augmentation by using the T2-w MR sequences as augmented images are illustrated in Tables 5.14, 5.15 and 5.16 for the bone, soft tissue, and air classes, respectively. The conducted experiments involve testing the model with T1-w and T2-w MR images to investigate the capability of the model to segment two types of MR sequences and the ability of the model to preserve the T1-w features after the addition of

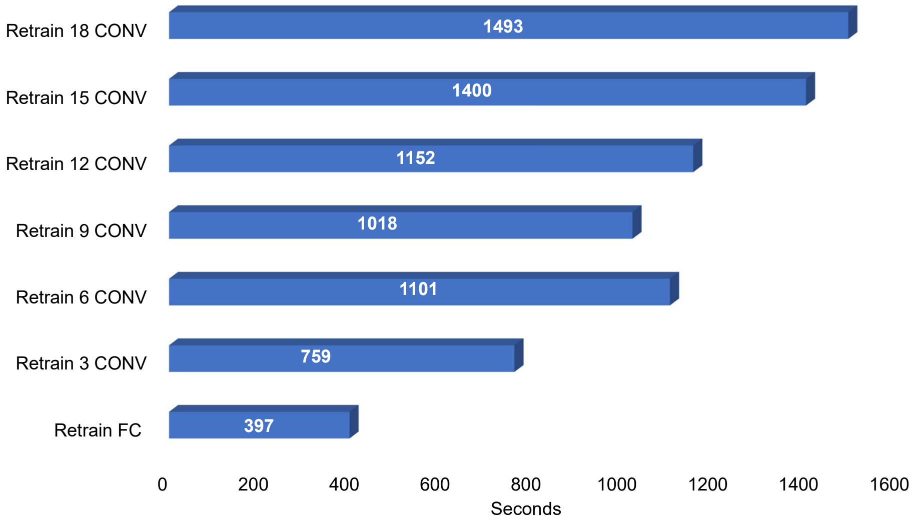


Figure 5.16: The total computation time of the conservative fine tuning approach by retraining different numbers of CNN's layers.

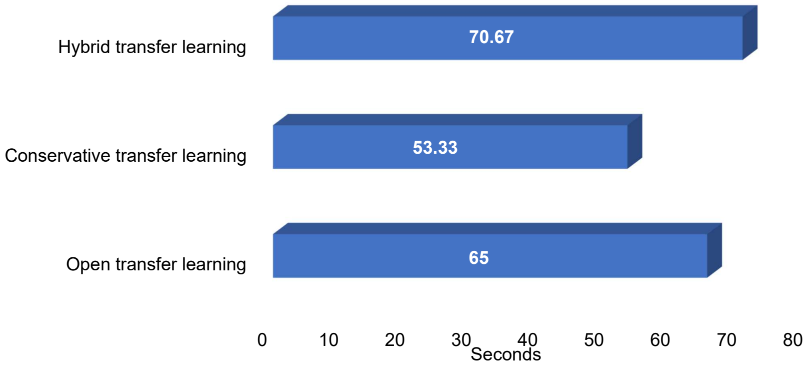


Figure 5.17: The computation time of a single epoch of different transfer learning approaches.

T2-w images.

Firstly, the model is trained with only T1-w images without the application of augmentation and tested to segment T1-w and T2-w MR images. The evaluation metrics of the three classes show that the model is not able to segment T2-w images properly with low DSC for the bone class (0.3841).

Secondly, the model is trained with thirty T1-w and eight T2-w MR images and tested to segment T1-w and T2-w MR images. The percentage increase of the segmentation results

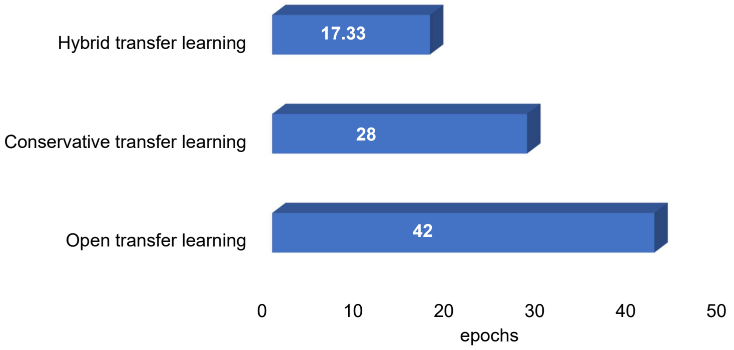


Figure 5.18: The required number of epochs to train different transfer learning approaches.

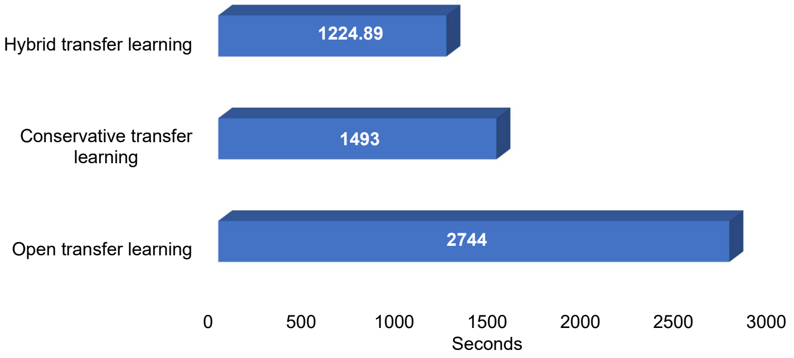


Figure 5.19: The total computation time of different transfer learning approaches.

of the T2-w bone class is high. This enhancement indicates the ability of the augmentation technique to include T2-w features which in turn enhance the segmentation results of T2-w images and build a multi-modality segmentation model. On the other hand, the T2-w based augmentation technique is not affecting the segmentation results of T1-w images. The lack of segmentation improvement indicates that the model is able to preserve the features of T1-w images even after training the model with two different MR sequences. The sensitivity of the soft tissue class remains stable while the precision of the air class is improved by 0.03% after the application of T2-w based augmentation.

Figure 5.20 illustrates the visual segmentation results of two randomly selected slices from the testing datasets of T1-w and T2-w MR images. The comparison between the segmen-

Table 5.14: The segmentation results of T1-w and T2-w bone class by applying T2-w based augmentation.

Bone tissue				
Model	PRE	REC	DSC	JSC
T1-w without augmentation	0.7261	0.5857	0.6484	0.4797
T1-w with augmentation	0.7180	0.5639	0.6317	0.4616
Percentage increase	-1.12%	-3.71%	-2.57%	-3.76%
T2-w without augmentation	0.5223	0.3038	0.3841	0.2377
T2-w with augmentation	0.7150	0.4338	0.5399	0.3698
Percentage increase %	36.90%	42.78%	40.56%	55.56%

Table 5.15: The segmentation results of T1-w and T2-w soft tissue class by applying T2-w based augmentation.

Soft tissue				
Model	PRE	REC	DSC	JSC
T1-w without augmentation	0.8962	0.9290	0.9123	0.8388
T1-w with augmentation	0.8917	0.9290	0.9100	0.8349
Percentage increase	-0.50%	0.00%	-0.25%	-0.47%
T2-w without augmentation	0.8445	0.8484	0.8464	0.7338
T2-w with augmentation	0.8595	0.8786	0.8690	0.7683
Percentage increase	1.78%	3.56%	2.66%	4.70%

Table 5.16: The segmentation results of T1-w and T2-w air class by applying T2-w based augmentation.

Air tissue				
Model	PRE	REC	DSC	JSC
T1-w without augmentation	0.9641	0.9626	0.9633	0.9293
T1-w with augmentation	0.9644	0.9618	0.9631	0.9288
Percentage increase	0.03%	-0.08%	-0.02%	-0.05%
T2-w without augmentation	0.9184	0.9481	0.9330	0.8744
T2-w with augmentation	0.9279	0.9460	0.9369	0.8813
Percentage increase	1.04%	-0.22%	0.42%	0.79%

tation results of T1-w MR images before and after the application of augmentation do not reveal a significant enhancement for the bone class. However, the segmentation of the air cavities before the augmentation is more accurate. Nevertheless, the segmentation of the T2-w images is highly improved after the T2-w based augmentation where many false positive of the bone and air classes are disappeared.

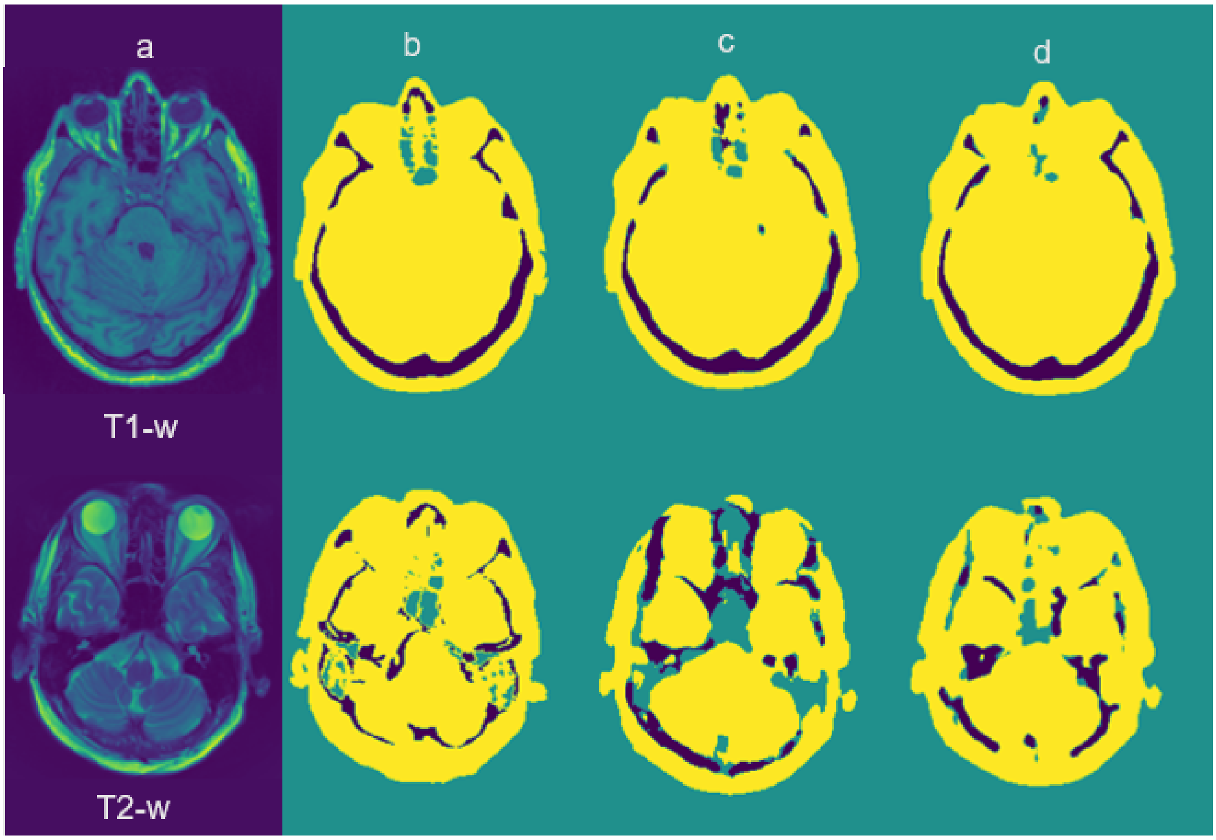


Figure 5.20: The segmentation results of two randomly selected slices from the testing datasets of T1-w and T2-w MR images. (a) MR images, (b) CT images which used as ground truth, (c) the segmentation results without the application of augmentation, and (d) the segmentation results using T2-w based augmentation.

5.5 Summary and Limitations

The application of transfer learning from T1-w to T2-w MR sequences has been studied in this chapter by exploring different mechanisms of domain adaptation. The knowledge from a pretrained model with T1-w MR images is transferred to segment T2-w images using four approaches which are: transfer learning without fine tuning, transfer learning using open fine tuning, transfer learning using conservative fine tuning, and hybrid transfer learning. The hybrid transfer learning solves the issue of overfitting when fine tuning the CNN with small dataset. This approach also enhances the segmentation of the bone class in T2-w MR images. The segmentation results of the soft tissue and air classes of T2-w MR images using the transfer learning from T1-w MR images without any fine tuning show the best performance with minimal computation time. The experiments of conservative fine tuning reveal that retraining only six convolutional layers is able to produce better segmentation results for air classes with acceptable computation time. On the other hand, the segmentation results of bone and soft tissue classes reveal better performance with fine tuning eighteen convolutional layers which requires more computation time.

The technique of T2-w based augmentation has not shown a significant impact on the performance of T1-w MR segmentation. However, this augmentation technique helps to improve the T2-w MR segmentation and build a mulitmodality or mulitsequences model which is able to segment different MR sequences. Generally, the performance of the segmentation of T1-w MR images is better than T2-w MR images because the size of T1-w dataset is larger than T2-w dataset. The increase of the size of the target training datasets helps the conservative fine tuning of CNN.

In summary, the hybrid transfer learning is useful to improve the bone segmentation. However, the segmentation of the soft tissue and air classes of T2-w images is performing well using the pretrained model without fine tuning and with minimal computation time and resources.

Chapter 6

Conclusions and Future Directions

6.1 Introduction

The PET attenuation correction is one challenging task for PET quantification and analysis. Recently, with the wide use of the hybrid PET/MR scanners, MR images with their various advantages are used to create pseudo CT to replace CT images. The generation of pseudo CT images is an essential step of the integrated process of MR based segmentation for PET attenuation correction.

The main challenge of the segmentation of brain MR images is the delineation of bone tissue especially in the nasal cavity areas. One key reason is the similarity of air and bone intensity values which makes the distinguish between these two tissues very challenging. Another reason related to machine learning modelling is the issue of classes unbalancing in the brain volume. The number of air and soft tissue voxels is much higher than the bone voxels. The classes unbalancing tends to build biased classification models.

This research has proposed multiple approaches to enhance the segmentation of brain MR images by investigating different mechanisms including CNN, attention based CNN, multiresolution analysis techniques with CNN, and transfer learning with different fine tuning mechanisms. The proposed approaches have shown their efficiency in enhancing the segmentation of the brain into three tissue classes (bone, soft tissue, air) compared to baseline approaches. Additionally, the proposed segmentation methods have been evaluated on clinical MR and CT images and validated with other similar methods in the literature. Moreover, this research has shown the applicability of transfer learning from different MR sequences. The objectives of this research have been achieved successfully and will be discussed in the next sections.

This chapter includes a general conclusion of the proposed contributions in this research work. The defined research objectives which are listed in Chapter 1 are revisited in Section 3 to ensure they have been addressed adequately. Section 4 illustrates the limitations and the future directions are suggested in Section 5.

6.2 General Conclusions

This research has investigated three directions of deep neural networks to enhance the segmentation of the bone class in brain MR images.

The first direction includes the combination of handcrafted features with CNN's features to enrich the set of features used for voxel's classification. The explored multiresolution handcrafted features are Shearlet and Contourlet transforms which are able to add the multiscale and multidirectional properties to the set of features. This type of multiresolution features has shown its capability to enhance the bone segmentation in brain MR images compared to other techniques proposed in the literature.

Although the results of the fusion of multiresolution features with CNN's features are promising, the task of handcrafted features extraction adds some complexity to the segmentation algorithm. Hence, the second direction of this research has proposed another approach by designing an attention based CNN that boosts the features during the features extraction process without the need to any additional handcrafted features. The attention based CNN uses a squeeze and excitation block that recalibrates the features adaptively on either channel or spatial domain. The two excitation blocks are integrated sequentially to reduce the model complexity. This approach has shown slightly better segmentation performance than the first proposed approach; Yet it has reduced the model complexity by eliminating the task of handcrafted features extraction and selection. Another enhancement that has been proposed under this direction is the application of an ensemble of attention based CNNs with the deployment of different activation and objective functions. The segmentation results of the ensemble of attention based CNNs outperform the results of a single attention based CNN.

The third direction of this research is the application of transfer learning from different MR sequences. The proposed method have applied different fine tuning mechanisms to transfer the knowledge from T1-w to T2-w MR sequences. The proposed mechanisms are transfer learning using open fine tuning, transfer learning using conservative fine tuning, and hybrid transfer learning. The results reveal that the domain adaptation is efficient with the particular application of brain MR segmentation. Moreover, the T2-w images have been used as an intensity based augmentation technique to enhance the segmentation results and build a multi-sequences segmentation model.

6.3 Research Objectives 'Revisited'

The main objectives of this research which are defined under Chapter 1 are revisited herein to ensure they have been successfully addressed. The objectives have been determined with the focus on the area of medical imaging segmentation using a variation of deep learning related techniques.

1. **Have the medical imaging modalities: MR and CT been explored and investigated?**

   The medical imaging modalities MR and CT have been investigated using MATLAB and Amide application. The 3D volume and 2D slices have been visualised to observe the difference between the two modalities and ensure their alignments.

2. **Has the 3D CNN been efficiently designed for brain segmentation?**

   The design of a 3D CNN that follows a convolutional encoder decoder architecture has been thoroughly investigated by conducting several experiments to explore the impact of different components of the network and select the optimal ones. The conducted experiments with the results analysis are illustrated in Chapter 3.

3. **Has the combination of the multiresolution analysis techniques such as Contourlet and Shearlet transforms with the CNN's features been investigated and successfully implemented?**

   A novel segmentation approach which fuses the multidirectional and multiscale features with CNN's features has been proposed and implemented. Different scales and directional coefficients have been investigated using a particular features selection method. The details of this approach with the implementation steps and the obtained results have been detailed in Chapter 3.

4. **Has a novel CNN that uses squeezing and excitation blocks to reduce the complexity been proposed?**

   A 3D attention based convolutional encoder decoder network has been proposed using two squeeze and excitation mechanisms which help to recalibrate the features adaptively and boost the most dominant features in order to improve the brain MR images segmentation. This network has been successfully implemented to decrease the model complexity by relying only on the CNN's features and avoiding the overhead of handcrafted features. The details of the implementation with the obtained results are presented in Chapter 4.

5. **Has an ensemble of different CNNs been proposed?**

   An ensemble of multiple attention based CNNs has been successfully proposed and implemented. The ensemble approach has shown an increase in the performance with the capability to build a reliable segmentation approach. The results of this objective are demonstrated in Chapter 4.

6. **Has the applicability of transfer learning from different MR sequences has been investigated?**

   The transfer learning has been applied using two MR sequences : T1-w and T2-w. Due to the small size of T2-w datasets, the knowledge is transferred from T1-w to T2-w images. The transfer learning has been successfully applied by exploring different fine tuning approaches. Additionally the use of T2-w MR images as an intensity based augmentation approach has been investigated and implemented to build multi-modality segmentation approach. The details of the transfer learning and multi-modality approach are illustrated in Chapter 5.

7. **Have the proposed segmentation approaches been evaluated?**

   Each of the proposed approaches has been evaluated using clinical datasets of fifty patients with different demographics and clinical diagnosis. Moreover, the performance of each approach has been assessed by calculating four quantitative evaluation metrics (precision, recall, DSC, and JSC) as well as the visual comparison with the ground truth.

8. **Have the proposed segmentation approaches been validated with other studies?**

   The first and second segmentation approaches have been compared with other works in the literature that applied deep learning to segment MR images for brain PET attenuation correction. The validation of the third approach which is related to transfer learning is not applicable.

6.4 Research Limitations

Although the determined objectives of this thesis have been successfully fulfilled, some limitations have been identified and categorised as follows:

6.4.1 Deep Learning

The main challenge of training a deep neural network is the need of big data to obtain a well-performed model. Another challenge is the lack of interpretability for certain behaviours of the deep learning. Additionally, the computational cost is a limitation when designing a deep network. It is crucial to use high performance computing resources such as graphical processing units with large memory capacity to train millions of parameters to segment 3D images. This is very challenging during the design stage of the deep network where many trials and experiments should be conducted given that each experiment can take up to eight hours with Tesla V100 GPU with 16 GB memory. The memory capacity of the GPU is another important resource that should be considered when designing a 3D deep network. One medical volume consists of thousands of voxels and in order to use the whole volume as an input to the network in order to preserve the contextual information, there is a need to use a GPU with memory size of 32 GB or higher.

6.4.2 Handcrafted Features

The process of extraction and selection of handcrafted features is cumbersome and time consuming. The selection of the Contourlet and Shearlet decomposition levels requires careful consideration by conducting more experiments to select the optimal decomposition levels. In this research, the decomposition levels and scales are randomly selected.

6.4.3 Generalised Model for Segmentation

The datasets consist of patients with different demographics and neurological diagnosis. The obtained segmentation results have revealed there is a high variability in the bone segmentation. Therefore, there is no guarantee that the proposed methods can be used for patients with different anatomical structures or neurological diseases. Moreover, this work focuses only on brain segmentation instead of whole body which is more challenging.

6.4.4 PET Attenuation Correction

Despite the promising performance of the deep learning guided MR segmentation for PET attenuation correction to identify the brain tissue classes, the predefined linear attenuation coefficient assignment to each tissue class is not very practical. The continuous patient-specific attenuation value estimation is a more promising approach which requires more consideration. The accurate bone delineation with patient specific attenuation value would lead to perfect attenuation correction and PET quantification. Hence, there is a need to find a direct mapping between MR intensity and tissue density (i.e., linear attenuation coefficient). Moreover, there is a lack of comprehensive comparison with existing techniques reported in the literature (e.g. CT synthesis) that generate pseudo CT images with continuous values.

6.5 Future Directions

6.5.1 Disease Specific Segmentation

Currently, the proposed segmentation CNN is trained with datasets which include patients with different clinical neurological diagnosis. This variation hinders the process of building an accurate and robust segmentation model since the deep learning is an application/data specific approach. Hence, training one model with a dataset that consists of patients with the same neurological diagnosis is a promising approach to achieve a robust and reliable segmentation model. The application of transfer learning is another possible solution to train models with datasets of patients with different brain diseases. A pretrained model with a specific dataset can be fine tuned to build another model that works with datasets of different diseases.

6.5.2 More Data

Deep learning is data hungry technique and always requires more data to increase the learning capability especially with more complicated learning algorithms. Data augmentation is a well-known technique to increase the size of the training dataset. The conventional data augmentation techniques such as rotation, flipping, and mirroring have been applied to enhance the segmentation performance, yet no enhancement has been observed. The application of different data augmentation techniques such as using generative adversarial network (GAN) to generate augmented images is a promising approach to be investigated in the future. On the other hand, the identification of the optimal training size is essential to build a successful model which is able to generalise without memorisation of the training data. The convergence of the objective function that measures the error of both training and validation is an indication of a sufficient sample size [280].

6.5.3 Further Evaluation for PET Images

The focus of this research is directed toward the enhancement of segmentation algorithm using deep learning technique. Brain segmentation is a critical task to generate pseudo CT image which is an essential part of the integrated process of PET attenuation correction. In order to fully evaluate the PET attenuation correction process, further evaluation is required by using the segmented images to generate the pseudo CT images. The pseudo CT images will be used to create the attenuation maps and reconstruct the corrected PET images. The evaluation and analysis of the reconstructed PET images is a future objective to ensure that the efficiency of the proposed segmentation approaches.

6.5.4 Pseudo CT Generation using Regression

MR image segmentation based approach for PET attenuation correction has some limitations as stated previously in Chapter 2. One of these limitation is the assignment of a predefined linear attenuation coefficient for each tissue class to generate the attenuation map. This predefined coefficient can lead to inaccurate attenuation value for each patient and eventually leads to PET quantification biases. As a promising solution, there are some recent studies [281, 282] which are proposing the mapping between MR intensity value and tissue linear attenuation coefficient without the need to apply any segmentation. This approach depends on the employment of regression technique using deep learning. Therefore, the application of regression task instead of segmentation task is an elegant and promising solution which can be investigated in the future.

Bibliography

[1] K. Shi, S. Fürst, L. Sun, et al. "Individual refinement of attenuation correction maps for hybrid PET/MR based on multi-resolution regional learning". In: *Computerized Medical Imaging and Graphics*. Computational Methods for Molecular Imaging 60 (Sept. 2017), pp. 50–57.

[2] J. Cabello, M. Lukas, S. Förster, et al. "MR-based attenuation correction using ultrashort-echo-time pulse sequences in dementia patients". In: *J Nucl Med* 56.3 (2015), pp. 423–429.

[3] A. Aitken, D Giese, C Tsoumpas, et al. "Improved UTE-based attenuation correction for cranial PET-MR using dynamic magnetic field monitoring". In: *Medical physics* 41.1 (2014).

[4] Y. Berker, J. Franke, A. Salomon, et al. "MRI-based attenuation correction for hybrid PET/MRI Systems: A 4-class tissue segmentation technique using a combined ultrashort-echo-time/Dixon MRI sequence". In: *Journal of Nuclear Medicine* 53.5 (2012), pp. 796–804.

[5] R. Bammer. "Basic principles of diffusion-weighted imaging". In: *European journal of radiology* 45.3 (2003), pp. 169–184.

[6] P. Khateri, H. Saligheh Rad, A. H. Jafari, et al. "Generation of a Four-Class Attenuation Map for MRI-Based Attenuation Correction of PET Data in the Head Area Using a Novel Combination of STE/Dixon-MRI and FCM Clustering". en. In: *Molecular Imaging and Biology* 17.6 (Dec. 2015), pp. 884–892.

[7] A Waldman, J. Rees, C. Brock, et al. “MRI of the brain with ultra-short echo-time pulse sequences”. In: *Neuroradiology* 45.12 (2003), pp. 887–892.

[8] M Weiger and K. Pruessmann. “MRI with zero echo time”. In: *eMagRes* (2012).

[9] N. Hylton. “Dynamic contrast-enhanced magnetic resonance imaging as an imaging biomarker”. In: *J Clin Oncol* 24.20 (2006), pp. 3293–3298.

[10] J. P. Mugler III and J. R. Brookeman. “Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE)”. In: *Magnetic Resonance in Medicine* 15.1 (1990), pp. 152–157.

[11] C. N. Ladefoged, D. Benoit, I. Law, et al. “Region specific optimization of continuous linear attenuation coefficients based on UTE (RESOLUTE): application to PET/MR brain imaging”. In: *Physics in Medicine & Biology* 60.20 (2015), p. 8047.

[12] *X-Ray Mass Attenuation Coefficients*. `https://physics.nist.gov/PhysRefData/XrayMassCoef/tab2.html`. [Online; accessed 26-February-2019].

[13] G. Wagenknecht, H.-J. Kaiser, F. M. Mottaghy, et al. “MRI for attenuation correction in PET: methods and challenges”. en. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 26.1 (Feb. 2013), pp. 99–113.

[14] N. Burgos, M. J. Cardoso, K. Thielemans, et al. “Attenuation correction synthesis for hybrid PET-MR scanners: application to brain studies.” In: *IEEE Trans. Med. Imaging* 33.12 (2014), pp. 2332–2341.

[15] E. Schreibmann, J. A. Nye, D. M. Schuster, et al. “MR-based attenuation correction for hybrid PET-MR brain imaging systems using deformable image registration”. In: *Medical physics* 37.5 (2010), pp. 2101–2109.

[16] J. Sjölund, D. Forsberg, M. Andersson, et al. “Generating patient specific pseudo-CT of the head from MR using atlas-based regression”. In: *Physics in Medicine & Biology* 60.2 (2015), p. 825.

[17] I. Mérida, N. Costes, R. A. Heckemann, et al. “Evaluation of several multi-atlas methods for PSEUDO-CT generation in brain MRI-PET attenuation correction”. In: *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*. IEEE. 2015, pp. 1431–1434.

[18] K. T. Chen, D. Izquierdo-Garcia, C. B. Poynton, et al. “On the accuracy and reproducibility of a novel probabilistic atlas-based generation for calculation of head attenuation maps on integrated PET/MR scanners”. In: *European journal of nuclear medicine and molecular imaging* 44.3 (2017), pp. 398–407.

[19] A. Mehranian, H. Arabi, and H. Zaidi. “Vision 20/20: Magnetic resonance imaging-guided attenuation correction in PET/MRI: Challenges, solutions, and opportunities”. en. In: *Medical Physics* 43.3 (Mar. 2016), pp. 1130–1155.

[20] K.-H. Su, L. Hu, C. Stehning, et al. “Generation of brain pseudo-CTs using an undersampled, single-acquisition UTE-mDixon pulse sequence and unsupervised clustering”. en. In: *Medical Physics* 42.8 (2015), pp. 4974–4986.

[21] A. Santos Ribeiro, E. Rota Kops, H. Herzog, et al. “Skull segmentation of UTE MR images by probabilistic neural network for attenuation correction in PET/MR”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. PET/MR and SPECT/MR: New Paradigms for Combined Modalities in Molecular Imaging 702 (Feb. 2013), pp. 114–116.

[22] P. Khateri, H. S. Rad, A. H. Jafari, et al. “A novel segmentation approach for implementation of MRAC in head PET/MRI employing Short-TE MRI and 2-point Dixon method in a fuzzy C-means framework”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. PSMR2013 - PET-MR and SPECT-MR: Current status of Instrumentation, Applications and Developments 734 (Jan. 2014), pp. 171–174.

[23] B. Fei, X. Yang, J. A. Nye, et al. “MR/PET quantification tools: Registration, segmentation, classification, and MR-based attenuation correction”. en. In: *Medical Physics* 39.10 (2012), pp. 6443–6454.

[24] S. L. S. Chan, R. L. Jeffree, M. Fay, et al. “Automated Classification of Bone and Air Volumes for Hybrid PET-MRI Brain Imaging”. In: *2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. Nov. 2013, pp. 1–8.

[25] T. Koesters, K. P. Friedman, M. Fenchel, et al. “Dixon Sequence with Superimposed Model-Based Bone Compartment Provides Highly Accurate PET/MR Attenuation Correction of the Brain”. en. In: *Journal of Nuclear Medicine* 57.6 (June 2016), pp. 918–924.

[26] A. Johansson, M. Karlsson, and T. Nyholm. “CT substitute derived from MRI sequences with ultrashort echo time”. In: *Medical physics* 38.5 (2011), pp. 2708–2714.

[27] S. Roy, W.-T. Wang, A. Carass, et al. “PET Attenuation Correction Using Synthetic CT from Ultrashort Echo-Time MR Imaging”. en. In: *Journal of Nuclear Medicine* 55.12 (Dec. 2014), pp. 2071–2077.

[28] Y. Wu, W. Yang, L. Lu, et al. “Prediction of CT Substitutes from MR Images Based on Local Diffeomorphic Mapping for Brain PET Attenuation Correction”. en. In: *Journal of Nuclear Medicine* 57.10 (Oct. 2016), pp. 1635–1641.

[29] L. Zhong, L. Lin, Z. Lu, et al. “Predict CT image from MRI data using KNN-regression with learned local descriptors”. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2016, pp. 743–746.

[30] W. Yang, L. Zhong, Y. Chen, et al. “Predicting CT Image From MRI Data Through Feature Matching With Learned Nonlinear Local Descriptors”. In: *IEEE Transactions on Medical Imaging* 37.4 (Apr. 2018), pp. 977–987.

[31] X. Yang, Y. Lei, H.-K. Shu, et al. “Pseudo CT estimation from MRI using patch-based random forest”. In: *Medical Imaging 2017: Image Processing*. Vol. 10133. International Society for Optics and Photonics, Feb. 2017, 101332Q.

[32] T. Huynh, Y. Gao, J. Kang, et al. “Estimating CT Image From MRI Data Using Structured Random Forest and Auto-Context Model”. In: *IEEE Transactions on Medical Imaging* 35.1 (Jan. 2016), pp. 174–183.

[33] L. Xiang, Q. Wang, D. Nie, et al. “Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image”. In: *Medical image analysis* 47 (2018), pp. 31–44.

[34] D. Pustina, H. B. Coslett, P. E. Turkeltaub, et al. “Automated segmentation of chronic stroke lesions using LINDA: Lesion identification with neighborhood data analysis”. en. In: *Human Brain Mapping* 37.4 (Apr. 2016), pp. 1405–1421.

[35] G. B. Praveen, A. Agrawal, P. Sundaram, et al. “Ischemic stroke lesion segmentation using stacked sparse autoencoder”. In: *Computers in Biology and Medicine* 99 (Aug. 2018), pp. 38–52.

[36] J. K. Boldsen, T. S. Engedal, S. Pedraza, et al. “Better Diffusion Segmentation in Acute Ischemic Stroke Through Automatic Tree Learning Anomaly Segmentation”. eng. In: *Frontiers in Neuroinformatics* 12 (2018), p. 21.

[37] M. Salem, M. Cabezas, S. Valverde, et al. “A supervised framework with intensity subtraction and deformation field features for the detection of new T2-w lesions in multiple sclerosis”. In: *NeuroImage: Clinical* 17 (Jan. 2018), pp. 607–615.

[38] S. Roy, Q. He, E. Sweeney, et al. “Subject-Specific Sparse Dictionary Learning for Atlas-Based Brain MRI Segmentation”. In: *IEEE Journal of Biomedical and Health Informatics* 19.5 (Sept. 2015), pp. 1598–1609.

[39] T. Brosch, Y. Yoo, L. Y. Tang, et al. “Deep convolutional encoder networks for multiple sclerosis lesion segmentation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, pp. 3–11.

[40] T. Brosch, L. Y. W. Tang, Y. Yoo, et al. “Deep 3D Convolutional Encoder Networks With Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation”. In: *IEEE Transactions on Medical Imaging* 35.5 (May 2016), pp. 1229–1239.

[41] A. van Opbroek, H. C. Achterberg, and M. de Bruijne. “Feature-Space Transformation Improves Supervised Segmentation Across Scanners”. en. In: *Machine Learning Meets*

Medical Imaging. Ed. by K. Bhatia and H. Lombaert. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 85–93.

[42] M. Liu, D. Cheng, K. Wang, et al. “Multi-modality cascaded convolutional neural networks for Alzheimer’s disease diagnosis”. In: *Neuroinformatics* (2018), pp. 1–14.

[43] L. Rundo, C. Militello, A. Tangherloni, et al. “NeXt for neuro-radiosurgery: A fully automatic approach for necrosis extraction in brain tumor MRI using an unsupervised machine learning technique”. en. In: *International Journal of Imaging Systems and Technology* 28.1 (Mar. 2018), pp. 21–37.

[44] F. P. Polly, S. K. Shil, M. A. Hossain, et al. “Detection and classification of HGG and LGG brain tumor using machine learning”. In: *2018 International Conference on Information Networking (ICOIN)*. Jan. 2018, pp. 813–817.

[45] A. F. I. Osman. “Automated Brain Tumor Segmentation on Magnetic Resonance Images and Patient’s Overall Survival Prediction Using Support Vector Machines”. en. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by A. Crimi, S. Bakas, H. Kuijf, et al. Lecture Notes in Computer Science. Springer International Publishing, 2018, pp. 435–449.

[46] F. W. Prior, S. J. Fouke, T. Benzinger, et al. “Predicting a multi-parametric probability map of active tumor extent using random forests”. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. July 2013, pp. 6478–6481.

[47] M. Havaei, A. Davy, D. Warde-Farley, et al. “Brain tumor segmentation with Deep Neural Networks”. In: *Medical Image Analysis* 35 (Jan. 2017), pp. 18–31.

[48] S. Sedlar. “Brain Tumor Segmentation Using a Multi-path CNN Based Method”. en. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by A. Crimi, S. Bakas, H. Kuijf, et al. Lecture Notes in Computer Science. Springer International Publishing, 2018, pp. 403–422.

[49] M. Soltaninejad, G. Yang, T. Lambrou, et al. "Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels". In: *Computer Methods and Programs in Biomedicine* 157 (Apr. 2018), pp. 69–84.

[50] J. R. Stone, E. A. Wilde, B. A. Taylor, et al. "Supervised learning technique for the automated identification of white matter hyperintensities in traumatic brain injury". en. In: *Brain Injury* 30.12 (Oct. 2016), pp. 1458–1468.

[51] Z. Wu, Y. Gao, F. Shi, et al. "Automatic Hippocampal Subfield Segmentation from 3T Multi-modality Images". en. In: *Machine Learning in Medical Imaging*. Ed. by L. Wang, E. Adeli, Q. Wang, et al. Lecture Notes in Computer Science. Springer International Publishing, 2016, pp. 229–236.

[52] S. Bonte, I. Goethals, and R. Van Holen. "Machine learning based brain tumour segmentation on limited data using local texture and abnormality". In: *Computers in Biology and Medicine* 98 (July 2018), pp. 39–47.

[53] M. Rincón, E. Díaz-López, P. Selnes, et al. "Improved Automatic Segmentation of White Matter Hyperintensities in MRI Based on Multilevel Lesion Features". en. In: *Neuroinformatics* 15.3 (July 2017), pp. 231–245.

[54] T. Magome, H. Arimura, S. Kakeda, et al. "Automated segmentation method of white matter and gray matter regions with multiple sclerosis lesions in MR images". en. In: *Radiological Physics and Technology* 4.1 (Jan. 2011), pp. 61–72.

[55] W. Fan, C. Hu, J. Du, et al. "A Novel Model-Based Approach for Medical Image Segmentation Using Spatially Constrained Inverted Dirichlet Mixture Models". en. In: *Neural Processing Letters* 47.2 (Apr. 2018), pp. 619–639.

[56] A. Ahmadvand, M. R. Daliri, and S. M. Zahiri. "Segmentation of brain MR images using a proper combination of DCS based method with MRF". en. In: *Multimedia Tools and Applications* 77.7 (Apr. 2018), pp. 8001–8018.

[57] C. Elliott, D. L. Arnold, D. L. Collins, et al. "A generative model for automatic detection of resolving multiple sclerosis lesions". In: *Bayesian and grAphical Models for Biomedical Imaging*. Springer, 2014, pp. 118–129.

[58] M. Tamajka and W. Benesova. “Automatic brain segmentation method based on supervoxels”. In: *2016 International Conference on Systems, Signals and Image Processing (IWSSIP)*. May 2016, pp. 1–4.

[59] A. De and C. Guo. “An image segmentation method based on the fusion of vector quantization and edge detection with applications to medical image processing”. en. In: *International Journal of Machine Learning and Cybernetics* 5.4 (Aug. 2014), pp. 543–551.

[60] K. Kothavari, B. Arunadevi, and S. N. Deepa. “A Hybrid DE-RGSO-ELM for Brain Tumor Tissue Categorization in 3D Magnetic Resonance Images”. en. In: *Mathematical Problems in Engineering* 2014 (2014), pp. 1–14.

[61] B. Yilmaz, A. Durdu, and G. D. Emlik. “A new method for skull stripping in brain MRI using multistable cellular neural networks”. en. In: *Neural Computing and Applications* 29.8 (Apr. 2018), pp. 79–95.

[62] A. Mekhmoukh and K. Mokrani. “Improved Fuzzy C-Means based Particle Swarm Optimization (PSO) initialization and outlier rejection with level set methods for MR brain image segmentation”. In: *Computer Methods and Programs in Biomedicine* 122.2 (Nov. 2015), pp. 266–281.

[63] Y. Xie and X. Tao. “White matter lesion segmentation using machine learning and weakly labeled MR images”. In: *Medical Imaging 2011: Image Processing*. Vol. 7962. International Society for Optics and Photonics, Mar. 2011, 79622G.

[64] M. Rajchl, M. C. H. Lee, O. Oktay, et al. “DeepCut: Object Segmentation From Bounding Box Annotations Using Convolutional Neural Networks”. In: *IEEE Transactions on Medical Imaging* 36.2 (Feb. 2017), pp. 674–683.

[65] S. Pedemonte, B. Bizzo, S. Pomerantz, et al. “Detection and Delineation of Acute Cerebral Infarct on DWI Using Weakly Supervised Machine Learning”. en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, et al. Lecture Notes in Computer Science. Springer International Publishing, 2018, pp. 81–88.

[66] N. Tomita, S. Jiang, M. E. Maeder, et al. “Automatic post-stroke lesion segmentation on MR images using 3D residual convolutional neural network”. en. In: *NeuroImage: Clinical* 27 (Jan. 2020), p. 102276.

[67] S. Valverde, M. Cabezas, E. Roura, et al. “Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach”. In: *NeuroImage* 155 (July 2017), pp. 159–168.

[68] G. Karayegen and M. F. Aksahin. “Brain tumor prediction on MR images with semantic segmentation by using deep learning network and 3D imaging of tumor region”. en. In: *Biomedical Signal Processing and Control* 66 (Apr. 2021), p. 102458.

[69] J. Zhang, J. Zeng, P. Qin, et al. “Brain tumor segmentation of multi-modality MR images via triple intersecting U-Nets”. en. In: *Neurocomputing* 421 (Jan. 2021), pp. 195–209.

[70] D. Zhang, G. Huang, Q. Zhang, et al. “Cross-modality deep feature learning for brain tumor segmentation”. en. In: *Pattern Recognition* 110 (Feb. 2021), p. 107562.

[71] P. Hambarde, S. Talbar, A. Mahajan, et al. “Prostate lesion segmentation in MR images using radiomics based deeply supervised U-Net”. en. In: *Biocybernetics and Biomedical Engineering* 40.4 (Oct. 2020), pp. 1421–1435.

[72] J. Dolz, C. Desrosiers, L. Wang, et al. “Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation”. en. In: *Computerized Medical Imaging and Graphics* 79 (Jan. 2020), p. 101660.

[73] N. Nogovitsyn, R. Souza, M. Muller, et al. “Testing a deep convolutional neural network for automated hippocampus segmentation in a longitudinal sample of healthy participants”. en. In: *NeuroImage* 197 (Aug. 2019), pp. 589–597.

[74] Y. Lemoigne and A. Caner. *Molecular Imaging: Computer Reconstruction and Practice*. Springer Science & Business Media, 2008.

[75] M. Defrise, P. E. Kinahan, and C. J. Michel. “Image reconstruction algorithms in PET”. In: *Positron Emission Tomography*. Springer, 2005, pp. 63–91.

[76] *Medical imaging modalities involving X-rays and medical radiation*. `https://openmedscience.com/medical-imaging/`. Accessed: 2021-08-17.

[77] A. Berger. “How does it work?: Positron emission tomography”. In: *BMJ: British Medical Journal* 326.7404 (2003), p. 1449.

[78] F. H. Fahey. “Data acquisition in PET imaging”. In: *Journal of nuclear medicine technology* 30.2 (2002), pp. 39–49.

[79] H. Watabe, K. Matsumoto, M. Senda, et al. “Performance of list mode data acquisition with ECAT EXACT HR and ECAT EXACT HR+ positron emission scanners”. In: *Annals of nuclear medicine* 20.3 (2006), pp. 189–194.

[80] A. M. Paans and A. van Waarde. “Positron emission tomography”. In: *METHODS-LONDON-A COMPANION TO METHODS IN ENZYMOLOGY-* 27.3 (2002), pp. 193–194.

[81] R. B. Buxton. *Introduction to functional magnetic resonance imaging: principles and techniques*. Cambridge university press, 2009.

[82] *MRI scan.* `https://snc2dmri.weebly.com/components--functions.html`. Accessed: 2021-08-25.

[83] *MRI pulse sequences.* `https://radiopaedia.org/articles/mri-pulse-sequences-1?lang=us`. Accessed: 2021-08-20.

[84] *MRI sequences (overview).* `https://radiopaedia.org/articles/mri-sequences-overview`. Accessed: 2021-08-20.

[85] *Magnetic Resonance Imaging (MRI) of the Brain and Spine: Basics.* `https://case.edu/med/neurology/NR/MRI\%20Basics.htm`. Accessed: 2021-08-20.

[86] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.

[87] A. Maier, S. Steidl, V. Christlein, et al., eds. *Medical Imaging Systems: An Introductory Guide*. en. Vol. 11111. Cham: Springer International Publishing, 2018.

[88] *X-ray/CT technique.* `https://www.startradiology.com/the-basics/x-rayct-technique/index.html`. Accessed: 2021-09-16.

[89] *What is the Difference Between CT, MRI, and PET Scans?* `https://aica.com/whats-the-difference-between-ct-mri-and-pet-scans/`. Accessed: 2022-01-30.

[90] *CT AND PET SCAN VS MRI*. `https://clinicalposters.com/blogs/health/radiology-comparison`. Accessed: 2022-01-30.

[91] M. K. Werner, H. Schmidt, and N. F. Schwenzer. "MR/PET: a new challenge in hybrid imaging". In: *American Journal of Roentgenology* 199.2 (2012), pp. 272–277.

[92] H. Zaidi and A. Del Guerra. "An outlook on future design of hybrid PET/MRI systems". In: *Medical physics* 38.10 (2011), pp. 5667–5689.

[93] *Hybrid Imaging: PET-CT and PET-MRI*. `https://radiologykey.com/hybrid-imaging-pet-ct-and-pet-mri/`. Accessed: 2021-08-25.

[94] A. F. Kazerooni, M. R. Ay, S. Arfaie, et al. "Single STE-MR Acquisition in MR-Based Attenuation Correction of Brain PET Imaging Employing a Fully Automated and Reproducible Level-Set Segmentation Approach". en. In: *Molecular Imaging and Biology* 19.1 (Feb. 2017), pp. 143–152.

[95] H. J. An, S. Seo, H. Kang, et al. "MRI-Based Attenuation Correction for PET/MRI Using Multiphase Level-Set Method". en. In: *Journal of Nuclear Medicine* 57.4 (Apr. 2016), pp. 587–593.

[96] C. Catana, A. v. d. Kouwe, T. Benner, et al. "Toward Implementing an MRI-Based PET Attenuation-Correction Method for Neurologic Studies on the MR-PET Brain Prototype". en. In: *Journal of Nuclear Medicine* 51.9 (Sept. 2010), pp. 1431–1438.

[97] G. Delso, K. Zeimpekis, M. Carl, et al. "Cluster-based segmentation of dual-echo ultra-short echo time images for PET/MR bone localization". In: *EJNMMI Physics* 1.1 (June 2014), p. 7.

[98] H. Jang, F. Liu, T. Bradshaw, et al. "Rapid dual-echo ramped hybrid encoding MR-based attenuation correction (dRHE-MRAC) for PET/MR". en. In: *Magnetic Resonance in Medicine* 79.6 (2018), pp. 2912–2922.

[99] M. Khalifé, B. Fernandez, O. Jaubert, et al. "Subject-specific bone attenuation correction for brain PET/MR: can ZTE-MRI substitute CT scan accurately?" en. In: *Physics in Medicine & Biology* 62.19 (2017), p. 7814.

[100] M. R. Juttukonda, B. G. Mersereau, Y. Chen, et al. "MR-based attenuation correction for PET/MRI neurological studies with continuous-valued attenuation coefficients for bone through a conversion from R2* to CT-Hounsfield units". In: *NeuroImage* 112 (May 2015), pp. 160–168.

[101] V. Keereman, Y. Fierens, T. Broux, et al. "MRI-Based Attenuation Correction for PET/MRI Using Ultrashort Echo Time Sequences". en. In: *Journal of Nuclear Medicine* 51.5 (May 2010), pp. 812–818.

[102] X. Yang and B. Fei. "Multiscale segmentation of the skull in MR images for MRI-based attenuation correction of combined MR/PET". en. In: *Journal of the American Medical Informatics Association* 20.6 (Nov. 2013), pp. 1037–1045.

[103] F. Liu, H. Jang, R. Kijowski, et al. "Deep learning MR imaging–based attenuation correction for PET/MR imaging". In: *Radiology* 286.2 (2017), pp. 676–684.

[104] H. Jang, F. Liu, G. Zhao, et al. "Technical Note: Deep learning based MRAC using rapid ultrashort echo time imaging". en. In: *Medical Physics* 45.8 (2018), pp. 3697–3704.

[105] H. Yang, X. Lu, S.-H. Wang, et al. "Synthesizing Multi-Contrast MR Images Via Novel 3D Conditional Variational Auto-Encoding GAN". en. In: *Mobile Networks and Applications* 26.1 (Feb. 2021), pp. 415–424.

[106] A. Abu-Srhan, I. Almallahi, M. A. M. Abushariah, et al. "Paired-unpaired Unsupervised Attention Guided GAN with transfer learning for bidirectional brain MR-CT synthesis". en. In: *Computers in Biology and Medicine* 136 (Sept. 2021), p. 104763.

[107] L. Tao, J. Fisher, E. Anaya, et al. "Pseudo CT Image Synthesis and Bone Segmentation From MR Images Using Adversarial Networks With Residual Blocks for MR-Based Attenuation Correction of Brain PET Data". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.2 (Mar. 2021), pp. 193–201.

[108] M. Hemsley, B. Chugh, M. Ruschin, et al. “Deep Generative Model for Synthetic-CT Generation with Uncertainty Predictions”. en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Ed. by A. L. Martel, P. Abolmaesumi, D. Stoyanov, et al. Cham: Springer International Publishing, 2020, pp. 834–844.

[109] A. Ranjan, D. Lalwani, and R. Misra. “GAN for synthesizing CT from T2-weighted MRI data towards MR-guided radiation treatment”. en. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* (Nov. 2021).

[110] R. Farjam, H. Nagar, X. K. Zhou, et al. “Deep learning-based synthetic CT generation for MR-only radiotherapy of prostate cancer patients with 0.35T MRI linear accelerator”. en. In: *Journal of Applied Clinical Medical Physics* n/a.n/a ().

[111] Y. Chen, C. Ying, M. M. Binkley, et al. “Deep learning-based T1-enhanced selection of linear attenuation coefficients (DL-TESLA) for PET/MR attenuation correction in dementia neuroimaging”. en. In: *Magnetic Resonance in Medicine* n/a.n/a ().

[112] K. Kläser, T. Varsavsky, P. Markiewicz, et al. “Imitation learning for improved 3D PET/MR attenuation correction”. en. In: *Medical Image Analysis* 71 (July 2021), p. 102079.

[113] V. Badrinarayanan, A. Kendall, and R. Cipolla. “Segnet: A deep convolutional encoder-decoder architecture for image segmentation”. In: *arXiv preprint arXiv:1511.00561* (2015).

[114] M. H. M. Zaini, M. I. Shapiai, A. R. Mohamed, et al. “Hippocampal segmentation using structured extreme learning machine with bag of features”. In: *2017 International Conference on Robotics, Automation and Sciences (ICORAS)*. Nov. 2017, pp. 1–5.

[115] B. Ayerdi, A. Savio, and M. Graña. “Meta-ensembles of classifiers for Alzheimer’s disease detection using independent ROI features”. In: *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer. 2013, pp. 122–130.

[116] M. F. Rachmadi, M. d. C. Valdés-Hernández, M. L. F. Agan, et al. “Segmentation of white matter hyperintensities using convolutional neural networks with global spatial

information in routine clinical brain MRI with none or mild vascular pathology”. In: *Computerized Medical Imaging and Graphics* 66 (June 2018), pp. 28–43.

[117] P. K. Roy, A. Bhuiyan, A. Janke, et al. “Automatic white matter lesion segmentation using contrast enhanced FLAIR intensity and Markov Random Field”. In: *Computerized Medical Imaging and Graphics* 45 (Oct. 2015), pp. 102–111.

[118] A. Serag, J. P. Boardman, A. G. Wilkinson, et al. “A sparsity-based atlas selection technique for multiple-atlas segmentation: Application to neonatal brain labeling”. In: *2016 24th Signal Processing and Communication Application Conference (SIU)*. May 2016, pp. 2265–2268.

[119] J. Amin, M. Sharif, M. Raza, et al. “Detection of Brain Tumor based on Features Fusion and Machine Learning”. en. In: *Journal of Ambient Intelligence and Humanized Computing* (Nov. 2018).

[120] V. Keereman, P. Mollet, Y. Berker, et al. “Challenges and current methods for attenuation correction in PET/MR”. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 26.1 (2013), pp. 81–98.

[121] Y. Chen and H. An. “Attenuation Correction of PET/MR Imaging”. In: *Magnetic Resonance Imaging Clinics of North America*. Hybrid PET/MR Imaging 25.2 (May 2017), pp. 245–255.

[122] T. J. Bradshaw, G. Zhao, H. Jang, et al. “Feasibility of Deep Learning–Based PET/MR Attenuation Correction in the Pelvis Using Only Diagnostic MR Images”. In: *Tomography* 4.3 (Sept. 2018), pp. 138–147.

[123] K. Kamnitsas, E. Ferrante, S. Parisot, et al. “DeepMedic for Brain Tumor Segmentation”. en. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by A. Crimi, B. Menze, O. Maier, et al. Lecture Notes in Computer Science. Springer International Publishing, 2016, pp. 138–149.

[124] S. Roy, J. A. Butman, and D. L. Pham. “Synthesizing CT from Ultrashort Echo-Time MR Images via Convolutional Neural Networks”. In: *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer. 2017, pp. 24–32.

[125] A. P. Leynes, J. Yang, F. Wiesinger, et al. "Zero-Echo-Time and Dixon Deep Pseudo-CT (ZeDD CT): Direct Generation of Pseudo-CT Images for Pelvic PET/MRI Attenuation Correction Using Deep Convolutional Neural Networks with Multiparametric MRI". In: *Journal of Nuclear Medicine* 59.5 (May 2018), pp. 852–858.

[126] X. Han. "MR-based synthetic CT generation using a deep convolutional neural network method". In: *Medical physics* 44.4 (2017), pp. 1408–1419.

[127] H. Emami, M. Dong, S. P. Nejad-Davarani, et al. "Generating synthetic CTs from magnetic resonance images using generative adversarial networks". en. In: *Medical Physics* 45.8 (2018), pp. 3627–3636.

[128] A. M. Dinkla, J. M. Wolterink, M. Maspero, et al. "MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network". en. In: *International Journal of Radiation Oncology*Biology*Physics* 102.4 (Nov. 2018), pp. 801–812.

[129] S. Chen, A. Qin, D. Zhou, et al. "Technical Note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning". en. In: *Medical Physics* 45.12 (2018), pp. 5659–5665.

[130] Y. Fu, T. R. Mazur, X. Wu, et al. "A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy". en. In: *Medical Physics* 45.11 (2018), pp. 5129–5137.

[131] *Brain Tumor Segmentation (BraTS)*. `http://braintumorsegmentation.org/`. Accessed: 2022-02-06.

[132] M. Antonelli, A. Reinke, S. Bakas, et al. "The medical segmentation decathlon". In: *arXiv preprint arXiv:2106.05735* (2021).

[133] *BrainMetShare*. `https://stanfordaimi.azurewebsites.net/datasets/ae0182f1-d5b6-451a-8177-d1f39f019962`. Accessed: 2022-02-06.

[134] *OASIS Brains Datasets*. `https://www.oasis-brains.org`. Accessed: 2022-02-06.

[135] K. Payette, P. de Dumast, H. Kebiri, et al. “An automatic multi-tissue human fetal brain segmentation benchmark using the Fetal Tissue Annotation Dataset”. In: *Scientific Data* 8.1 (2021), pp. 1–14.

[136] R. Souza, O. Lucena, J. Garrafa, et al. “An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement”. In: *NeuroImage* 170 (2018), pp. 482–494.

[137] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “ImageNet classification with deep convolutional neural networks”. en. In: *Communications of the ACM* 60.6 (May 2017), pp. 84–90.

[138] C. Szegedy, V. Vanhoucke, S. Ioffe, et al. “Rethinking the Inception Architecture for Computer Vision”. In: 2016, pp. 2818–2826.

[139] K. He, X. Zhang, S. Ren, et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 770–778.

[140] K. Simonyan and A. Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *arXiv:1409.1556 [cs]* (Apr. 2015). arXiv: 1409.1556.

[141] J. Long, E. Shelhamer, and T. Darrell. “Fully Convolutional Networks for Semantic Segmentation”. In: *arXiv:1411.4038 [cs]* (Mar. 2015). arXiv: 1411.4038.

[142] F. Yu, V. Koltun, and T. Funkhouser. “Dilated residual networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 472–480.

[143] O. Ronneberger, 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, pp. 234–241.

[144] F. Milletari, N. Navab, and S.-A. Ahmadi. “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation”. In: *arXiv:1606.04797 [cs]* (June 2016). arXiv: 1606.04797.

[145] J. Hu, L. Shen, and G. Sun. “Squeeze-and-excitation networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7132–7141.

[146] C. Yu, J. Wang, C. Peng, et al. “Bisenet: Bilateral segmentation network for real-time semantic segmentation”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 325–341.

[147] C. Yu, J. Wang, C. Peng, et al. “Learning a discriminative feature network for semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 1857–1866.

[148] Q. Cheng, C. Zhang, Z. Li, et al. “Attention and Multi-layer Fusion for Real-time Semantic Segmentation”. In: *2019 12th International Symposium on Computational Intelligence and Design (ISCID)*. Vol. 1. ISSN: 2473-3547. Dec. 2019, pp. 142–145.

[149] J. Bian and Y. Liu. “Dual Channel Attention Networks”. en. In: *Journal of Physics: Conference Series* 1642 (Sept. 2020), p. 012004.

[150] J. Deng, Y. Ma, D.-a. Li, et al. “Classification of breast density categories based on SE-Attention neural networks”. en. In: *Computer Methods and Programs in Biomedicine* 193 (Sept. 2020), p. 105489.

[151] L. Rundo, C. Han, Y. Nagano, et al. “USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets”. en. In: *Neurocomputing* 365 (Nov. 2019), pp. 31–43.

[152] A. G. Roy, N. Navab, and C. Wachinger. “Concurrent Spatial and Channel ‘Squeeze & Excitation’ in Fully Convolutional Networks”. en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 421–429.

[153] A. G. Roy, N. Navab, and C. Wachinger. “Recalibrating Fully Convolutional Networks With Spatial and Channel “Squeeze and Excitation” Blocks”. In: *IEEE Transactions on Medical Imaging* 38.2 (Feb. 2019), pp. 540–549.

[154] H. Wu, J. Pan, Z. Li, et al. “Automated Skin Lesion Segmentation Via an Adaptive Dual Attention Module”. In: *IEEE Transactions on Medical Imaging* 40.1 (2020), pp. 357–370.

[155] P. Liu, Q. Dou, Q. Wang, et al. “An Encoder-Decoder Neural Network With 3D Squeeze-and-Excitation and Deep Supervision for Brain Tumor Segmentation”. In: *IEEE Access* 8 (2020), pp. 34029–34037.

[156] B. Hou, X. Xu, G. Kang, et al. “Hybrid Attention Densely Connected Ensemble Framework for Lesion Segmentation From Magnetic Resonance Images”. In: *IEEE Access* 8 (2020), pp. 188564–188576.

[157] Y. Han, C. Wei, R. Zhou, et al. “Combining 3D-CNN and squeeze-and-excitation networks for remote sensing sea ice image classification”. In: *Mathematical Problems in Engineering* 2020 (2020).

[158] M. Kaur and D. Singh. “Multi-modality medical image fusion technique using multi-objective differential evolution based deep neural networks”. en. In: *Journal of Ambient Intelligence and Humanized Computing* 12.2 (Feb. 2021), pp. 2483–2493.

[159] S. Liu, L. Gao, Y. Lei, et al. “SAR Speckle Removal Using Hybrid Frequency Modulations”. In: *IEEE Transactions on Geoscience and Remote Sensing* 59.5 (May 2021), pp. 3956–3966.

[160] A. Khan, A. Chefranov, and H. Demirel. “Image scene geometry recognition using low-level features fusion at multi-layer deep CNN”. en. In: *Neurocomputing* 440 (June 2021), pp. 111–126.

[161] X. Gao, Z. Tang, Y. Xie, et al. “A layered working condition perception integrating handcrafted with deep features for froth flotation”. en. In: *Minerals Engineering* 170 (Aug. 2021), p. 107059.

[162] J. M. Fajardo, O. Gomez, and F. Prieto. “EMG hand gesture classification using handcrafted and deep features”. en. In: *Biomedical Signal Processing and Control* 63 (Jan. 2021), p. 102210.

[163] S. Devulapalli, A. Potti, R. Krishnan, et al. “Experimental evaluation of unsupervised image retrieval application using hybrid feature extraction by integrating deep learning and handcrafted techniques”. en. In: *Materials Today: Proceedings* (June 2021).

[164] F. B. Tesema, H. Wu, M. Chen, et al. “Hybrid channel based pedestrian detection”. In: *Neurocomputing* 389 (2020), pp. 1–8.

[165] S. Shang, J. Sun, Z. Yue, et al. “Multi-parametric MRI based radiomics with tumor subregion partitioning for differentiating benign and malignant soft-tissue tumors”. en. In: *Biomedical Signal Processing and Control* 67 (May 2021), p. 102522.

[166] G. F. Roberto, A. Lumini, L. A. Neves, et al. “Fractal Neural Network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images”. en. In: *Expert Systems with Applications* 166 (Mar. 2021), p. 114103.

[167] M. Choudhary, V. Tiwari, and V. U. “Iris anti-spoofing through score-level fusion of handcrafted and data-driven features”. en. In: *Applied Soft Computing* 91 (June 2020), p. 106206.

[168] Z. Golrizkhatami and A. Acan. “ECG classification using three-level fusion of different feature descriptors”. en. In: *Expert Systems with Applications* 114 (Dec. 2018), pp. 54–64.

[169] D. S. Shibu and S. S. Priyadharsini. “Multi scale decomposition based medical image fusion using convolutional neural network and sparse representation”. en. In: *Biomedical Signal Processing and Control* 69 (Aug. 2021), p. 102789.

[170] Z. Wang, X. Li, H. Duan, et al. “Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform”. en. In: *Expert Systems with Applications* 171 (June 2021), p. 114574.

[171] M. Liu, L. Jiao, X. Liu, et al. “C-CNN: Contourlet Convolutional Neural Networks”. In: *IEEE Transactions on Neural Networks and Learning Systems* 32.6 (June 2021). Conference Name: IEEE Transactions on Neural Networks and Learning Systems, pp. 2636–2649.

[172] L. Fang, H. Zhang, J. Zhou, et al. “Image classification with an RGB-channel nonsubsampled contourlet transform and a convolutional neural network”. en. In: *Neurocomputing* 396 (July 2020), pp. 266–277.

[173] L. Li, L. Ma, L. Jiao, et al. “Complex Contourlet-CNN for polarimetric SAR image classification”. en. In: *Pattern Recognition* 100 (Apr. 2020), p. 107110.

[174] A.-A. Nahid and Y. Kong. “Histopathological Breast-Image Classification Using Local and Frequency Domains by Convolutional Neural Network”. en. In: *Information* 9.1 (Jan. 2018). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 19.

[175] Y. Zhao, Y. Li, and B. Yang. “Low-Frequency Desert Noise Intelligent Suppression in Seismic Data Based on Multiscale Geometric Analysis Convolutional Neural Network”. In: *IEEE Transactions on Geoscience and Remote Sensing* 58.1 (Jan. 2020), pp. 650–665.

[176] Y. Liu, C. Zhang, J. Cheng, et al. “A multi-scale data fusion framework for bone age assessment with convolutional neural networks”. en. In: *Computers in Biology and Medicine* 108 (May 2019), pp. 161–173.

[177] G. Yang, J. Yang, Z. Lu, et al. “A convolutional neural network with sparse representation”. en. In: *Knowledge-Based Systems* 209 (Dec. 2020), p. 106419.

[178] M. Imani. “Texture feed based convolutional neural network for pansharpening”. en. In: *Neurocomputing* 398 (July 2020), pp. 117–130.

[179] Y. Li, L.-M. Po, C.-H. Cheung, et al. “No-Reference Video Quality Assessment With 3D Shearlet Transform and Convolutional Neural Networks”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 26.6 (June 2016). Conference Name: IEEE Transactions on Circuits and Systems for Video Technology, pp. 1044–1057.

[180] H. Hermessi, O. Mourali, and E. Zagrouba. “Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain”. en. In: *Neural Computing and Applications* 30.7 (Oct. 2018), pp. 2029–2045.

[181] R. Hou, D. Zhou, R. Nie, et al. “Brain CT and MRI medical image fusion using convolutional neural networks and a dual-channel spiking cortical model”. en. In: *Medical & Biological Engineering & Computing* 57.4 (Apr. 2019), pp. 887–900.

[182] H. Hermessi, O. Mourali, and E. Zagrouba. “Deep feature learning for soft tissue sarcoma classification in MR images via transfer learning”. en. In: *Expert Systems with Applications* 120 (Apr. 2019), pp. 116–127.

[183] H. Rezaeilouyeh, A. Mollahosseini, and M. H. Mahoor. “Microscopic medical image classification framework via deep learning and shearlet transform”. In: *Journal of Medical Imaging* 3.4 (Oct. 2016).

[184] M. Liang, Z. Ren, J. Yang, et al. “Identification of Colon Cancer Using Multi-Scale Feature Fusion Convolutional Neural Network Based on Shearlet Transform”. In: *IEEE Access* 8 (2020). Conference Name: IEEE Access, pp. 208969–208977.

[185] E. Jabason, M. O. Ahmad, and M. N. S Swamy. “Shearlet based Stacked Convolutional Network for Multiclass Diagnosis of Alzheimer’s Disease using the Florbetapir PET Amyloid Imaging Data”. en. In: *2018 16th IEEE International New Circuits and Systems Conference (NEWCAS)*. Montreal, QC: IEEE, June 2018, pp. 344–347.

[186] M. Galar, A. Fernandez, E. Barrenechea, et al. “A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42.4 (2011), pp. 463–484.

[187] M. S. Sharif, M. Abbod, A. Al-Bayatti, et al. “An Accurate Ensemble Classifier for Medical Volume Analysis: Phantom and Clinical PET Study”. In: *IEEE Access* 8 (2020), pp. 37482–37494.

[188] S. Bakas, M. Reyes, A. Jakab, et al. “Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge”. In: *arXiv preprint arXiv:1811.02629* (2018).

[189] S. Paisitkriangkrai, C. Shen, and A. van den Hengel. “Pedestrian detection with spatially pooled features and structured ensemble learning”. In: *IEEE transactions on pattern analysis and machine intelligence* 38.6 (2015), pp. 1243–1257.

[190] S. Dey, R. Bhattacharya, F. Schwenker, et al. “Median Filter Aided CNN Based Image Denoising: An Ensemble Approach”. en. In: *Algorithms* 14.4 (Apr. 2021), p. 109.

[191] G. Pandey and U. Ghanekar. “Single image super-resolution using multi-scale feature enhancement attention residual network”. en. In: *Optik* 231 (Apr. 2021), p. 166359.

[192] S. Xia, Y. Xia, H. Yu, et al. “Transferring Ensemble Representations Using Deep Convolutional Neural Networks for Small-Scale Image Classification”. In: *IEEE Access* 7 (2019), pp. 168175–168186.

[193] Z. Chen, J. Duan, L. Kang, et al. “Class-Imbalanced Deep Learning via a Class-Balanced Ensemble”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2021). Conference Name: IEEE Transactions on Neural Networks and Learning Systems, pp. 1–15.

[194] M. Carranza-García, P. Lara-Benítez, J. García-Gutiérrez, et al. “Enhancing object detection for autonomous driving by optimizing anchor generation and addressing class imbalance”. en. In: *Neurocomputing* 449 (Aug. 2021), pp. 229–244.

[195] Z. Zheng, H. Qi, L. Zhuang, et al. “Automated rail surface crack analytics using deep data-driven models and transfer learning”. en. In: *Sustainable Cities and Society* 70 (July 2021), p. 102898.

[196] W. A. Haque, S. Arefin, A. S. M. Shihavuddin, et al. “DeepThin: A novel lightweight CNN architecture for traffic sign recognition without GPU requirements”. en. In: *Expert Systems with Applications* 168 (Apr. 2021), p. 114481.

[197] M. Khanramaki, E. Askari Asli-Ardeh, and E. Kozegar. “Citrus pests classification using an ensemble of deep learning models”. en. In: *Computers and Electronics in Agriculture* 186 (July 2021), p. 106192.

[198] E. Ayan, H. Erbay, and F. Varçın. “Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks”. en. In: *Computers and Electronics in Agriculture* 179 (Dec. 2020), p. 105809.

[199] E. Tasci. “Voting combinations-based ensemble of fine-tuned convolutional neural networks for food image recognition”. en. In: *Multimedia Tools and Applications* 79.41 (Nov. 2020), pp. 30397–30418.

[200] V.-E. Neagoe and P. Diaconescu. “An Ensemble of Deep Convolutional Neural Networks for Drunkenness Detection Using Thermal Infrared Facial Imagery”. In: *2020 13th International Conference on Communications (COMM)*. June 2020, pp. 147–150.

[201] P. He, H. Li, and H. Wang. “Detection of Fake Images Via The Ensemble of Deep Representations from Multi Color Spaces”. In: *2019 IEEE International Conference on Image Processing (ICIP)*. ISSN: 2381-8549. Sept. 2019, pp. 2299–2303.

[202] A. Abayomi-Alli, O. Atinuke, S. A. Onashoga, et al. “Facial Image Quality Assessment using an Ensemble of Pre-Trained Deep Learning Models (EFQnet)”. In: *2020 20th International Conference on Computational Science and Its Applications (ICCSA)*. July 2020, pp. 1–8.

[203] D. Liu, J. Chen, Z. Huang, et al. “A deep attention-based ensemble network for real-time face hallucination”. en. In: *Journal of Real-Time Image Processing* 17.6 (Dec. 2020), pp. 1927–1937.

[204] D. M. Vo, D. M. Nguyen, and S.-W. Lee. “Deep softmax collaborative representation for robust degraded face recognition”. en. In: *Engineering Applications of Artificial Intelligence* 97 (Jan. 2021), p. 104052.

[205] S. Mohammadi, S. G. Majelan, and S. B. Shokouhi. “Ensembles of Deep Neural Networks for Action Recognition in Still Images”. In: *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*. ISSN: 2643-279X. Oct. 2019, pp. 315–318.

[206] A. Das, P. Sil, P. K. Singh, et al. “MMHAR-EnsemNet: A Multi-Modal Human Activity Recognition Model”. In: *IEEE Sensors Journal* 21.10 (May 2021). Conference Name: IEEE Sensors Journal, pp. 11569–11576.

[207] X. He and Y. Chen. “Transferring CNN Ensemble for Hyperspectral Image Classification”. In: *IEEE Geoscience and Remote Sensing Letters* 18.5 (May 2021), pp. 876–880.

[208] B. Fielding and L. Zhang. "Evolving Deep DenseBlock Architecture Ensembles for Image Classification". en. In: *Electronics* 9.11 (Nov. 2020). Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, p. 1880.

[209] B. Fielding, T. Lawrence, and L. Zhang. "Evolving and Ensembling Deep CNN Architectures for Image Classification". In: *2019 International Joint Conference on Neural Networks (IJCNN)*. ISSN: 2161-4407. July 2019, pp. 1–8.

[210] A. Afifi, N. E. Hafsa, M. A. S. Ali, et al. "An Ensemble of Global and Local-Attention Based Convolutional Neural Networks for COVID-19 Diagnosis on Chest X-ray Images". en. In: *Symmetry* 13.1 (Jan. 2021). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 113.

[211] F. Ahmad, A. Farooq, and M. U. Ghani. "Deep Ensemble Model for Classification of Novel Coronavirus in Chest X-Ray Images". en. In: *Computational Intelligence and Neuroscience* 2021 (Jan. 2021). Publisher: Hindawi, e8890226.

[212] A. K. Das, S. Ghosh, S. Thunder, et al. "Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network". en. In: *Pattern Analysis and Applications* (Mar. 2021).

[213] M. Turkoglu. "COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble". en. In: *Applied Intelligence* 51.3 (Mar. 2021), pp. 1213–1226.

[214] H. Liz, M. Sánchez-Montañés, A. Tagarro, et al. "Ensembles of Convolutional Neural Network models for pediatric pneumonia diagnosis". en. In: *Future Generation Computer Systems* 122 (Sept. 2021), pp. 220–233.

[215] R. Elakkiya, P. Vijayakumar, and M. Karuppiah. "COVID_SCREENET: COVID-19 Screening in Chest Radiography Images Using Deep Transfer Stacking". en. In: *Information Systems Frontiers* (Mar. 2021).

[216] M. Shorfuzzaman. "An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection". en. In: *Multimedia Systems* (Apr. 2021).

[217] S. A. A. Ahmed, B. Yanikoğlu, Göksu, et al. “Skin Lesion Classification With Deep CNN Ensembles”. In: *2020 28th Signal Processing and Communications Applications Conference (SIU)*. ISSN: 2165-0608. Oct. 2020, pp. 1–4.

[218] L. Wei, K. Ding, and H. Hu. “Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network”. In: *IEEE Access* 8 (2020), pp. 99633–99647.

[219] S. Osowski and T. Les. “Deep Learning Ensemble for Melanoma Recognition”. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. ISSN: 2161-4407. July 2020, pp. 1–7.

[220] P. Tang, Q. Liang, X. Yan, et al. “GP-CNN-DTEL: Global-Part CNN Model With Data-Transformed Ensemble Learning for Skin Lesion Classification”. In: *IEEE Journal of Biomedical and Health Informatics* 24.10 (Oct. 2020), pp. 2870–2882.

[221] C. M. Vasile, A. L. Udriștoiu, A. E. Ghenea, et al. “Intelligent Diagnosis of Thyroid Ultrasound Imaging Using an Ensemble of Deep Learning Methods”. en. In: *Medicina* 57.395 (Apr. 2021). Publisher: MDPI AG, p. 395.

[222] M. Byra, A. Han, A. S. Boehringer, et al. “Liver Fat Assessment in Multiview Sonography Using Transfer Learning With Convolutional Neural Networks”. en. In: *Journal of Ultrasound in Medicine* n/a.n/a ().

[223] N. Chouhan, A. Khan, J. Z. Shah, et al. “Deep convolutional neural network and emotional learning based breast cancer detection using digital mammography”. en. In: *Computers in Biology and Medicine* 132 (May 2021), p. 104318.

[224] S. J. Malebary and A. Hashmi. “Automated Breast Mass Classification System Using Deep Learning and Ensemble Learning in Digital Mammogram”. In: *IEEE Access* 9 (2021), pp. 55312–55328.

[225] S. Bhattacharjee, C.-H. Kim, D. Prakash, et al. “An Efficient Lightweight CNN and Ensemble Machine Learning Classification of Prostate Tissue Using Multilevel Feature Analysis”. en. In: *Applied Sciences* 10.8013 (Nov. 2020). Publisher: MDPI AG, p. 8013.

[226] R. Sanyal, D. Kar, and R. Sarkar. “Carcinoma type classification from high-resolution breast microscopy images using a hybrid ensemble of deep convolutional features and gradient boosting trees classifiers”. In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2021). Conference Name: IEEE/ACM Transactions on Computational Biology and Bioinformatics, pp. 1–1.

[227] B. Zhang, S. Qi, P. Monkam, et al. “Ensemble Learners of Multiple Deep CNNs for Pulmonary Nodules Classification Using CT Images”. In: *IEEE Access* 7 (2019), pp. 110358–110371.

[228] A. Basher, B. C. Kim, K. H. Lee, et al. “Volumetric Feature-Based Alzheimer’s Disease Diagnosis From sMRI Data Using a Convolutional Neural Network and a Deep Neural Network”. en. In: *IEEE Access* 9 (Jan. 2021). Publisher: IEEE, pp. 29870–29882.

[229] L. Nanni, M. Interlenghi, S. Brahnam, et al. “Comparison of Transfer Learning and Conventional Machine Learning Applied to Structural Brain MRI for the Early Diagnosis and Prognosis of Alzheimer’s Disease”. en. In: *Frontiers in Neurology* 11 (Nov. 2020). Publisher: Frontiers Media S.A.

[230] J. Kim and L. Tran. “Ensemble Learning Based on Convolutional Neural Networks for the Classification of Retinal Diseases from Optical Coherence Tomography Images”. In: *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. ISSN: 2372-9198. July 2020, pp. 532–537.

[231] I. A. Khan, A. Sajeeb, and S. A. Fattah. “An Automatic Ocular Disease Detection Scheme from Enhanced Fundus Images Based on Ensembling Deep CNN Networks”. In: *2020 11th International Conference on Electrical and Computer Engineering (ICECE)*. Dec. 2020, pp. 491–494.

[232] A. Tiba, Z. Bartik, H. Toman, et al. “Detecting outlier and poor quality medical images with an ensemble-based deep learning system”. In: *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*. ISSN: 1849-2266. Sept. 2019, pp. 99–104.

[233] Z. Jahanshiri, M. S. Abadeh, and H. Sajedi. “Brain Age Estimation based on Brain MRI by an Ensemble of Deep Networks”. In: *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*. Jan. 2021, pp. 1–6.

[234] S. Wang, X. Wang, Y. Shen, et al. “An Ensemble-Based Densely-Connected Deep Learning System for Assessment of Skeletal Maturity”. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2020), pp. 1–12.

[235] L. mao, T. Yumeng, and C. Lina. “Pneumonia Detection in chest X-rays: a deep learning approach based on ensemble RetinaNet and Mask R-CNN”. In: *2020 Eighth International Conference on Advanced Cloud and Big Data (CBD)*. Dec. 2020, pp. 213–218.

[236] H. Aetesam and S. K. Maji. “Noise dependent training for deep parallel ensemble denoising in magnetic resonance images”. en. In: *Biomedical Signal Processing and Control* 66 (Apr. 2021), p. 102405.

[237] P. Coupé, B. Mansencal, M. Clément, et al. “AssemblyNet: A novel deep decision-making process for whole brain MRI segmentation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2019, pp. 466–474.

[238] R. Karthik, R. Menaka, M. Hariharan, et al. “Ischemic Lesion Segmentation using Ensemble of Multi-Scale Region Aligned CNN”. en. In: *Computer Methods and Programs in Biomedicine* 200 (Mar. 2021), p. 105831.

[239] H. Mei, W. Lei, R. Gu, et al. “Automatic segmentation of gross target volume of nasopharynx cancer using ensemble of multiscale deep neural networks with spatial attention”. en. In: *Neurocomputing* 438 (May 2021), pp. 211–222.

[240] M. Goyal, A. Oakley, P. Bansal, et al. “Skin Lesion Segmentation in Dermoscopic Images With Ensemble Deep Learning Methods”. In: *IEEE Access* 8 (2020). Conference Name: IEEE Access, pp. 4171–4181.

[241] T. Y. Tan, L. Zhang, C. P. Lim, et al. “Evolving Ensemble Models for Image Segmentation Using Enhanced Particle Swarm Optimization”. In: *IEEE Access* 7 (2019). Conference Name: IEEE Access, pp. 34004–34019.

[242] I. Banerjee, A. Crawley, M. Bhethanabotla, et al. “Transfer learning on fused multiparametric MR images for classifying histopathological subtypes of rhabdomyosarcoma”. In: *Computerized Medical Imaging and Graphics* 65 (2018), pp. 167–175.

[243] K. Kushibar, S. Valverde, S. González-Villà, et al. “Supervised Domain Adaptation for Automatic Sub-cortical Brain Structure Segmentation with Minimal User Interaction”. In: *Scientific Reports* 9.1 (2019).

[244] S. Vesal, N. Ravikumar, and A. Maier. “Automated multi-sequence cardiac MRI segmentation using supervised domain adaptation”. In: *International workshop on statistical atlases and computational models of the heart*. Springer. 2019, pp. 300–308.

[245] M. Ghafoorian, A. Mehrtash, T. Kapur, et al. “Transfer learning for domain adaptation in mri: Application in brain lesion segmentation”. In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2017, pp. 516–524.

[246] A. Chen, T. Zhou, I. Icke, et al. “Transfer learning for the fully automatic segmentation of left ventricle myocardium in porcine cardiac cine MR images”. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer. 2017, pp. 21–31.

[247] D. Kessler, J. MacKay, V. Crowe, et al. “The optimisation of deep neural networks for segmenting multiple knee joint tissues from MRIs”. In: *Computerized Medical Imaging and Graphics* 86 (2020).

[248] X. Zhao, P. Xie, M. Wang, et al. “Deep learning–based fully automated detection and segmentation of lymph nodes on multiparametric-mri for rectal cancer: A multicentre study”. In: *EBioMedicine* 56 (2020).

[249] D. Ataloglou, A. Dimou, D. Zarpalas, et al. “Fast and Precise Hippocampus Segmentation Through Deep Convolutional Neural Network Ensembles and Transfer Learning”. en. In: *Neuroinformatics* 17.4 (Oct. 2019), pp. 563–582.

[250] C. N. Ladefoged, A. E. Hansen, O. M. Henriksen, et al. “AI-driven attenuation correction for brain PET/MRI: Clinical evaluation of a dementia cohort and importance of the training group size”. en. In: *NeuroImage* 222 (Nov. 2020), p. 117221.

[251] M. Arshad, M. Qureshi, O. Inam, et al. “Transfer learning in deep neural network based under-sampled MR image reconstruction”. en. In: *Magnetic Resonance Imaging* 76 (Feb. 2021), pp. 96–107.

[252] F. Eitel, E. Soehler, J. Bellmann-Strobl, et al. “Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation”. en. In: *NeuroImage: Clinical* 24 (Jan. 2019), p. 102003.

[253] F. Grimm, F. Edl, S. Kerscher, et al. “Semantic segmentation of cerebrospinal fluid and brain volume with a convolutional neural network in pediatric hydrocephalus—transfer learning from existing algorithms”. In: *Acta Neurochirurgica* 162.10 (2020), pp. 2463–2474.

[254] S. Cui, L. Mao, J. Jiang, et al. “Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network”. In: *Journal of Healthcare Engineering* 2018 (2018).

[255] A. Kuzina, E. Egorov, and E. Burnaev. “Bayesian generative models for knowledge transfer in MRI semantic segmentation problems”. In: *Frontiers in Neuroscience* 13.JUL (2019).

[256] R. Mehrotra, M. A. Ansari, R. Agrawal, et al. “A Transfer Learning approach for AI-based classification of brain tumors”. en. In: *Machine Learning with Applications* 2 (Dec. 2020), p. 100003.

[257] M. A. Naser and M. J. Deen. “Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images”. en. In: *Computers in Biology and Medicine* 121 (June 2020), p. 103758.

[258] Z. N. K. Swati, Q. Zhao, M. Kabir, et al. “Brain tumor classification for MR images using transfer learning and fine-tuning”. en. In: *Computerized Medical Imaging and Graphics* 75 (July 2019), pp. 34–46.

[259] M. Talo, U. B. Baloglu, Yıldırım, et al. "Application of deep transfer learning for automated brain abnormality classification using MR images". en. In: *Cognitive Systems Research* 54 (May 2019), pp. 176–188.

[260] X. Zhao, H. Zhang, Y. Zhou, et al. "Gibbs-ringing artifact suppression with knowledge transfer from natural images to MR images". en. In: *Multimedia Tools and Applications* 79.45-46 (Dec. 2020), pp. 33711–33733.

[261] Y. Xia, N. Ravikumar, J. P. Greenwood, et al. "Super-Resolution of Cardiac MR Cine Imaging using Conditional GANs and Unsupervised Transfer Learning". en. In: *Medical Image Analysis* 71 (July 2021), p. 102037.

[262] E. Ölmez, V. Akdoğan, M. Korkmaz, et al. "Automatic Segmentation of Meniscus in Multispectral MRI Using Regions with Convolutional Neural Network (R-CNN)". en. In: *Journal of Digital Imaging* 33.4 (Aug. 2020), pp. 916–929.

[263] S. Klein, M. Staring, K. Murphy, et al. "Elastix: a toolbox for intensity-based medical image registration". In: *IEEE transactions on medical imaging* 29.1 (2010), pp. 196–205.

[264] S. A. Taghanaki, K. Abhishek, J. P. Cohen, et al. "Deep semantic segmentation of natural and medical images: A review". In: *Artificial Intelligence Review* 54.1 (2021), pp. 137–178.

[265] X. Glorot and Y. Bengio. "Understanding the difficulty of training deep feedforward neural networks". In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings. 2010, pp. 249–256.

[266] K. He, X. Zhang, S. Ren, et al. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1026–1034.

[267] H. Yu, M. Oliver, K. Leszczynski, et al. "Tissue segmentation-based electron density mapping for MR-only radiotherapy treatment planning of brain using conventional T1-

weighted MR images". In: *Journal of Applied Clinical Medical Physics* 20.8 (2019), pp. 11–20.

[268] A. L. Da Cunha, J. Zhou, and M. N. Do. "The nonsubsampled contourlet transform: theory, design, and applications". In: *IEEE transactions on image processing* 15.10 (2006), pp. 3089–3101.

[269] *Arthur Cunha, Nonsubsampled Contourlet Toolbox, MATLAB Central File Exchange.* `https://www.mathworks.com/matlabcentral/fileexchange/10049-nonsubsampled-contourlet-toolbox`. Accessed: 2021-02-15.

[270] S. Biswas and J. Sil. "An efficient face recognition method using contourlet and curvelet transform". In: *Journal of King Saud University-Computer and Information Sciences* 32.6 (2020), pp. 718–729.

[271] *Entropy*. `https://www.mathworks.com/help/images/ref/entropy.html#responsive_offcanvas`. Accessed: 2022-02-10.

[272] Z. Qu, Y. Xing, and Y. Song. "An image enhancement method based on non-subsampled shearlet transform and directional information measurement". In: *Information* 9.12 (2018), p. 308.

[273] P Subramanian, N. Alamelu, and M Aramudhan. "Image fusion method using non-subsampled shearlet transform and fuzzy and simple fuzzy neural network algorithms". In: *The Journal of Defense Modeling and Simulation* 13.1 (2016), pp. 23–33.

[274] G. Kutyniok, W.-Q. Lim, and R. Reisenhofer. "Shearlab 3D: Faithful digital shearlet transforms based on compactly supported shearlets". In: *ACM Transactions on Mathematical Software (TOMS)* 42.1 (2016), pp. 1–42.

[275] R. R. Kelly, S. J. Sidles, and A. C. LaRue. "Effects of Neurological Disorders on Bone Health". English. In: *Frontiers in Psychology* 11 (2020). Publisher: Frontiers.

[276] X. Glorot, A. Bordes, and Y. Bengio. "Deep sparse rectifier neural networks". In: *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings. 2011, pp. 315–323.

[277] A. L. Maas, A. Y. Hannun, and A. Y. Ng. "Rectifier nonlinearities improve neural network acoustic models". In: *Proc. icml*. Vol. 30. 1. Citeseer. 2013, p. 3.

[278] B. Xu, N. Wang, T. Chen, et al. "Empirical evaluation of rectified activations in convolutional network". In: *arXiv preprint arXiv:1505.00853* (2015).

[279] N. Abraham and N. M. Khan. "A novel focal tversky loss function with improved attention u-net for lesion segmentation". In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE. 2019, pp. 683–687.

[280] H. Zaidi and I. El Naqa. "Quantitative molecular positron emission tomography imaging using advanced deep learning techniques". In: *Annual Review of Biomedical Engineering* 23 (2021).

[281] D. Nie, R. Trullo, J. Lian, et al. "Medical image synthesis with deep convolutional adversarial networks". In: *IEEE Transactions on Biomedical Engineering* 65.12 (2018), pp. 2720–2730.

[282] H. Arabi, G. Zeng, G. Zheng, et al. "Novel adversarial semantic structure deep learning for MRI-guided attenuation correction in brain PET/MRI". In: *European journal of nuclear medicine and molecular imaging* 46.13 (2019), pp. 2746–2759.