Machine Learning and Quantitative Imaging for the Management of Brain Metastasis

Seyed Ali Jalalifar

A Dissertation Submitted to
the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy in
Electrical Engineering and Computer Science

Graduate Program in
Electrical Engineering and Computer Science
York University
Toronto, Ontario

December 2022

Abstract

Significantly affecting patients' clinical course and quality of life, a growing number of cancer cases are diagnosed with brain metastasis annually. Although a considerable percentage of cancer patients survive for several years if the disease is discovered at an early stage while it is still localized, when the tumour is metastasized to the brain, the median survival decreases considerably. Early detection followed by precise and effective treatment of brain metastasis may lead to improved patient survival and quality of life. A main challenge to prescribe an effective treatment regimen is the variability of tumour response to treatments, *e.g.*, radiotherapy as a main treatment option for brain metastasis, despite similar cancer therapy, due to many patient-related factors. Stratifying patients based on their predicted response and consequently assessing their response to therapy are challenging yet crucial tasks. While risk assessment models with standard clinical attributes have been proposed for patient stratification, the imaging data acquired for these patients as a part of the standard-of-care are not computationally analyzed or directly incorporated in these models. Further, therapy response monitoring and assessment is a cumbersome task for patients with brain metastasis that requires longitudinal tumour delineation on MRI volumes before and at multiple follow-up sessions after treatment. This is aggravated by the time-sensitive nature of the disease. In an effort to address these challenges, a number of machine learning frameworks and computational techniques in areas of automatic tumour segmentation, radiotherapy outcome assessment, and therapy outcome prediction have been introduced and investigated in this dissertation. Powered by advanced machine learning algorithms, a complex attention-guided segmentation framework is introduced and investigated for segmenting brain tumours on serial MRI. The experimental results demonstrate that the proposed framework can achieve a dice score of 91.5% and 84.1% to 87.4% on the baseline and follow-up scans, respectively. This framework is then applied in a proposed system that follows

standard clinical criteria based on changes in tumour size at post-treatment to assess tumour response to radiotherapy automatically. The system demonstrates a very good agreement with expert clinicians in detecting local response, with an accuracy of over 90%. Next, innovative machine-learning-based solutions are proposed and investigated for radiotherapy outcome prediction before or early after therapy, using MRI radiomic models and novel deep learning architectures that analyze treatment-planning MRI with and without standard clinical attributes. The developed models demonstrate an accuracy of up to 82.5% in predicting radiotherapy outcome before the treatment initiation. The ground-breaking machine learning platforms presented in this dissertation along with the promising results obtained in the conducted experiments are steps forward towards realizing important decision support tools for oncologists and radiologists and, can eventually, pave the way towards the personalized therapeutics paradigm for cancer patients.

Originality Statement

This dissertation has been written by Seyed Ali Jalalifar under the supervision of Dr. Ali Sadeghi-Naini. Parts of the material presented in this thesis have been published in peer-reviewed journals or conference proceedings, as listed below. The research presented in each publication has been performed by the principal author, supervised/supported by the senior author, and in collaboration with other co-authors who participated in project design and/or data acquisition/preprocessing.

- The material in Chapter 1 has partly been presented in:
    - S. A. Jalalifar, H. Soliman, A. Sahgal, A. Sadeghi-Naini. *Radiomics and its role in predicting radiotherapy response and outcome in brain metastasis*. In: A. Sahgal, S. S. Lo, L. Ma, J. P. Sheehan. editors. Image-guided hypofractionated stereotactic radiosurgery for brain and spinal tumours, Second Edition. Boca Raton, FL, USA: CRC Press; 2021. chapter 33. ISBN: 9781003037095.
    - S. A. Jalalifar, H. Soliman, M. Ruschin, A. Sahgal, A. Sadeghi-Naini. *A brain tumor segmentation framework based on outlier detection using one-class support vector machine*. 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). July 2020; Montreal, QC, Canada. pp. 1067-1070.
- The material in Chapter 2 has been presented in:
    - S. A. Jalalifar, H. Soliman, A. Sahgal, A. Sadeghi-Naini. *Automatic assessment of stereotactic radiation therapy outcome in brain metastasis using longitudinal tumour segmentation on serial MRI*. IEEE Journal of Biomedical and Health Informatics, 2023. DOI: 10.1109/JBHI.2023.3235304.
- The material in Chapter 3 has been presented in:
    - S. A. Jalalifar, H. Soliman, A. Sahgal, A. Sadeghi-Naini. *Impact of tumour segmentation accuracy on efficacy of quantitative MRI biomarkers of radiotherapy outcome in brain metastasis.* Cancers, vol. 14, no. 20, p. 5133, 2022. DOI: 10.3390/cancers14205133.

- The material in Chapter 4 has been presented in:
    - S. A. Jalalifar, H. Soliman, A. Sahgal, A. Sadeghi-Naini. *Predicting the outcome of radiotherapy in brain metastasis by integrating the clinical and MRI-based deep learning features*. Medical Physics, vol. 49, no. 11, pp. 7167–7178, 2022. DOI: 10.1002/mp.15814.

- The material in Chapter 5 has been presented in:
    - S. A. Jalalifar, H. Soliman, A. Sahgal, A. Sadeghi-Naini. *A self-attention-guided 3D deep residual network with big transfer to predict local failure in brain metastasis after radiotherapy using multi-channel MRI*. IEEE Journal of Translational Engineering in Health and Medicine, vol. 11, pp. 13–22, 2023. DOI: 10.1109/JTEHM.2022.3219625.

Acknowledgments

First and foremost, I want to thank my supervisor Dr. Ali Sadeghi-Naini. It has been an honor to be his Ph.D. student. I appreciate all Dr. Sadeghi-Naini's contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating.

I would also like to thank my supervisory committee members, Dr. Hui Jiang and Dr. Michael Brown for their ideas, suggestions, and guidance which lead to improving the quality of my work.

I am fortunate to have the support of amazing friends who make my long Ph.D. years memorable. My most special thanks go to my parents and sister for their unconditional love and supports during these years.

Table of Contents

Abstract……ii

Originality Statement……iv

Acknowledgement……vi

List of Tables……x

List of Figures……xii

List of Appendices……xvi

List of Acronyms and Abbreviations……xvii

Chapter 1
Introduction 1

1.1. Brain Metastasis: Background 1

1.2. Treatment Options 3

1.3. Imaging in Brain Metastasis Management 6

1.4. Therapy Response Evaluation 8

1.5. Prognostic Factors and Biomarkers for Risk Assessment 10

1.5.1. Biomarkers of Cancer Therapy Response 11

1.6. Literature Review 12

1.6.1. Brain Image Segmentation 12

1.6.2. Automated Radiotherapy Outcome Assessment 15

1.6.3. Quantitative Imaging for Therapy Outcome Prediction 16

1.7. Objectives 19

1.8. Dissertation Outline 20

1.8.1. Chapter 2 20

1.8.2. Chapter 3 21

1.8.3. Chapter 4 21

1.8.4. Chapter 5 21

1.8.5. Chapter 6 22

Chapter 2
Automatic Assessment of Stereotactic Radiation Therapy Outcome in Brain Metastasis using Longitudinal Segmentation on Serial MRI 23

2.1. Introduction 23

2.2. Methods 27

2.2.1. Date Acquisition and pre-processing 27

2.2.2. Tumour Segmentation Framework 29

2.2.3. Training and Evaluation of the System 32

2.2.4. Procedure and Criteria for Automatic Assessment of Tumour Size Status, Local Response, and ARE Outcome 35

2.3. Results 38

2.4. Discussion and Conclusion 47

Chapter 3
Impact of Tumour Segmentation Accuracy on Efficacy of Quantitative MRI Biomarkers of Radiotherapy Outcome in Brain Metastasis 55

3.1. Introduction 55

3.2. Materials and Methods 60

3.2.1. Data Acquisition 60

3.2.2. System Overview 61

3.2.3. Segmentation Module 62

3.2.4. Radiomic Feature Extraction 62

3.2.5. Feature Extraction 63

3.2.6. Classifier 63

3.3. Results 64

3.4. Discussion and Conclusion 70

Chapter 4
Predicting the Outcome of Radiotherapy in Brain Metastasis by Integrating the Clinical and MRI-Based Deep Learning Features 73

4.1. Introduction 73

4.2. Materials and Methods 77

4.2.1. Study Protocol and Data Acquisition 77
4.2.2. Preprocessing 78
4.2.3. Clinical Features 79
4.2.4. Deep Learning and Visualization Framework 80
4.2.4.1. System Overview 80
4.2.4.2. Model Training 82
4.2.5. Visualization of Network Decision Basis 84
4.3. Results 85
4.4. Discussion and Conclusion 91
Chapter 5
A Self-Attention-Guided 3D Deep Residual Network with Big Transfer to Predict Local Failure in Brain Metastasis after Radiotherapy using Multi-Channel MRI 94
5.1. Introduction 94
5.2. Methods and Procedures 98
5.2.1. Data Acquisition 98
5.2.2. Preprocessing 99
5.2.3. Network Overview 100
5.2.4. Big Transfer and Training Details 103
5.2.5. Visualization of Network Decision Basis 105
5.3. Results 107
5.4. Discussion and Conclusion 113
Chapter 6
Conclusion and Future Work 115
6.1. Conclusion 115
6.2. Limitations and Future Work 118
References 121
Appendices 137

List of Tables

Table 2.1- Dice similarity coefficient (DSC), Hausdorff distance (HD), and volume estimation error (VEE) for segmentation of brain metastasis at the baseline and follow-up scans using different network architectures. 41

Table 2.2- The 95% confidence intervals associated with the results in Table 2.1 The confidence intervals were estimated for the dice similarity coefficient (DSC), Hausdorff distance (HD), and the volume estimation error (VEE) based on the sample mean and standard error of the mean in each experiment using the t test. 42

Table 2.3- Experimental results of the proposed segmentation framework with different hyperparameters. DSC: dice similarity coefficient; CI: 95% confidence interval 42

Table 2.4- Dice similarity coefficient (DSC), Hausdorff Distance (HD), and Volume Estimation Error (VEE) for segmentation of brain metastasis various sizes (measured at baseline) using different network architectures. the results are associated with the patients of test set. 43

Table 2.5- The 95% confidence intervals associated with the results reported in Table 2.4. The confidence intervals were estimated for the dice similarity coefficient (DSC), Hausdorff distance (HD), and the volume estimation error (VEE) based on the sample mean and standard error of the mean in each experiment using the t distribution. 44

Table 2.6- Results of statistical comparison on the percental tumour size changes at each follow-up session relative to the baseline obtained from the automatically generated segmentation masks compared to those from the ground-truth masks. 45

Table 2.7- Results of detecting tumour size status at follow-up sessions after SRT for the patients of test set using different segmentation models. 45

Table 2.8- Results of detecting RANO-BM response categories at individual follow-up scans for the patients of test set using different segmentation models, validated based on the response categories identified from the ground-truth segmentation masks. 46

Table 2.9- Results of detecting the LC/LF and are outcomes for the patients of test set based on the RANO-BM criteria using different segmentation models. Sens: sensitivity; Spec: specificity; Acc: accuracy. 46

Table 3.1- Dice similarity coefficient (DSC), the 95% confidence interval (CI) of DSC, Hausdorff distance (HD), and volume estimation error (VEE) for segmentation of brain metastasis using different network architectures. 65

Table 3.2- List of the top seven selected features using the mRMR feature selection method for different segmentation models. Bold features are the shared ones with those obtained using the ground-truth segmentation masks..... 66

Table 3.3- Results of therapy outcome prediction using the radiomic models developed with different segmentation modules. 68

Table 4.1- Patient Characteristics..... 86

Table 4.2- Results of therapy outcome prediction in terms of accuracy, sensitivity, specificity, and AUC for different models..... 87

Table 4.3- Experimental results of the proposed framework on Validation Set. For the input of LSTM, Seq2Seq, and Transformer networks, the extracted features from the best InceptionResNet were concatenated with clinical features and then fed to these networks. 88

Table 5.1- Results of radiotherapy outcome prediction for different models. Acc: Accuracy; Sens: sensitivity; Spec: specificity. 110

Table 5.2- Experimental results of the proposed framework on the validation set for 3D residual network, 3D residual network + CBAM attention, and 3D residual network + self-attention with different hyperparameters and pretraining settings. 112

List of Figures

Figure 1.1- An illustration of the process of metastasis [11]. Red lymph nodes represent the formation of metastatic tumour cell lines at primary locations such as the breast, lung, and skin (melanoma). Metastasis from these initial locations spreads to the brain via the vascular system (red arrows), as well as to nearby sites such as the liver, bone, lungs, and lymph nodes (black arrows). 2

Figure 1.2- MRI scans are usually acquired at multiple sessions before and after the treatment for patients diagnosed with brain metastasis, Tumours are better visible on T1-weighted MRI (left), while edema is better seen on T2-FLAIR imaging. 7

Figure 1.3- Gross tumour volume (GTV, red contour) and planning tumour volume (PTV, green contour) on, left to right, contrast-enhanced T1-weighted, T2-FALIR, and CT images. GTV is the extent of the tumour visible on imaging and PTV allows for uncertainties in planning for treatment delivery. The CT image is used for radiation simulation and dose delivery planning as it provides patient- and tissue-specific attenuation and electron density information. 8

Figure 1.4- A series of contrast-enhance T1-weighted images acquired at the baseline (pre-treatment) and five follow-up sessions after radiotherapy from a patient diagnosed with brain metastasis. The tumour shows an initial increase in size on the first follow-up image (F1) but diminishes in size on the second follow-up image (F2), which is a sign of adverse radiation effect. While on the third follow-up image (F3) the tumour size remains stable, it increases on the fourth (F4), and again, on the fifth follow-up image (F5), indicting a local failure outcome in the last patient follow-up. 9

Figure 1.5- Segmentation results using OC-SVM for five patients (rows 1-5): (a) the contrast-enhanced T1-w image, (b) T2-FLAIR image, (c) predicted tumour mask from the OC-SVM model trained on the T1w image, (d) predicted lesion mask from the model trained on T2-FLAIR image, (e) the ground-truth mask (blue), final segmentation mask (red) and the overlap region of the two masks (purple) [61]. 14

Figure 1.6- Schematic of a radiomic framework. Image segmentation (either manual or automatic) is required to delineate regions of interest on the acquired images. The radiomic features are extracted from the images within the segmented regions of interest. Finally, analysis is done on radiomic features based on the target of the study. 17

Figure 1.7- Schematic of a general deep learning architecture for medical imaging analysis (classification). Usually, the feature extraction and processing, as well as the classification are done in a single network. 19

Figure 2.1- (a) Overview of the proposed segmentation framework. For a volumetric input image (contrast-enhanced T1-weighted, 512×512×128 voxels), first all slices are fed to a 2D UNet one by one. The generated masks from the 2D UNet are used to finds an approximate tumour position *(x, y)*. The volumetric input image is then cropped around *(x, y)* into a 256×256×128 voxel volume. A similar procedure is performed to reduce the size of volumetric image containing the tumour to 128×128×128 voxels. This volume is then fed into a 3D UNet for segmentation with no patching. The slices of this volume are also fed to MSGA, after concatenation with the co-registered T2-FLAIR image, and the segmentation masks of MSGA are then fused with those of 3D UNet. (b) The MSGA network structure. Features extracted at different scales from Resnet-101 are concatenated and convolved and then self-concatenated and fed into guided attention module. The resulting self-guided features are fed into the guided loss. 31

Figure 2.2- Training and validation loss of the 3D UNet and MSGA network in the proposed segmentation framework. 35

Figure 2.3- Contrast-enhanced T1-weighted images acquired at the baseline (1), and the first (2), second (3), and third (4) follow-ups after SRT from three representative patients with brain metastasis demonstrating local control (a), local failure (b), and ARE (c) after treatment. The arrow in the baseline image shows the location of brain metastasis. LC/LF/ARE is evaluated based on the changes in longest diameter. In (c) an initial growth in first follow-up is followed by a decrease in the second, and then third follow-ups. 39

Figure 2.4- Tumour segmentation masks generated by the cascaded 2D UNets (1), 3D UNet (2), cascaded 2D & 3D UNets (3), and Cascaded 2D & 3D UNets + MSGA (4) for five representative patients (a-e) in the test set. The images are acquired at the baseline. The arrow in the first image of each row shows the location of brain metastasis. The ground-truth and automatic segmentation masks (middle slide) are shown in blue and red, respectively, with the purple area showing the overlap region in each case. 40

Figure 2.5- Kaplan-Meier curves for comparative time to event analysis between the clinical radiotherapy outcome assessment and the assessment performed by the proposed automatic system on the test set. The plots have been shown for the LF (left) and ARE (right) events. The time to event for each tumour was calculated from the date of radiotherapy to the date an LF/ARE was detected clinically or by the automatic system using the proposed segmentation framework. 47

Figure 3.1- The overall system for SRT outcome prediction. The segmentation module segments tumours on the baseline (BL) and the first follow-up (FU1) MR images. Using the MR images and segmentation masks, different radiomic features are extracted. The features are then sorted based on their relevance/redundancy and a subset of them identified through feature selection are used to train a classifier for predicting the local outcome. 61

Figure 3.2- Representative parametric maps of the top four selected features based on the ground-truth (Table XVI) overlayed on the associated MR images acquired at the baseline and the first follow-up from a representative tumour with an LF outcome. The parametric maps are shown for the ground-truth mask and the segmentation masks generated automatically using different models. Δmean for each feature is the mean relative change from the baseline at the first follow-up that is calculated for different masks. 67

Figure 3.3- Kaplan-Meier progression-free survival curves for two cohorts of patients stratified based on the outcome prediction by the radiomic models developed with different segmentation modules: (a) Cascaded 2D UNets, (b) 3D UNets, (c) Cascaded 2D & 3D UNets, (d) Cascaded 2D & 3D UNets + MSGA, and (e) Ground-truth masks. Cohort 2 includes the patients in the independent test set who had at least one lesion with a predicted outcome of local failure, and cohort 1 includes all other patients in the independent test set. 69

Figure 4.1- System overview. (a) Slice by slice feature extraction from MRI volumes using the InceptionResnetV2, (b) a two-layer LSTM, (c) a Seq2Seq network, and (d) a transformer network. The input feature vectors for all three networks are a concatenation of the features coming from the InceptionResNet and the clinical features..... 81

Figure 4.2- The ROC curves on the independent test set for different predictive models. 88

Figure 4.3- Kaplan-Meier progression-free survival curves for two cohorts of patients stratified at the baseline based on the outcome prediction by different models: (a) MLP with clinical features only, (b) InceptionResNet, (c) InceptionResNet + LSTM, and (d) InceptionResNet + LSTM with clinical feature fusion. Cohort 2 includes the patients in the independent test set who had at least one lesion with a predicted outcome of local failure, and cohort 1 includes all other patients in the independent test set. 89

Figure 4.4- Treatment-Planning MRI and deep learning visualization heatmaps acquired for four representative patients with metastatic lesions treated with SRT, two with an LC and the other two with an LF outcome. The heatmaps are overlaid on the ROIs input to the InceptionResNet and demonstrate the impact level of different areas on each image for the network's prediction, as calculated through a modified PDA. The network predicted the outcome of all four lesions correctly. 90

Figure 5.1- Architecture of (a) vanilla 3D residual network (baseline), (b) the 3D residual network with CBAM attention, (c) the proposed self-attention-guided 3D residual network, (d) the residual block in 3D residual network, consisting of residual connections, (e) the CBAM attention block consisting of the channel and spatial attention modules, and (f) the self-attention block consisting of the key, query, and value tensors that generates the final self-attention feature tensor. 101

Figure 5.2- The procedure for creating explorable 3D brain models and visualization heatmaps from a set of individual slices. (a) initially, the coordinates (x, y, z) of each voxel center in the MRI volume is determined. Since the number of slices is often limited, the resulting point cloud consists of multiple clusters with the same z and different x and y which is visually undesirable. To mitigate this issue, the points between slices are randomly interpolated, (b) the point cloud after the inter-slice interpolation, (c) the 3D brain model after assigning an intensity to each point in the point cloud and surface reconstruction, (d) applying the same procedure to generate a color-coded 3D visualization heatmap of importance for a lesion within the brain. 107

Figure 5.3- Training loss of the three models investigated in this study before and after applying BiT training scheme. Following the BiT training recipe generally led to faster convergence, smaller loss, and better performance overall. 108

Figure 5.4- The ROC curves for (a) vanilla 3D residual network, 3D residual network + CBAM, and 3D residual network + self-attention, and (b) the same models trained with the BiT scheme. 110

Figure 5.5- 3D visualization heatmaps corresponding to the two input channels (T1w and T2-FLAIR) of the 3D residual network with self-attention and BiT training for two representative lesions, one with an LF (top) and the other one with an LC (bottom) outcome. The user can inspect any desired area on the lesion/margin surface or inside the volumetric ROI and their correspondence with the MRI channels. 112

List of Appendices

Appendix A: Details of Deep-Learning Modules and Architectures……..137

List of Acronyms and Abbreviations

2D	Two-dimensional
3D	Three-dimensional
AI	Artificial Intelligence
ARE	Adverse Radiation Effect
AUC	Area Under Curve
BERT	Bidirectional Encoder Representations from Transformers
BiT	Big Transfer
CBAM	Convolutional Block Attention Module
CNN	Convolutional Neural Network
CNS	Central Nervous System
CT	Computed Tomography
DNA	Deoxyribonucleic Acid
DSC	Dice Similarity Coefficient
FLAIR	Fluid Attenuation Inversion Recovery
GBM	Glioblastoma Multiforme
GLCM	Gray-level Co-Occurrence Matrix
GPA	Graded Prognostic Assessment
GPT-3	Generative Pre-trained Transformer 3
GTV	Gross Tumour Volume
HD	Hausdorff Distance
LC	Local Control
LF	Local Failure
LSTM	Long Short-term Memory
MIQ	Mutual Information Quotient
ML	Machine Learning
MLP	Multi-Layer Perceptron
MRI	Magnetic Resonance Imaging/Image

mRMR	Minimum Redundancy Maximum Relevance
MSGA	Multi-scale Self-guided Attention
NSCLC	Non-small Cell Lung Cancer
PDA	Prediction Difference Analysis
PTV	Planning Tumour Volume
qMRI	Quantitative Magnetic Resonance Imaging
RANO	Response Assessment in Neuro-oncology
RELU	Rectified Linear Unit
RNN	Recurrent Neural Network
ROI	Region of Interest
RPA	Recursive Partitioning Analysis
RT	Radiation Therapy
Seq2Seq	Sequence to Sequence
SGD	Stochastic Gradient Descent
SNR	Signal-to-Noise Ratio
SRS	Stereotactic Radiosurgery
SRT	Stereotactic Radio Therapy
SVM	Support Vector Machines
T1w	T1-weighted
US	Ultrasound
VEE	Volume Estimation Error
WBRT	Whole-brain Radiation Therapy

Chapter 1

Introduction

1.1. Brain Metastasis: Background

Occurring in up to 30% of all cancer patients, brain metastases are the most common malignancy of the central nervous system (CNS) [1], with an increase in the number of incidents due to improvements in controlling systemic disease, enhanced diagnostic tools, and prolonged survival of cancer patients [2]. The estimated incidence of new brain metastases in the United States is between 7–14 per 100,000 persons based on population studies [3]. The prevalence of metastases to the brain is expected to continue to increase in the future [2].

Metastatic brain tumours represent an important cause of morbidity and mortality in cancer patients. Whereas a significant proportion of cancer patients survive for many years if the cancer is identified at an early stage while it is still localized [4]–[7], when the tumour is metastasized to the brain, the median survival ranges from as short as 5 months to up to 4 years, based on the subgroup and origin of cancer [8]. Early diagnosis and precise treatment of brain metastasis may lead to the reduction of brain symptoms and enhance the quality of life and survival of the patients [8]–[10].

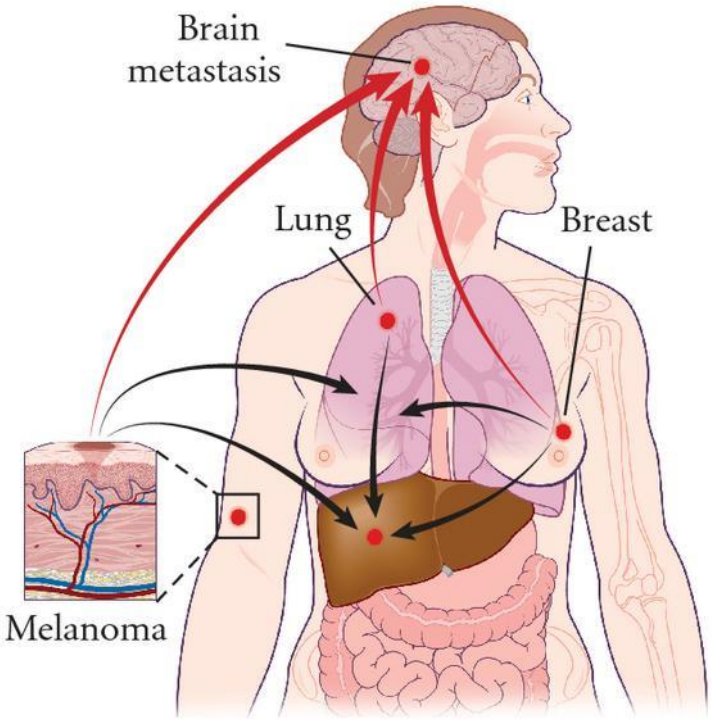

Figure 1.1- An illustration of the process of metastasis [11]. Red lymph nodes represent the formation of metastatic tumour cell lines at primary locations such as the breast, lung, and skin (melanoma). Metastasis from these initial locations spreads to the brain via the vascular system (red arrows), as well as to nearby sites such as the liver, bone, lungs, and lymph nodes (black arrows).

As discussed above, while in primary brain cancer the tumours cells start to grow inside the brain, in brain metastasis, cancer cells initiate in another part of the body, *e.g.*, lung, breast, and melanoma, and then migrate to the brain (Figure 1.1). Brain metastasis may present as single or multiple tumours. Clinical records indicate that approximately 29% of brain metastases arise as single tumours, 35% as 2-3 tumours, and 36% as more than three tumours [12].

A population-based study by Cagney *et al.* has provided generalizable estimates of the incidence and prognosis for patients with brain metastases in the United States [6]. Out of 1,302,166 patients diagnosed with nonhematologic malignancies originating outside of the CNS between 2010 and 2013, 26,430 patients were identified with brain metastasis. Patients diagnosed

with small cell lung cancer (SCLC) (16%) and lung adenocarcinoma (14%) had an incidence proportion of more than 10%. Patients with prostate cancer, bronchioloalveolar carcinoma, and breast cancer as the primary cancer had the longest median survivals (12, 10, and 10 months, respectively). Several studies by Sperduto *et al.* on patients with brain metastasis have also reported overall median survival times for non-small cell lung cancer (NSCLC), SCLC, melanoma, and breast cancer as 12, 5, 10, and 16 months respectively, where higher clinical prognostic scores were associated with longer survival time, *e.g.*, as long as 4 years for NSCLC [7], [13], [14].

1.2. Treatment Options

Reduced brain symptoms, improved quality of life, and increased survival rates may be achieved for cancer patients through early detection and precise treatment of brain metastasis [2]. Many factors including the origin of cancer, associated symptoms, and size/number of metastases are considered in planning a treatment strategy for the patient. Radiation therapy, chemotherapy, immunotherapy, and surgery are the main treatment options for the management of metastatic brain tumours. Available options for radiotherapy include whole-brain radiation therapy (WBRT), hypo-fractionated stereotactic radiotherapy (SRT), and single-fraction stereotactic radiosurgery (SRS).

Important factors in deciding whether a patient should proceed with surgical resection of brain metastasis include the accessibility and size of the tumour, its relative proximity to eloquent brain areas, the degree of mass effect, the patient's age, and presence of other extracranial diseases [15]. Surgical resection is mainly recommended for patients with single brain metastasis in an accessible location when the tumour is large (> 3-4 cm) and/or causing a considerable mass effect

[16]. Neurosurgical intervention is also used in case a pathological diagnosis is required [17]. In addition, surgery is preferred in the presence of significant peri-tumoural edema to relieve and potentially reverse the associated neurological complications [18]. In most cases, surgical resection of the tumour is combined with postoperative (adjuvant) radiation therapy. For the treatment of a single metastatic tumour, a combination of surgical resection and adjuvant radiotherapy has been shown to outperform radiotherapy alone [19]. In a randomized trial, Patchell *et al.* investigated whether postoperative radiotherapy can increase survival and improve the neurological control of the disease [19]. Ninety-five patients participated in this study and the outcome demonstrated that the recurrence of metastasis anywhere in the brain was less frequent in the radiotherapy group compared to the observation group (18% in radiotherapy group compared to 70% in the observation group). In a more recent randomized trial, Mahajan *et al.* investigated whether SRS after the surgical resection of brain metastasis improves the time to local recurrence [20]. 132 patients were randomly assigned after surgery to the observation (n=68) or cavity SRS (n = 64). The findings of the study indicate that in patients who had complete resection of one, two, or three brain metastases (the majority of the study population had one metastasis), adjuvant SRS significantly decreased local recurrence compared to the observation cohort. A multicenter randomized controlled phase 3 trial by Brown *et al.* studied the efficacy of postoperative SRS compared to WBRT for resected brain metastasis [21]. 194 adult patients from 48 institutions with at least one resected brain metastasis of less than 5 cm in maximum diameter were randomly assigned to either postoperative SRS (n=98) or WBRT (n=96). Cognitive-deterioration-free survival was longer in patients assigned to SRS (median = 3.7) compared to WBRT (median= 3).

The median survival rate was 12.2 months for the SRS and 11.6 months for WBRT cohorts, respectively.

For WBRT, the radiation dose is delivered to the whole brain in 5-20 fractions over a period of 1-4 weeks. Common fractionation schedules (total radiation dose in Gy / number of fractions) include 20/5, 30/10, 37.5/15, 40/20. Conventionally, WBRT has been the treatment of choice for patients with multiple brain metastases [22]. WBRT, however, is associated with adverse side effects which typically include mild fatigue, mild dermatitis, temporary alopecia, and otitis media or externa [22]. Multiple clinical trials have shown cognitive deterioration after WBRT [21], [23], [24]. Apart from the detrimental effects of WBRT, some studies have also shown no benefit with respect to survival [25]. Most recently, the QUARTZ trial in patients with non-small-cell lung cancer, brain metastasis, and poor performance status observed no survival advantage following WBRT versus best supportive care (median survival of 9.2 weeks for patients who receive optimal supportive care plus WBRT and 8.5 weeks for patients who only receive optimal supportive care) [25]. It should be noted that despite the evidence against WBRT, this treatment is still recommended in certain situations such as when patients present with innumerable metastases and/or diffuse leptomeningeal or pachymeningeal disease.

There has been a gradual shift in the past 2 decades away from WBRT to stereotactic radiosurgery (SRS), particularly in patients with limited brain metastases. SRS delivers a focused ablative (intense-dose) radiation treatment with sub-millimeter precision to the tumour localized in three-dimensions in a single fraction. This treatment modality is usually used as the sole treatment for patients who have less than four metastatic brain tumours [26]. SRS demonstrated to have lower toxic effects over the non-tumour areas within the brain compared to WBRT [27].

However, several studies have shown that using WBRT reduces the risk of development of new brain metastasis compared to SRS [27].

In some cases, due to the large size or location of the tumour applying the prescribed radiation dose in a single fraction is not recommended. In such cases, high-dose radiation is frequently delivered in very few fractions using the same precisely targeted method. This approach is known as hypo-fractioned stereotactic radiation therapy (SRT). Because of the number of SRT fractions that typically range from 3 to 5, a thermoplastic mask, or a more sophisticated non-invasive mask system, is used as the immobilization device and the precision of therapy is coupled to image-guidance when using such systems as compared to an invasive frame.

1.3. Imaging in Brain Metastasis Management

Brain malignancies are often detected with the help of imaging. As a part of the standard of care, imaging is used to detect metastases in patients with known primary malignancies, neurological symptoms, or even for asymptomatic patients when there is a strong clinical suspicion of metastatic presence. It is also used for screening CNS involvement in patients diagnosed with cancer. Computed tomography (CT) and magnetic resonance imaging (MRI), mainly (contrast-enhanced) T1-weighted, T2-weighted, and T2-weighted-fluid-attenuation-inversion-recovery (T2-FLAIR) imaging (Figure 1.2), are two main imaging modalities applied in the diagnosis and management of brain metastasis.

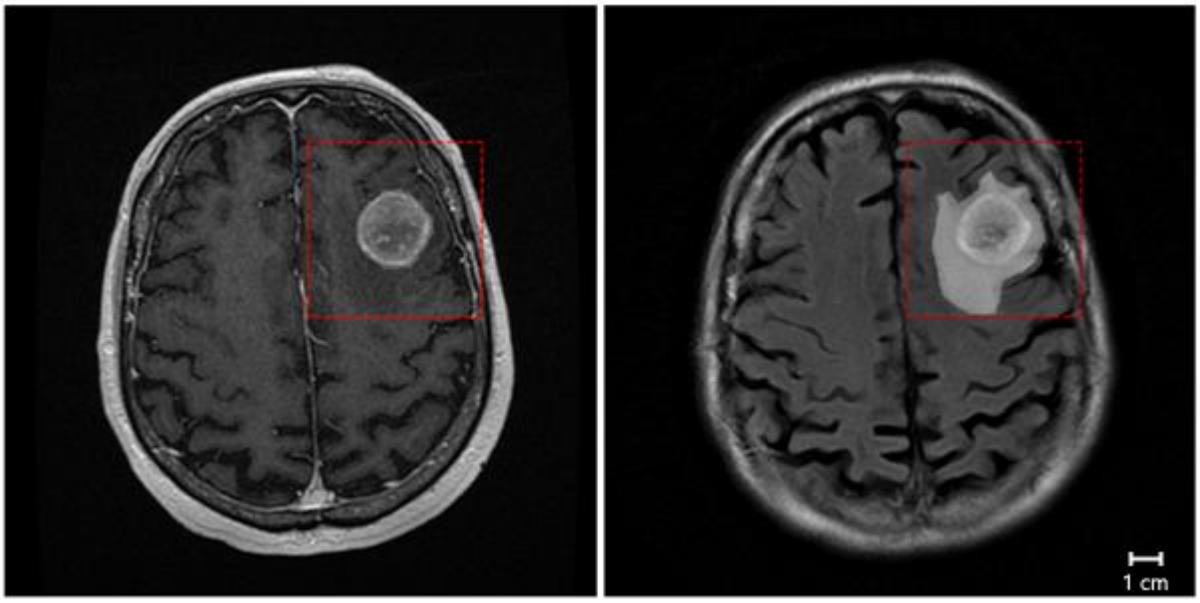

Figure 1.2- MRI scans are usually acquired at multiple sessions before and after the treatment for patients diagnosed with brain metastasis, Tumours are better visible on T1-weighted MRI (left), while edema is better seen on T2-FLAIR imaging.

The images acquired before, and after the treatment of brain metastasis are used clinically for treatment planning and evaluating the treatment outcome through post-treatment follow-ups. There are specific regions within and around the tumour on imaging that are of particular interest for treatment planning and outcome assessment. Contouring those regions of interest (ROIs) within images is an initial step in these procedures (Figure 1.3). To avoid the laborious task of manually contouring lesions on several 2D planes of 3D MRI acquired for each patient at multiple scan sessions, an automatic yet accurate ROI segmentation framework is highly desirable in these applications.

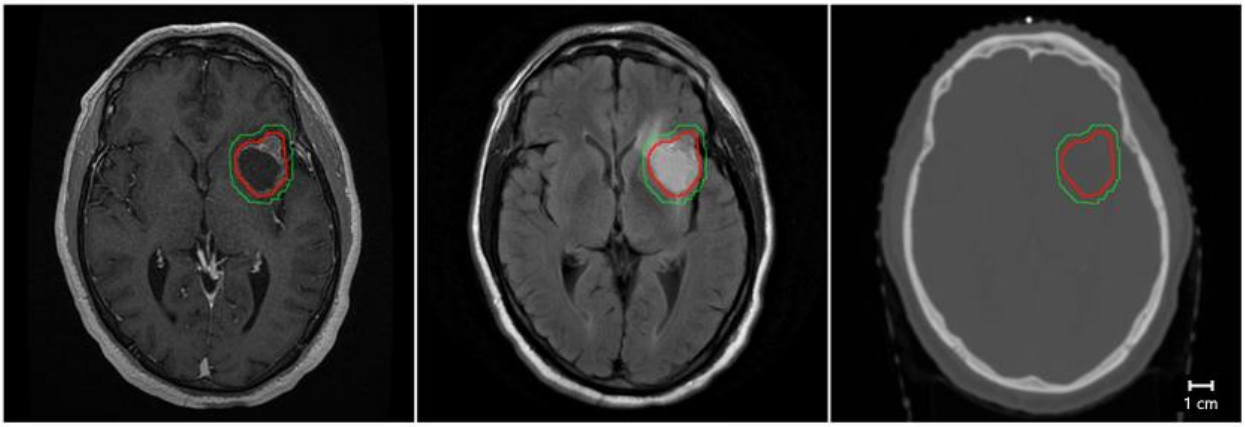

Figure 1.3- Gross tumour volume (GTV, red contour) and planning tumour volume (PTV, green contour) on, left to right, contrast-enhanced T1-weighted, T2-FALIR, and CT images. GTV is the extent of the tumour visible on imaging and PTV allows for uncertainties in planning for treatment delivery. The CT image is used for radiation simulation and dose delivery planning as it provides patient- and tissue-specific attenuation and electron density information.

1.4. Therapy Response Evaluation

As mentioned above, assessing tumour response to radiotherapy is performed using serial imaging acquired before (baseline) and at multiple follow-up sessions after the treatment. Evaluation of radiotherapy outcome in brain metastasis on serial MRI is mainly performed based on the standard criteria presented by the response assessment in neuro-oncology–brain metastases (RANO-BM) group [28]. The RANO-BM criteria are principally based on changes in the longest diameter of the target tumour in the axial, coronal, and sagittal planes compared to baseline or nadir (smallest tumour size on the previous scans) to specify its response to therapy. The four categories of therapy response based on the RANO-BM criteria include complete response (CR; no target tumour remaining), partial response (PR; more than 30% reduction in the longest diameter compared to baseline), stable disease (SD; less than 30% decrease compared to baseline but also less than 20% increase in the longest diameter compared to nadir), or progressive disease (PD; more than 20% increase in the longest diameter compared to nadir). Tumour enlargement on MRI after radiotherapy may also become apparent due to adverse radiation effect (ARE), also referred to as

pseudo-progression, that should be differentiated from progressive disease (local progression), as quite different therapies should be administrated for them. Evident tumour enlargements on MRI due to ARE often become stable or followed by a decrease in tumour size on subsequent imaging follow-ups. The local response to radiation therapy is often classified as either local control (LC; stable or shrinking tumour that is indicative of a stable disease, partial response or complete response,) or local failure (LF; enlarging tumour associated with a progressive disease) based on tumour size changes on follow-up structural serial imaging [28]. Figure 1.4 demonstrates the serial MRI acquired from a patient diagnosed with brain metastasis, showing different patterns of change in tumour size on follow-up images.

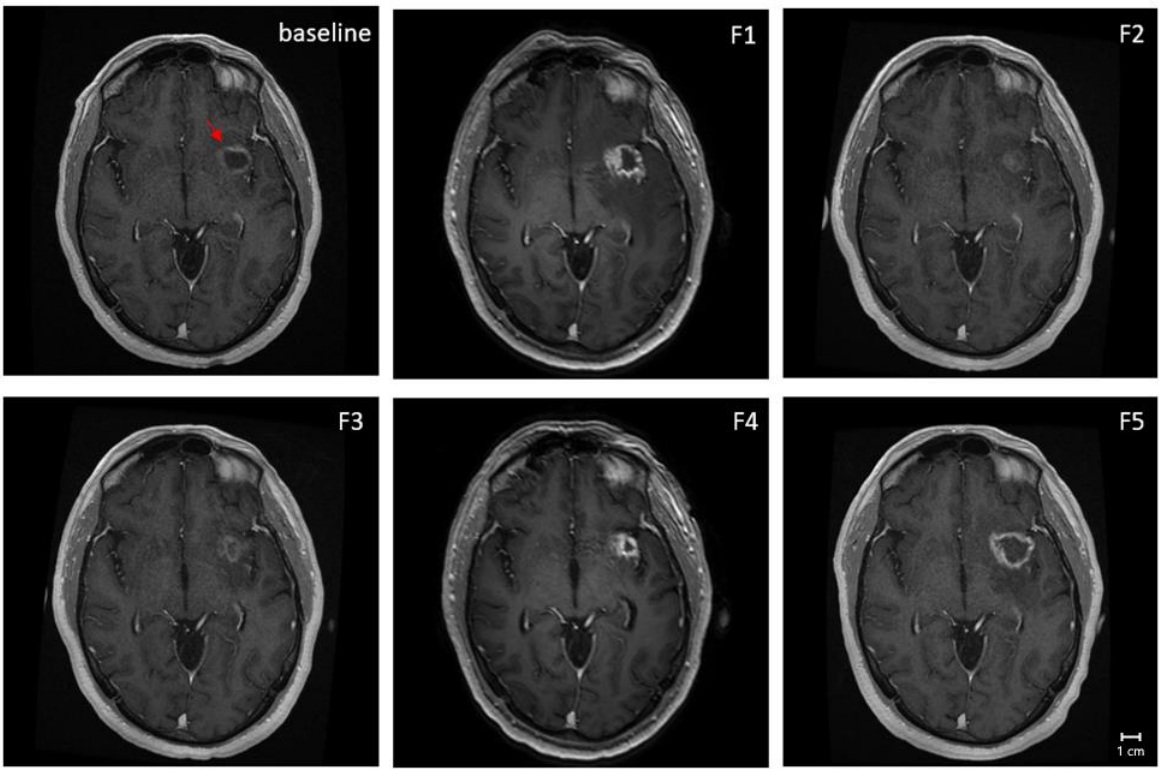

Figure 1.4- A series of contrast-enhance T1-weighted images acquired at the baseline (pre-treatment) and five follow-up sessions after radiotherapy from a patient diagnosed with brain metastasis. The tumour shows an initial increase in size on the first follow-up image (F1) but diminishes in size on the second follow-up image (F2), which is a sign of adverse radiation effect. While on the third follow-up image (F3) the tumour size remains stable, it increases on the fourth (F4), and again, on the fifth follow-up image (F5), indicting a local failure outcome in the last patient follow-up.

1.5. Prognostic Factors and Biomarkers for Risk Assessment

Tumour response to treatment is variable in patients with brain metastasis despite similar cancer therapy. There are many patient-related factors contributing to therapy response including genetics, age, nutrition, health status, environmental exposure, and epigenetic factors [29]. Predicting the efficacy of a treatment for individual patients before, or early after, therapy initiation could lead to improvements in prognosis and clinical outcome.

While response to radiation treatment in brain metastasis is evaluated based on changes in tumour size using serial imaging, it could take months for a local response to be visible on follow-up scans. Given that the median survival of these patients following radiotherapy can be between 5 months and 4 years [14], *a priori* risk assessment in terms of LF for tumours receiving radiation therapy potentially permits effective adjustments in treatment that result in enhanced therapy outcomes, patients' survival and their quality of life [2].

As the treatment outcomes are vastly different, an essential step in clinical trials is to classify patients based on measurable prognostic scores. Recursive partitioning analysis (RPA), a statistical method for multivariant analysis, was one of the first methods used to prognosticate patients into separate classes based on age, performance status, control of primary and extent of extracranial disease [30]. It was a major development at the time but overly simplistic, RPA is now replaced by more sophisticated classification methods such as diagnosis-specific graded prognostic assessment (DS-GPA) [31]. Proposed by Sperduto et al. [32], significant prognostic factors were used to define the DS-GPA with a GPA of 4.00 and 0.0 for the best and worst prognosis, respectively. The prognosis factors vary based on the diagnostic information regarding

the primary cancer. For example, lung cancer prognostic factors included the Karnofsky performance score (an index which classifies patients based on their functional impairment [33]), age, presence of extracranial metastases, and the number of brain metastases, in addition to the presence of epidermal growth factor receptor (EGFR) and anaplastic lymphoma kinase (ALK) mutations in patients with adenocarcinoma [34]. Similarly, for Gastrointestinal cancers, prognostic factors included the Karnofsky performance score, age, extracranial metastases, and the number of brain metastases [5].

1.5.1. Biomarkers of Cancer Therapy Response

Biological markers are any measurable indicators of the severity or presence of a disease. Biomarkers are formally defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or biological responses to a therapeutic intervention. [35]." Different types of biomarkers can potentially provide beneficial diagnostic and prognostic information for therapy response prediction and, in combination with available therapeutic options, can pave the way for the paradigm of personalized medicine.

The methodologies for cancer therapy response prediction could be either invasive or non-invasive depending on the type of applied biomarkers. Important invasive biomarkers of cancer therapy response include those obtained via biopsy or blood samples, including the histopathologic, molecular, genomic, and proteomic data, or circulating tumour DNA (ctDNA) [36]–[38]. One limitation associated with these biomarkers is that they are acquired through, more or less, invasive approaches that in some cases are challenging or even technically not feasible. Further, processing technologies required for the genomic and proteomic biomarkers are not always accessible. Another limitation is that these biomarkers may not be robust as they may not

represent the entirety of the tumour due to the heterogeneity and high degree of genomic diversity within the tumour [39].

Non-invasive biomarkers are often derived from images or raw data acquired using different imaging modalities [40]. With the advancements in imaging technologies and improvements in standard equipment and protocols, the development of methodologies to extract, standardize, and mine non-invasive biomarkers for cancer characterization and therapy response prediction have become the focus of many works in the literature [41]–[44]. Some of these works are reviewed in Section 1.6.3.

1.6. Literature Review

Advancements in machine learning have motivated many studies to investigate its application for solving various problems in different areas from engineering to medical sciences. Recent studies show the potential of machine learning algorithms in medical image segmentation [45]–[48], therapy outcome assessment [49], [50], and treatment outcome prediction [51]–[55] for different cancer sites. In this section, recent studies on brain image segmentation, automated radiotherapy outcome assessment, and imaging-based cancer therapy outcome prediction are presented.

1.6.1. Brain Image Segmentation

Brain lesion segmentation aims to separate regions of different abnormalities such as tumour, and edema from normal brain tissues, *i.e.*, White matter (WM), Gray matter (GM), and Cerebrospinal Fluid (CSF). Due to its clinical utility as part of the standard of care, and the good contrast provided by MRI for such abnormalities, brain lesion segmentation on MRI images has gained popularity in the literature in recent years.

Conventional segmentation algorithms for this task include simple threshold-based techniques, *e.g.*, skull stripping by thresholding [56], region-based techniques such as region growing [57] and watershed [58], model-based techniques such as level sets [59], and pixel classification techniques such as clustering [60]. Jalalifar et al. have developed an outlier-detection-based segmentation framework to delineate brain tumours in magnetic resonance (MR) images automatically [61]. The proposed method considers the tumour and edema pixels in MR images as outliers compared to the pixels associated with the healthy tissue. The texture characteristics of tumour and edema regions are often different from healthy tissues in contrast-enhanced T1-weighted and T2-FLAIR images. The framework applies such dissimilarity to detect the tumour and lesion regions automatically and utilizes morphological and logical operators to generate a fused tumour mask. More specifically, the framework generates two outlier masks using independent one-class support vector machines (OC-SVMs) that operate on contrast-enhanced T1-weighted and T2-FLAIR images. The outlier masks are subsequently refined and fused using morphological and logical operators to generate a tumour mask for each image slice. Representative results for five patients are shown in Figure 1.5.

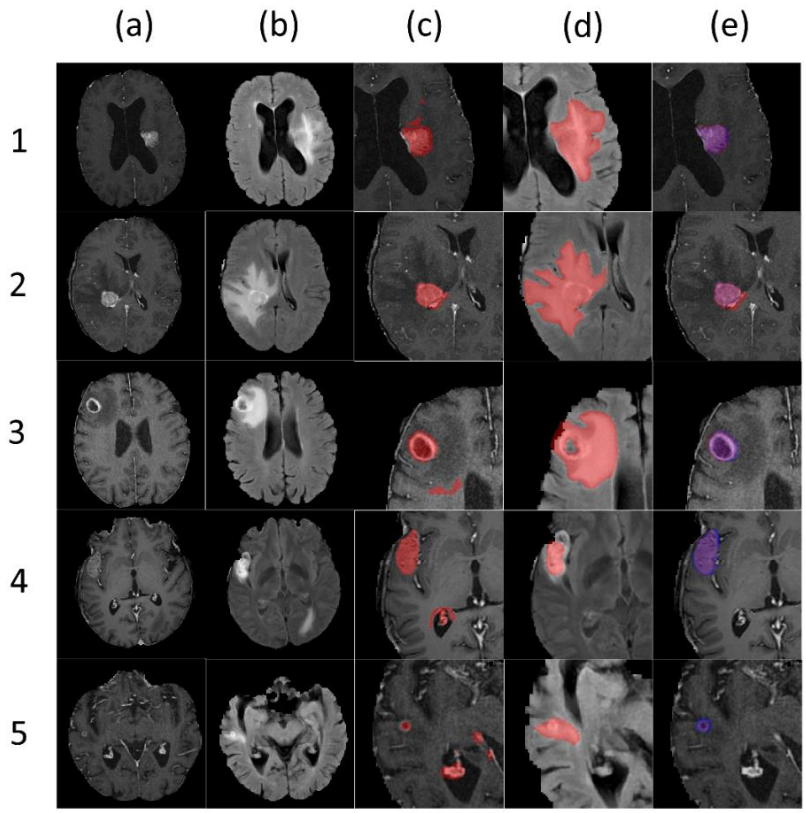

Figure 1.5- Segmentation results using OC-SVM for five patients (rows 1-5): (a) the contrast-enhanced T1-w image, (b) T2-FLAIR image, (c) predicted tumour mask from the OC-SVM model trained on the T1w image, (d) predicted lesion mask from the model trained on T2-FLAIR image, (e) the ground-truth mask (blue), final segmentation mask (red) and the overlap region of the two masks (purple) [61].

In recent years, with the advancement of deep learning, many convolutional neural network architectures have been introduced for the task of medical image segmentation. Among these architectures, 2D and 3D UNet [62], [63] gained widespread popularity because of their robustness on different modalities. 2D UNet has an encoder-decoder network structure for medical image segmentation consisting of 19 fully-convolutional layers, deconvolution layers [64], and skip connections to overcome the trade-off between localization and context. 3D UNet borrows the same idea from 2D UNet but unlike 2D UNet, it applies 3D operations and performs dense 3D segmentation on adjacent 2D slices. Other variants of UNet have also been introduced such as Attention UNet [65], a 2D UNet with attention gates to trim irrelevant features, Inception UNet

[66], in which inception layer are connected in the UNet fashion, and UNet++ [67], a 2D UNet with dense network of skip connections.

Encoder-decoder architectures such as UNet have the drawback of using a multi-scale approach which leads to redundancy since similar low-level features are extracted multiple times at multiple scales and also long-range feature dependencies are not efficiently modeled [43]. To mitigate this problem, Sinha et al. have proposed a multi-scale guided self-attention approach to integrate local features with their corresponding global dependencies, as well as adaptively highlight interdependent channel maps [43].

1.6.2. Automated Radiotherapy Outcome Assessment

Objective assessment of tumour response to therapy has been the basis for many investigations in cancer therapeutics during recent years [68]. The RANO-BM [28] criteria and recommendations were proposed to establish a basis for standard response assessment in clinical trials for brain metastasis. Few studies have recently investigated the potential of deep-learning-based methods in automatic brain tumour delineation followed by assessment of tumour size changes in response to treatment. Xue et al. [49] proposed a cascade of modified 3D UNet architecture for the detection and segmentation of brain metastases on 3D T1 MPRAGE images. They proposed the utility of automatically generated segmentation masks for facilitating radiotherapy treatment planning and post-treatment monitoring of tumour size, where they demonstrated example results for one case. Cho et al. [50] developed a CAD system for automated detection of brain metastasis on MRI using a UNet based cascaded model and applied it for categorizing tumour size changes at two follow-up sessions separately, where they achieved a moderate agreement with RANO-BM criteria. To our knowledge, no previous study has been performed to investigate the efficacy of deep-learning-

based segmentation frameworks for automatic radiotherapy outcome assessment comprehensively.

1.6.3. Quantitative Imaging for Therapy Outcome Prediction

A quantitative imaging biomarker can be defined as "an objective characteristic derived from an in vivo image measured on a ratio or interval scale as indicators of normal biological processes, pathogenic processes, or a response to a therapeutic intervention [69]." The tumour volume measured based on volumetric CT or MRI could be considered as an imaging biomarker for response monitoring, for example, to describe a patient response to therapy. Tissue radioactivity concentration in positron emission tomography (PET) scans is another example of imaging biomarkers [69]. Different quantitative imaging biomarkers have been utilized for diagnosis of abnormalities and tissue characterization, such as differentiation of benign versus malignant lesions in breast [70] or low- and high-grade malignancies in brain using various textural biomarkers derived from MRI [71]. In a study by Tran *et al.*, histopathological findings were predicted based on the characteristic proton spectra for 5 common adult supratentorial brain tumours: low-grade and anaplastic astrocytomas, glioblastoma multiforme (GBM), meningioma and metastasis [72].

Radiomic features are a prominent example of non-invasive imaging biomarkers, being explored to discover quantitative diagnostic and prognostic biomarkers through the high-throughput mining of high-dimensional medical imaging data [73]. In radiomic analysis, hundreds of quantitative features are extracted from an image and its transformed versions within different regions of interest including the intra- and peri-tumoural regions. Data mining is performed in conjunction with advanced statistical analyses on these high-dimensional data under the hypothesis

that an appropriate low-dimensional combination of these features, possibly along with clinical data, can represent crucial tissue properties, useful for diagnosis, prognosis, or treatment planning of individual patients [74]. A general schematic of the radiomic framework is shown in Figure 1.6.

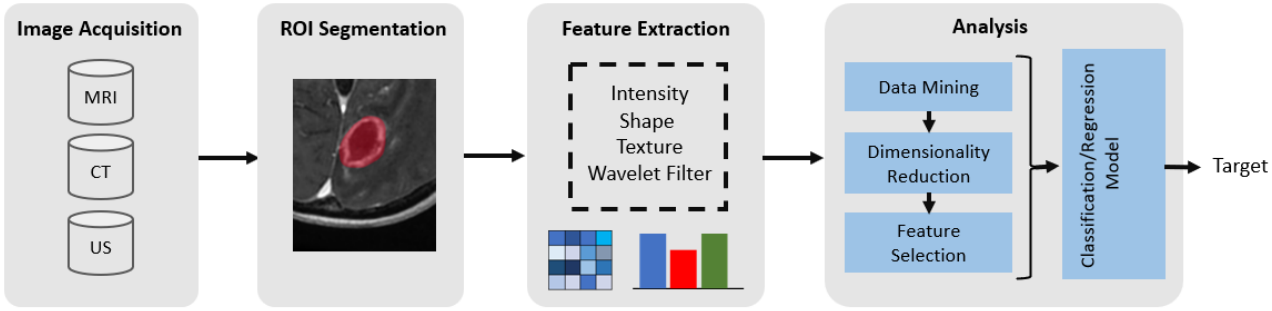

Figure 1.6- Schematic of a radiomic framework. Image segmentation (either manual or automatic) is required to delineate regions of interest on the acquired images. The radiomic features are extracted from the images within the segmented regions of interest. Finally, analysis is done on radiomic features based on the target of the study.

Recent studies demonstrate correlations between the radiomic signature of tumours and their phenotypes and genomic and proteomic profiles. Aerts *et al.* have presented a radiomic analysis of 440 features quantifying the intensity, shape and texture of tumours in CT imaging data of 1,019 patients with lung or head-and-neck cancer [75]. The results of that study suggest that radiomics are capable of identify a general prognostic phenotype existing in both lung and head-and-neck cancer, and that these features have prognostic power in independent datasets of lung and head-and-neck cancer patients. Extracted radiomic features from CT images of the lung and head and neck tumours also show correlations between radiomic features and underlying gene-expression patterns [75].

Considering the potential association of the radiomic features with tissue/tumour phenotypes and its genomic and proteomic profiles, several studies have explored the potential of radiomic features in conjunction with machine learning methods for risk stratification in cancer

patients and therapy response prediction. Bae et al. [76] investigated the effectiveness of MRI radiomic features for improving survival prediction in patients with GBM and achieved an AUC of 0.65. Karami et al. [77] have demonstrated that radiomic MRI biomarkers can be adapted to predict stereotactic radiotherapy outcome in brain metastasis using support vector machines. Mouraviev et al. [78] exploited the power of random forest classifier with radiomic features to predict the local control of brain metastasis after stereotactic radiosurgery and achieved an AUC of 0.79.

While radiomic methods applies handcrafted features usually with conventional machine learning techniques for prediction modeling, deep learning algorithms could be used to extract relevant distinguishing imaging features automatically. Deep learning models have proven to be quite effective at identifying important and distinct characteristics, particularly in image data [79], [80]. Deep models have the capability to identify regions of interest automatically, capture textural variations within a lesion, discriminate between cancerous and non-cancerous cells, and potentially extract distinctive information from lesions to be applied for the task of therapy outcome prediction [54], [55], [81]–[88]. Diamant et al. [53] hypothesized that convolutional neural networks could enhance the performance of traditional radiomics, by detecting image patterns that may not be covered by a traditional radiomic framework. They tested their hypothesis for the task of head & neck cancer therapy outcome prediction, where their results show that deep models can explicitly recognize traditional radiomic features and perform accurate outcome prediction. A recent study by Cho et al. [89] suggests that using deep learning methods instead of classic machine learning results in brain metastasis detection with a lower false-positive rate. Figure 1.7 shows the general scheme of deep learning-based framework for medical imaging classification.

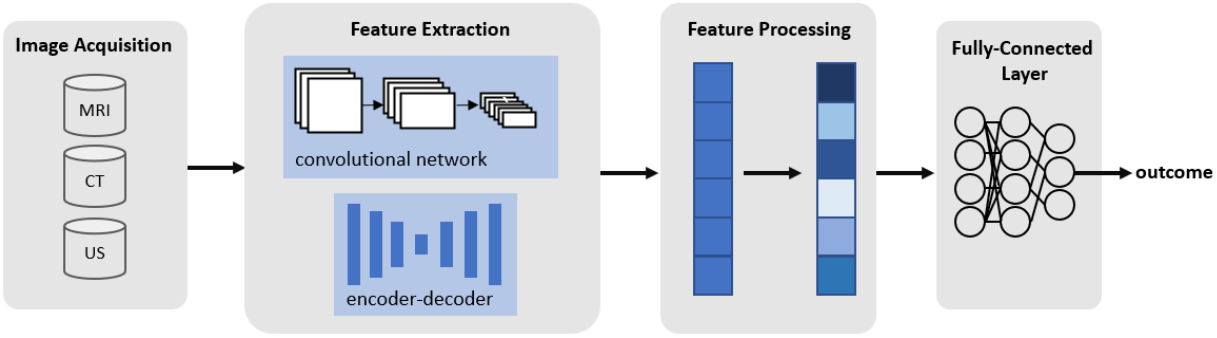

Figure 1.7- Schematic of a general deep learning architecture for medical imaging analysis (classification). Usually, the feature extraction and processing, as well as the classification are done in a single network.

1.7. Objectives

The overarching goal of this dissertation project is to develop, optimize and investigate novel machine learning and quantitative imaging methodologies that can eventually improve the clinical management of brain metastasis. This goal is perused through four objectives.

As the first objective, a novel deep-learning-based framework is proposed and investigated for automatic brain tumour segmentation on standard MRI. With a robust segmentation framework in hand, as the second objective, we introduce and comprehensively investigate an automatic system for assessing radiotherapy outcome in brain metastasis based on standardized criteria of changes in tumour size on serial MRI, *e.g.*, RANO-BM criteria. An automatic outcome assessment system, if approved after rigorous evaluations by regulatory bodies, can streamline the radiation therapy workflow in cancer centres considerably.

ROI segmentation is the heart of any automated system aiming on mining information form a lesion and its surrounding regions on imaging. With a variety of automatic segmentation models in hand, as the third objective in this dissertation, we investigate the impact of tumour segmentation

accuracy on the performance and robustness of hand-crafted radiomic biomarkers and associated models for radiotherapy outcome prediction in brain metastasis.

Predicting the outcome of prescribed anti-cancer therapies before or early after the treatment is highly desirable in the clinic and can pave the way for personalized medicine paradigm in cancer therapeutics. As the fourth objective of this dissertation project, we propose and investigate novel convolutional deep neural network architectures for predicting radiotherapy outcome in brain metastasis *a priori* (at baseline) using treatment-planning images, with and without standard clinical attributes available at pre-treatment. We incorporate novel attention mechanisms in these architectures to investigate their potential for improving the model performance in identifying important regions on MRI for therapy outcome prediction.

1.8. Dissertation Outline

The objectives described above are expanded and presented in four chapters in this dissertation, followed by a concluding chapter. The main subject of each chapter is summarized below.

1.8.1. Chapter 2

In chapter 2, we introduce a novel system for automatic assessment of radiation therapy outcome in brain metastasis using standard serial MRI. A new multi-step framework is proposed and investigated for automatic brain tumour segmentation. This framework is applied for delineation of tumours before and at multiple imaging follow-ups after the radiotherapy to assess the therapy outcome automatically based on the RANO-BM criteria. We discuss the methods for therapy response assessment thoroughly and evaluate the performance of the system comprehensively using several relevant clinical metrics.

1.8.2. Chapter 3

In chapter 3, we explore the impact of using less-accurate automatically generated segmentation masks on the efficacy of radiomic features for radiotherapy outcome prediction in brain metastasis. Specifically, we investigate if automatic segmentation masks with different levels of accuracy compared to ground-truth tumour contours delineated manually by clinicians would result in similar radiomic signatures with comparable prediction models of radiotherapy outcome in terms of performance and risk assessment.

1.8.3. Chapter 4

In chapter 4, a novel multi-modal deep learning architecture is introduced to predict the outcome of LC/LF in brain metastasis treated with stereotactic radiation therapy using treatment-planning MRI and standard clinical attributes. At the core of the proposed architecture is an InceptionResentV2 network to extract distinct features from each MRI slice for local outcome prediction. A recurrent or transformer network is integrated into the architecture to incorporate spatial dependencies between MRI slices, as well as standard clinical attributes, into the predictive modeling. We investigate whether the deep-learning-based MRI features can complement standard clinical attributes for patient prognostication and risk assessment in terms of progression-free survival.

1.8.4. Chapter 5

In chapter 5, we propose and investigate a new explainable attention-guided deep-learning architecture to predict the radiotherapy outcome in brain metastasis *a priori*. Specifically, a novel self-attention-guided 3D residual network is introduced for predicting the outcome of LC/LF after

radiotherapy using the baseline treatment-planning MRI. The proposed model will be compared to a vanilla 3D residual network and 3D residual network with convolutional block attention module (CBAM) attention in terms of performance in outcome prediction. A new 3D visualization algorithm is developed to accompany the outcome prediction framework and illustrate how different areas within the volumetric ROI on the input images contribute to the prediction of the network for each lesion. The 3D heatmaps of importance generated by this module potentially aid clinicians to examine the lesion volume thoroughly and inspect impactful regions for a predicted outcome which can eventually support their decision in assessment, and treatment planning decisions.

1.8.5. Chapter 6

Chapter 6, summarizes and concludes the findings in chapters 2 through 5. Future directions and study limitations are also discussed in chapter 6.

Chapter 2

Automatic Assessment of Stereotactic Radiation Therapy Outcome in Brain Metastasis using Longitudinal Segmentation on Serial MRI*

2.1. Introduction

About 10% to 30% of all cancer patients develop brain metastasis [3], with a higher risk for melanoma, lung, and breast cancer patients [90]. Metastatic brain tumours represent an important cause of morbidity and mortality in cancer patients. Whereas a significant proportion of cancer patients survive for many years if the cancer is identified at an early stage while it is still localized [4], when the tumour is metastasized to the brain, the median survival ranges from as short as 5 months to up to 4 years, based on the subgroup and origin of cancer [5]–[7], [34]. Early diagnosis and precise treatment of brain metastasis may lead to the reduction of brain symptoms and may enhance the quality of life and survival of the patients [8]–[10].

Brain metastasis may occur as a single tumour (approximately 29% of cases), two-three tumours (35% of cases), and more than three tumours (36% of cases) [12]. Treatment planning for

*A version of the material presented in this chapter has been published in IEEE Journal of Biomedical and Health Informatics, 2023. DOI: 10.1109/JBHI.2023.3235304.

patients diagnosed with metastatic brain tumours depends on many factors including the origin of cancer, symptoms, number of metastases, and location of the tumour.

Magnetic resonance imaging (MRI) is the main imaging modality for diagnosis, treatment planning, and therapy outcome evaluation in brain metastasis. MRI scans are acquired before (baseline) and at multiple follow-up sessions after the radiation therapy as part of the standard treatment planning and outcome assessment procedure. The procedure requires accurate delineation of the tumour that is often performed by expert radiation oncologists and neuro-radiologists. Evaluation of radiotherapy outcome in brain metastasis on serial MRI is mainly performed based on the standard criteria presented by the response assessment in neuro-oncology–brain metastases (RANO-BM) group [28]. The RANO-BM criteria are principally based on changes in the longest diameter of the target tumour in the axial, coronal, and sagittal planes compared to baseline or nadir (smallest tumour size on the previous scans) to specify its response to therapy. The four categories of therapy response based on the RANO-BM criteria include complete response (CR; no target tumour remaining), partial response (PR; more than 30% reduction in the longest diameter compared to baseline), stable disease (SD; less than 30% decrease compared to baseline but also less than 20% increase in the longest diameter compared to nadir), or progressive disease (PD; also referred to as local failure; more than 20% increase in the longest diameter compared to nadir). Tumour enlargement on MRI after radiotherapy may also become apparent due to adverse radiation effect (ARE). Such evident tumour enlargements on MRI often become stable or followed by a decrease in tumour size on subsequent imaging follow-ups. Differentiating between tumour progression and ARE is crucial for radiotherapy response evaluation. The standard approaches to diagnose ARE include serial MRI (including the use of

T1-weighted, T2-weighted, and perfusion imaging), and where applicable, histology on resected specimens [91]–[93].

In order to calculate the tumour size changes on serial imaging, precise delineation of tumour is required for each imaging session. Manual segmentation of tumour on volumetric images acquired at several follow-up sessions for each patient is a tedious and time-consuming job. An automatic and robust tumour segmentation framework is highly desirable in the clinic and could streamline radiation therapy outcome evaluation workflow considerably. Because of many applications of automatic tumour segmentation, intense research has been carried out on this topic [94]–[96]. The existing segmentation algorithms include those that apply traditional methods such as region-based [61], [97] and model-based techniques [57], with more recent methodologies based on deep neural networks [47], [98], [99]. Deep learning-based image segmentation is now very popular in the literature and has demonstrated to outperform the traditional methods [63], [67], [100]. The deep networks for image segmentation generally consist of stacked convolutional layers and occasionally fully connected layers. Among many networks introduced for the task of segmentation, 2D and 3D UNet gained widespread popularity because of their robustness in different modalities [63], [101]. However, 2D UNet has the drawback of extracting similar features multiple times throughout the network in addition to inefficient modeling of long-range spatial dependencies. The main limitation associated with 3D UNet is that it often cannot handle large input sizes due to memory limitations with the complex architecture of the network.

Deep learning-based techniques have demonstrated promising performance in brain tumour segmentation [102]–[104]. Despite previous research on the application of these techniques in feature extraction frameworks for classifying brain tumour subtypes and predicting

clinical outcomes such as survival, their clinical efficacy in longitudinal monitoring of changes in tumour physical dimensions has not been investigated thoroughly. Cabezas et al. proposed an ensemble of 3D UNets to segment different sub-regions of gliomas on the BraTS dataset [48] to extract quantitative features for predicting the overall survival of patients. Gates et al. [105] proposed a multi-scale convolutional neural network based on the DeepMedic to segment glioma sub-volumes on MRI, and applied the features extracted from the segmented images and clinical data for predicting the overall survival. Pei et al. [106] proposed a context-aware deep learning model for brain tumour segmentation on MRI, followed by deep learning models for subtype classification and survival prediction using the tumour segments. Zhu et al. [107] developed a semi-automatic segmentation software for quantitative clinical evaluation of glioblastoma multiforme on MRI. While their results demonstrate a good correlation between the manual and semi-automatic segmentation, the developed method was not evaluated on serial MRI to quantify changes in tumour size.

In this chapter, a novel deep-learning-based system is introduced for automatic radiotherapy outcome assessment in brain metastases. A multi-step framework is proposed for automatic brain tumour segmentation that is applied for delineation of tumours before and at multiple imaging follow-ups after the radiotherapy to assess the therapy outcome automatically based on the RANO-BM criteria. To the best of our knowledge, this is the first time that a deep-learning-based segmentation framework is adapted and investigated comprehensively for automatic radiotherapy outcome assessment in brain malignancies.

2.2. Methods

2.2.1. Data Acquisition and pre-processing

This study was conducted in accordance with institutional research ethics approval from Sunnybrook Health Sciences Centre (SHSC), Toronto, Canada. The imaging and clinical data were collected from 116 patients (152 tumours; average size at baseline: 2.4 ± 1.0 cm, range: 0.5-7 cm) diagnosed with brain metastasis and treated with hypo-fractionated SRT. The patients (40.2% male, 59.8% female) were aged between 29 and 91 years (average age: 62 ± 15 years).

Among the 116 patients, 86 patients had one, 24 patients had two, and 6 patients had three or more brain metastasis tumours. The primary tumour histology included lung cancer (76 tumours, 50%), breast cancer (36 tumours, 23.7%), melanoma (15 tumours, 9.9%), colorectal cancer (7 tumours, 4.6%), renal cell carcinoma (6 tumours, 3.9%), and other cancers (12 tumours, 7.9%). Lesions with prior resection were excluded. Any salvage therapy was administrated after identifying tumour progression clinically that was the endpoint of this study. The imaging data included gadolinium-contrast-enhanced T1-weighted and T2-weighted-fluid-attenuation-inversion-recovery (T2-FLAIR) images acquired, as part of standard of care, before (baseline) and at up to 9 follow-ups after the treatment (average number of imaging follow-ups: 4). All available follow-up imaging data were used for post-treatment monitoring in this study. The dataset also included treatment-planning gross tumour volume (GTV) contours for each patient. All GTVs were contoured by an expert CNS radiation oncologist and reviewed by at least one other CNS radiation oncologist and a neuroradiologist. The GTVs were used to generate ground-truth tumour masks for the baseline and follow-up scans under the supervision of expert oncologists. the MRI scans were acquired using a 1.5 T Ingenia system (Philips Healthcare, Best, Netherlands) and a

1.5 T Signa HDxt system (GE Healthcare, Milwaukee, WI, USA). The scan sequences were 3D T1w TFE (repetition time: 9.4 ms, echo time: 2.3 ms, imaging frequency: 127.77 MHz) and 3D T1w FSPGR (repetition time: 8.548 ms, echo time: 4.2 ms, imaging frequency: 63.86 MHz) for the T1-weighted images and T2 FLAIR CLEAR (repetition time: 9000 ms, echo time: 125 ms, imaging frequency: 127.77 MHz) and T2 FLAIR PROPELLER (repetition time: 8600 ms, echo time: 117 ms, imaging frequency: 63.86 MHz) for the T2-FLAIR images. The in-plane image resolution and the slice thickness were 0.5 and 1.5 mm for T1-weighted and 0.5 and 5 mm for T2-FLAIR images, respectively. All images were resampled with a voxel size of $0.5 \times 0.5 \times 1$ mm^3. The voxel intensities in each image were normalized to be between 0 and 1. The normalization was done on voxel level (*vox_intst*) using Equation (2.1).

$$normalized\ voxel_{intensity} = \frac{\left(voxel_{intensity} - min_{intensity}\right)}{max_{intensity} - min_{intensity}} \tag{2.1}$$

In this equation, the $min_{intensity}$ and $max_{intensity}$ are the minimum and maximum intensity values in the corresponding 3D image. The T2-FLAIR images were co-registered on their corresponding T1-weighted images using an affine registration. Among the 116 patients, 96 patients (130 tumours) were randomly selected for training the models, and the remaining 20 patients (22 tumours) were kept as an unseen test set for independent evaluation.

The tumours were monitored longitudinally on MRI after SRT and the pattern of changes in tumour size as well as the ground-truth local control/failure (LC/LF) outcome for each tumour was determined by a radiation oncologist using the follow-up imaging data. The follow-up scans were performed every 2-3 months for all patients until they transitioned to palliative care or passed

away. The ground-truth tumour size status (decrease/stable/increase) was determined for each follow-up scan. Specifically, the tumour size status was determined as decrease/increase if a measurable (≥2 mm) decrease/increase was evident in the longest diameter of the tumour in the axial plane compared to the previous scan, otherwise, it was determined as stable. The RANO-BM criteria were used to determine an outcome of LC (complete response, partial response, or stable disease) or LF (progressive disease) for each tumour separately [28]. Adverse radiation effect (ARE) was diagnosed and differentiated from local progression based on the report by Sneed et al. [91]. The ARE cases were diagnosed clinicoradiologically based on serial imaging, including the use of perfusion MRI (rCBV cut-off = 2) and chemical exchange saturation transfer (CEST) imaging, and/or through histological confirmation (available for 50% of tumours diagnosed with ARE) [92], [108].

2.2.2. Tumour Segmentation Framework

Figure 2.1 presents a scheme of the proposed framework for automatic segmentation of brain tumours on MRI. The framework consists of two cascaded 2D UNets to find the approximate position of the tumour. Once the approximate tumour position is found, the image is cropped around the tumour to make the size of input image smaller for the next network. Specifically, the size of input T1-weighted images for the first and second 2D UNets is 512 × 512 and 256 × 256 pixels, respectively. The need for cropping images stems from the fact that both the 3D UNet and multi-scale self-guided attention (MSGA) network [109] adapted in the framework have memory limitation which makes their training process challenging. If the input size for the 3D UNet is the original image size (512 × 512 × 128 voxels) without cropping, one needs to patch or resize the input volume to meet the memory limitations of the network. Patching the volume leads to losing

contextual information (*e.g.*, tumour tears apart in different patches) while resizing it results in losing detailed local information. Similarly, and due to its complex architecture, training the MSGA network on the original 2D images (512 × 512 pixels) with two channels of T1-weighted and T2-FLAIR requires limiting the batch size. With cropping, it would be possible to preserve both local and contextual information using the approximate position of the tumour estimated with the cascaded 2D UNets. The output of each 2D UNet for a patient is a set of 128 2D masks with size of 512 × 512 pixels for the first and 256 × 256 pixels for the second 2D UNet. To find the approximate position of the tumour from these masks, a logical OR operation is applied on all the 2D masks to create a single mask presenting an upper-bound of the tumour areas in different slices. Subsequently, the connected components are identified in the single mask and the center of each connected component is regarded as the approximate center of the corresponding tumour. The approximated centers are used to crop the image around the tumour region. In cases where there is more than one tumour in an MRI volume, the tumours are treated separately, and the final masks are fused at the end. At the core of the framework there are two segmentation networks including a 3D UNet and a MSGA network. The 3D UNet is fed with the cropped T1-weighted volumetric images (128 × 128 × 128 voxels). The MSGA network is fed with cropped two-channel T1-weighted and T2-FLAIR co-registered image slices (128 × 128 pixels each). The output of these two networks is fused at the end using slice-wise averaging over their output probability maps. The final output masks are generated by thresholding the averaged probability maps with a threshold level of 0.5.

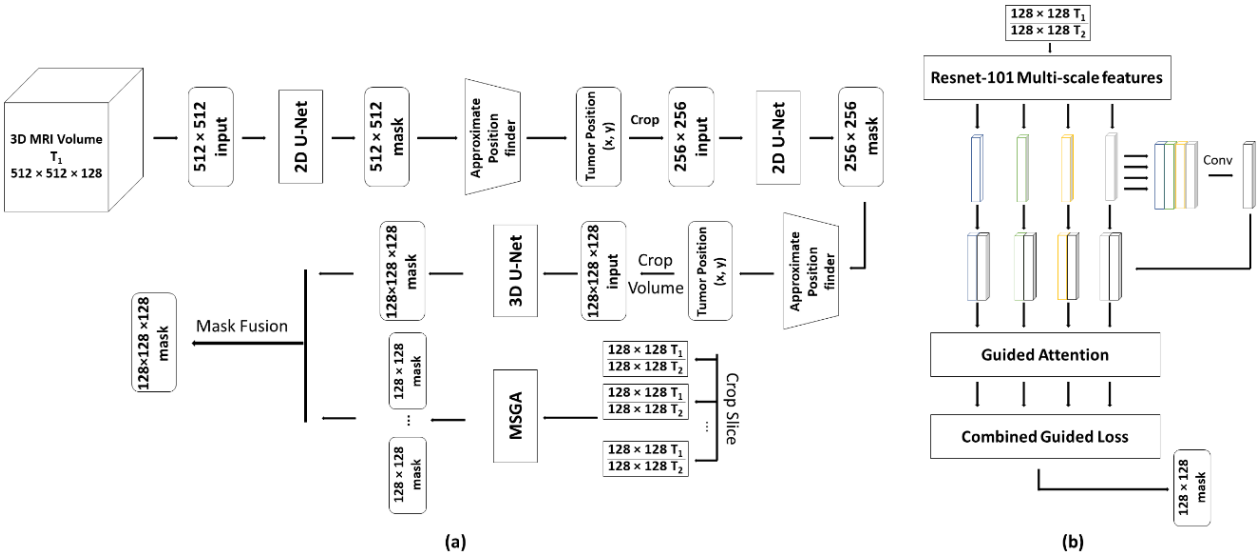

Figure 2.1- (a) Overview of the proposed segmentation framework. For a volumetric input image (contrast-enhanced T1-weighted, 512×512×128 voxels), first all slices are fed to a 2D UNet one by one. The generated masks from the 2D UNet are used to finds an approximate tumour position *(x, y)*. The volumetric input image is then cropped around *(x, y)* into a 256×256×128 voxel volume. A similar procedure is performed to reduce the size of volumetric image containing the tumour to 128×128×128 voxels. This volume is then fed into a 3D UNet for segmentation with no patching. The slices of this volume are also fed to MSGA, after concatenation with the co-registered T2-FLAIR image, and the segmentation masks of MSGA are then fused with those of 3D UNet. (b) The MSGA network structure. Features extracted at different scales from Resnet-101 are concatenated and convolved and then self-concatenated and fed into guided attention module. The resulting self-guided features are fed into the guided loss.

The choice of a combination of 2D UNet, 3D UNet, and MSGA network is to take advantage of their features, while simultaneously mitigating their limitations. More specifically, whereas the 2D UNets can effectively localize the region of interest even for smaller tumours to crop the large input image, it can not generate precise segmentation masks for all tumours. On the other hand, a localized input for the 3D UNet and MSGA network reduces irrelevant information and enhances the model focus on the region of interest, leading to considerable improvements in their performance in generating precise segmentation masks. The good performance of the 2D UNet architecture in various segmentation tasks is due to its capability to capture context and enable localization, using a contracting path and a symmetric expanding path, with skip connections in-between the two paths [101]. Such architecture enables the network to share features from multiple layers and overcome the trade-off between localization accuracy and

context utilization. The drawback of 2D UNet, however, is that it does not consider the 3D spatial dependencies between the voxels, and consequently, loses a considerable amount of useful information for segmentation. To overcome this, Çiçek et al. proposed the 3D UNet as a volumetric image segmentation network [63], which maintains the benefits of the 2D UNet architecture but also considers the voxel dependencies. Considering 3D spatial dependencies comes at the cost of high memory consumption because of the huge input size. A cascaded network of 2D UNet and 3D UNet could benefit from the advantages of 3D UNet while the redundant information could be filtered out using 2D UNet to meet the memory limitations of the 3D UNet. The two main drawbacks of the encoder-decoder architectures such as 2D and 3D UNet include deriving redundant information, and more importantly, inefficient modeling of long-range feature dependencies in these networks. Sinha *et al.* [109] proposed a multi-scale self-guided attention network to overcome these limitations. The MSGA network enables capturing richer contextual dependencies and neglecting irrelevant information by using an attention mechanism. Also, the utilization of interdependent channel maps which enable the network to integrate local features with their corresponding global dependencies makes it efficient in our application, where the network is fed with two channels of T1-weighted and T2-FLAIR images. Further details on the MSGA architecture have been provided in Appendix A.

2.2.3. Training and Evaluation of the System

In order to train and evaluate the tumour segmentation framework, the data associated with samples of the training and test sets were completely separated at patient level. The networks in the framework were trained independently using the data acquired from the training samples. The second 2D UNet, the 3D UNet, and the MSGA network were trained using the manually cropped

data from the training set. The networks were only trained on the images acquired at the baseline. This was done to permit evaluating the framework's performance on the training set at the first follow-up and compare it with the performance on the independent test set. The framework was initially evaluated in terms of segmentation accuracy, using the images of the independent test set acquired at the baseline and follow-up scans. The Dice similarity coefficient, Hausdorff distance, and the tumour volume estimation error were used for this evaluation. The performance of the system was subsequently evaluated in monitoring the tumour size status after SRT and automatic assessment of therapy outcome using the imaging data of the independent test set acquired at the baseline and all follow-ups available for each patient. For comparison, experiments were conducted using seven different models following a similar training and evaluation procedure. The first model included two cascaded 2D UNets, the second model consisted of a 3D UNet, and the third model included a 3D UNet along with an MSGA network. For training and testing the standalone 3D UNet and 3D UNet + MSGA in the second and third models, each 512 × 512 × 128 voxel volume was patched into 16 input patches of 128 × 128 × 128 voxels and the associated masks were concatenated together at the end. The fourth model included two cascaded 2D UNets followed by a 3D UNet, the fifth model utilized the framework proposed in this study inputting the T1-weighted image only, and the sixth model incorporated the complete framework proposed (Figure 2.1). The seventh model utilized the well-recognized nnUNet framework [110] for further comparison (details on the nnUNet framework have been provided in Appendix A). The nnUNet framework input the co-registered T1-weighted and T2-Flair images (512 × 512 × 128 voxels) as two channels, where each image was down-sampled to 128 × 128 × 128 voxels for the first 3D UNet in the framework. Pre-training of the networks for weight initialization was performed using

the data from the brain tumour segmentation (BraTS) dataset [48]. A set of 9 tumours from the training samples was used as the validation set for tuning the network hyperparameters in the training phase. A batch size of 4 and 2 was used for training the 2D UNets and nnUNet, respectively. The batch size for the 3D UNet and MSGA network was tuned to one based on the experimental results on the validation set. The training was performed with a learning rate of 0.0001 for all networks. A dice and a cross-entropy based loss function was used for the 2D and 3D UNets, respectively. The loss function for the nnUNet was defined as the sum of the dice and cross-entropy losses. The dice loss function was defined as $(1 - dice\ coefficient)$, with the $dice\ coefficient$ obtained using Equation (2.2).

$$dice\ coefficient = \frac{2 \times TP}{2TP + FN + FP + smooth} \tag{2.2}$$

In this equation, TP, FP, and FN refer to true positive, false positive, and false negative in segmentation on pixel/voxel level, respectively. A smoothing term was added in the equation to prevent division by zero. Instead of setting Boolean intensity values for the ground-truth and the automatically generated masks and performing Boolean operations, the mask intensities were defined as continuous values to make the dice loss differentiable. The cross-entropy loss was defined using Equation (2.3).

$$cross - entropy\ loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{J=1}^{M}(y_{ij}.log(p_{ij})) \tag{2.3}$$

In this equation, *N* and *M* are the number of pixels and classes (in our case two, tumour vs. normal tissue), respectively. The loss function for the MSGA network was defined using Equation (2.4).

$$MSGA\ loss = Lseg_{total} + LG_{total} + Lrec_{total} \quad (2.4)$$

In this equation, $Lrec_{total}$ is the mean squared error between the original input and output features of the encoder-decoder network in the attention module. LG_{total} is the mean squared error between the encoded representation of features in the encoder-decoder network inside the attention module. Finally, $Lseg_{total}$ is the cross-entropy between the ground-truth and network output masks. The training and validation loss for the 3D UNet and MSGA networks over the training epochs are presented in Figure 2.2. The framework was developed on an Nvidia GeForce RTX 2080 Ti with 12 GB of Memory. All models were developed in Python and trained and tested using Keras with TensorFlow backend.

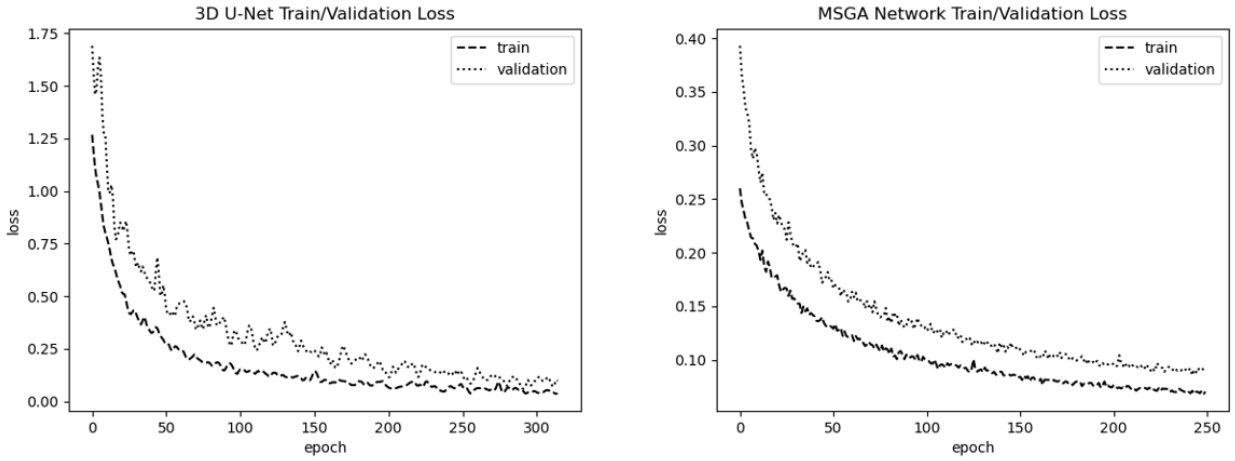

Figure 2.2- Training and validation loss of the 3D UNet and MSGA network in the proposed segmentation framework.

2.2.4. Procedure and Criteria for Automatic Assessment of Tumour Size Status, Local Response, and ARE Outcome

The segmentation masks generated by the deep learning models were used to estimate the size of tumour in each scan and, subsequently, the tumour size changes after SRT. The tumour size status,

local response, and the ARE outcome were then assessed automatically based on the estimated changes in tumour size using the procedure and criteria described below.

A typical SRT outcome evaluation workflow in the clinic consists of determining the tumour size status at each follow-up scan compared to the previous scan. For automatic assessment of tumour size status, following the protocol applied in the clinic, the longest diameter of tumour in the axial plane was calculated for all scans using the automatic segmentation masks.

Tumour size status at each follow-up scan was labeled as increase or decrease if a measurable increase or decrease ($\geq$ 2 mm) was estimated, respectively, in the tumour's longest diameter compared to the previous scan. Otherwise, it was labeled as stable. The tumour size status labels identified automatically were compared with the ground-truth labels to evaluate the performance of automatic labeling in terms of accuracy, precision, and recall. It should be noted that this step was only to evaluate the performance of the network in automatic labeling of tumour size status and not the local response (discussed below).

The SRT outcome in terms of LC/LF and ARE was evaluated for each tumour automatically based on the RANO-BM criteria. Using the automatic segmentation masks, the longest diameter of tumour in the axial, coronal, and sagittal planes was estimated for the baseline and all follow-ups. The relative change in the longest diameter of tumour was calculated at each follow-up compared to the baseline and nadir. The change in the tumour diameter at each follow-up was categorized into three categories of shrinkage, steady, and enlargement when more than 30% decrease compared to baseline, less than 30% decrease compared to baseline but also less than 20% increase compared to nadir, and more than 20% increase compared to nadir was detected

in the tumour longest diameter, respectively [28]. Further, the relative change in tumour volume was calculated at each follow-up compared to the baseline and nadir. The change in the tumour volume at each follow-up scan was categorized into three categories of shrinkage, steady, and enlargement based on the volumetric response assessment criteria proposed by Oft et al. [111] which is an extension of RANO-BM guideline recommendations for volumetric response assessment. Specifically, shrinkage at a follow-up scan was defined as more than 65% reduction in tumour volume compared to baseline, steady as less than 65% reduction compared to baseline but also less than 72.8% increase compared to nadir, and enlargement as more than 72.8% increase in tumour volume compared to nadir. The categories detected at each follow-up scan using the automatic segmentation models were compared to those identified from the ground-truth segmentation masks to evaluate the performance of automatic response categorization at individual follow-ups in terms of accuracy, precision, and recall.

The shrinkage/steady/enlargement patterns determined based on the longest diameter for each tumour at the follow-up scans were used for automatic detection of LC/LF and ARE outcome. Any tumour demonstrating a sequence of steady or shrinkage patterns at follow-ups with no enlargement was classified with an LC outcome. When an enlargement was detected in the pattern of size changes, the change in the tumour longest diameter at the next follow-up was calculated compared to the scan in which the enlargement was detected. The tumour was classified with an LF outcome if its size increased again (more than 2 mm to account for measurement errors) compared to the previous scan. If the tumour size decreased or remained stable after the initial enlargement, the tumour was classified as LC but with ARE. As a tumour with ARE could possibly progress later and be classified as LF, detection of LC/LF and ARE outcome was performed and

evaluated independently for each tumour. The outcomes identified automatically were compared with the ground-truth outcome for each tumour to evaluate the performance of the automatic outcome assessment in terms of accuracy, sensitivity, and specificity.

2.3. Results

Figure 2.3 demonstrates contrast-enhanced T1-weighted images acquired from three representative brain metastasis patients with an outcome of LC, LF, and ARE after SRT, respectively. In Figure 2.3(a), the tumour has consistently shrunk after SRT (follow-ups 1-3), demonstrating an LC outcome. In Figure 2.3(b) the tumour has continued to grow after the first follow-up, showing an LF outcome. In Figure 2.3(c), initial growth in the first follow-up stopped immediately in the second follow-up, followed by further shrinkage in the third follow-up, which is evidence for ARE.

Figure 2.4 shows the ground-truth and automatic tumour segmentation masks generated by different deep learning models for five representative patients of the test set. The images show a step-by-step improvement in the automatic segmentation masks generated by the cascaded 2D UNets, 3D UNet, cascaded 2D & 3D UNets, and the complete segmentation framework proposed in this paper (cascaded 2D & 3D UNets + MSGA). Specifically, the proposed frameworks could achieve a close to perfect segmentation for cases (a) and (c), while for cases (b) and (d) it slightly under-segmented the tumour, and for case (e) the results indicate over-segmentation. In general, the results demonstrate that the model is not biased towards under- or over-segmentation.

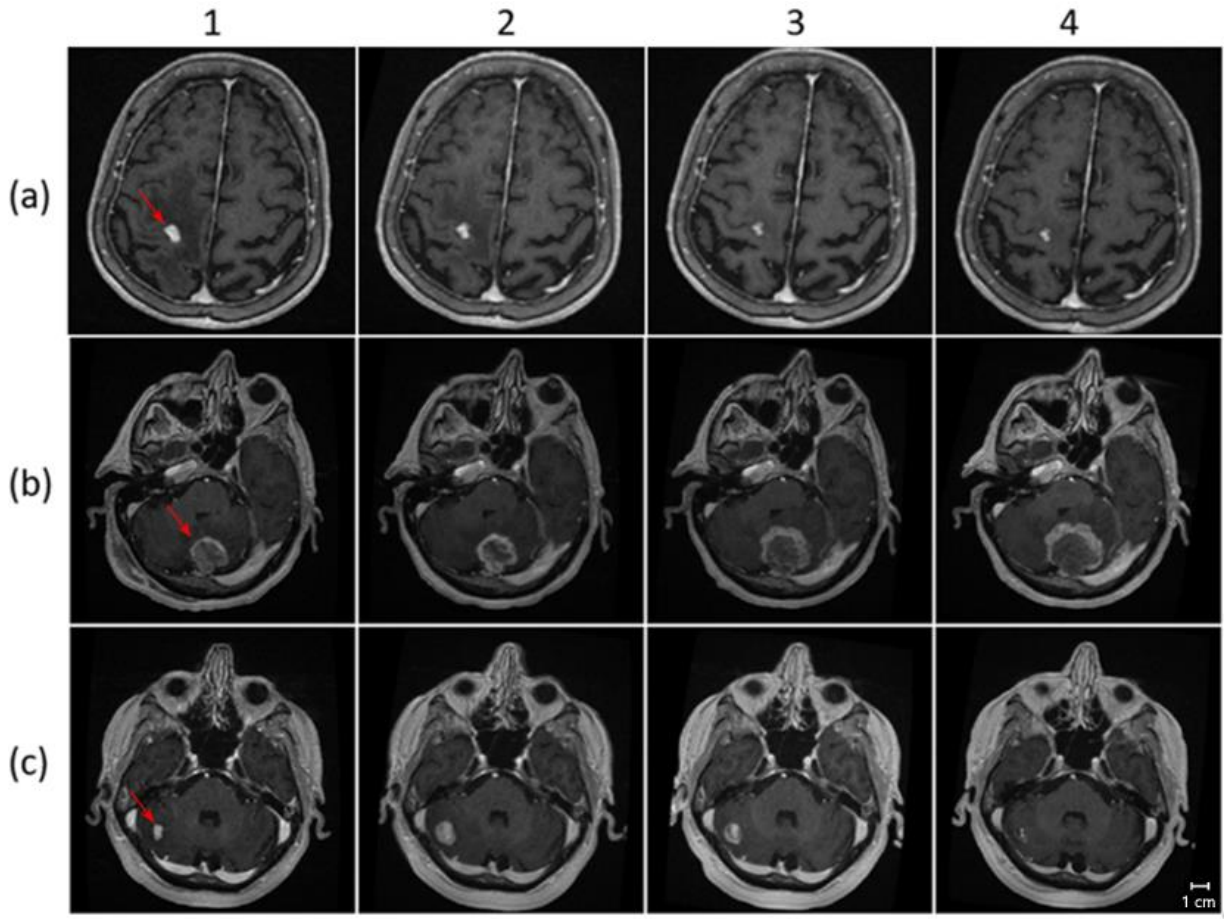

Figure 2.3- Contrast-enhanced T1-weighted images acquired at the baseline (1), and the first (2), second (3), and third (4) follow-ups after SRT from three representative patients with brain metastasis demonstrating local control (a), local failure (b), and ARE (c) after treatment. The arrow in the baseline image shows the location of brain metastasis. LC/LF/ARE is evaluated based on the changes in longest diameter. In (c) an initial growth in first follow-up is followed by a decrease in the second, and then third follow-ups.

A detailed comparison between the segmentation results of different networks at the baseline and follow-up sessions is given in Table 2.1 in terms of dice similarity coefficient, Hausdorff distance, and tumour volume estimation error. A consistent step-by-step improvement is observed in different criteria of segmentation accuracy, with the best results associated with the cascaded 2D & 3D UNets + MSGA architecture inputting both the T1-weighted and T2-FLAIR images. The networks demonstrate a similar performance of the training and test sets, implying very good generalizability for tumour segmentation of new unseen cases.

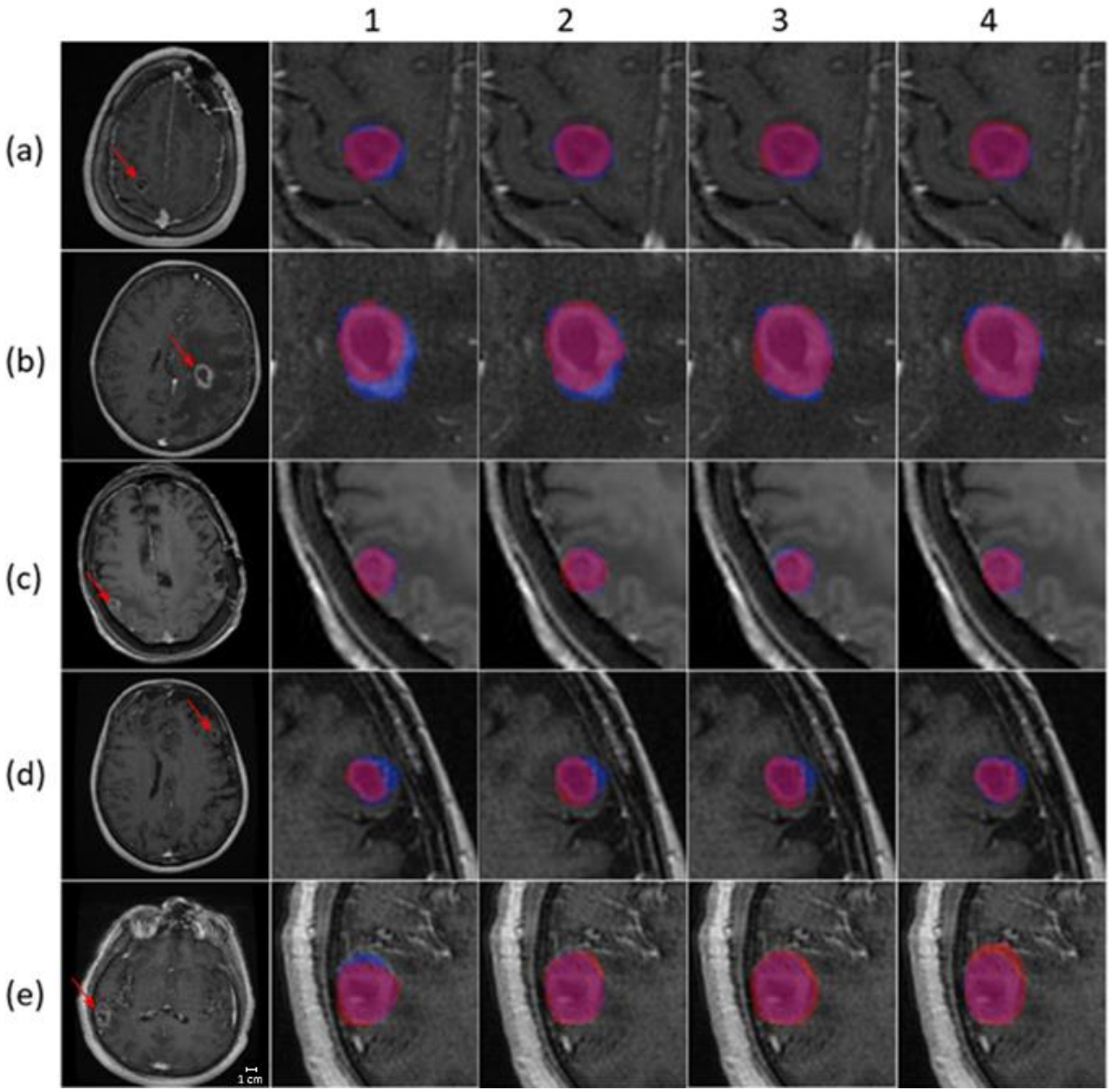

Figure 2.4- Tumour segmentation masks generated by the cascaded 2D UNets (1), 3D UNet (2), cascaded 2D & 3D UNets (3), and Cascaded 2D & 3D UNets + MSGA (4) for five representative patients (a-e) in the test set. The images are acquired at the baseline. The arrow in the first image of each row shows the location of brain metastasis. The ground-truth and automatic segmentation masks (middle slide) are shown in blue and red, respectively, with the purple area showing the overlap region in each case.

Further, the segmentation results of the proposed framework are comparable between the baseline and follow-up scans. The relative deterioration in the segmentation performance metrics at the follow-up scans can be attributed to the gradual tumour shrinkage observed in many cases after the radiotherapy. It should be noted that in the experiments conducted in this study, no data from

the follow-ups were used for training the networks. Specifically, the networks were solely trained using the data of the training set patients acquired at the baseline, but subsequently evaluated using the follow-up data from the patients of the training and test sets, separately. Table 2.2 reports the 95% confidence intervals associated with the results reported in Table 2.1. The confidence intervals were estimated based on the sample mean and standard error of the mean in each experiment using the t distribution. Experimental results with different hyperparameters have been presented in Table 2.3.

Table 2.1- Dice similarity coefficient (DSC), Hausdorff distance (HD), and volume estimation error (VEE) for segmentation of brain metastasis at the baseline and follow-up scans using different network architectures.

Segmentation Model	**Metric**	**Baseline**		**1st Follow-up**		**2nd Follow-up**	**3rd Follow-up**	**4th Follow-up**	**5th Follow-up**
		Training Set	**Test Set**	**Training Set**	**Test Set**	**Test Set**	**Test Set**	**Test Set**	**Test Set**
Cascaded 2D UNets	DSC	86.5 ± 5.8	85.4 ± 7	82.8 ± 6	81.3 ± 5.9	79.7 ± 11	77.2± 9.7	76.1 ± 10.6	74.3 ± 11.2
	HD (mm)	2.8 ± 0.4	3 ± 0.6	3.2 ± 0.6	3.7 ± 0.5	4.6 ± 0.7	4.4 ± 0.6	4.3 ± 0.7	4.5 ± 0.9
	VEE (cc)	0.55 ± 0.5	0.58 ± 0.5	0.64 ± 0.5	0.67 ± 0.5	0.82 ± 0.7	0.78 ± 0.7	0.75 ± 0.6	0.75 ± 0.7
	VEE (%)	15.8 + 7	16.4 ± 9	18.3 ± 8.4	19.2 ± 8.3	23.7 ± 11	24.5 ± 12.5	26.1 ± 11.9	30.2 ± 14
3D UNet	DSC	88.8 ± 4.5	87.2 ± 5.4	84.8 ± 5.5	83.7 ± 6.4	83 ± 8	81.6 ± 7.7	80 ± 8	79.8 ± 8
	HD (mm)	2.4 ± 0.7	2.6 ± 0.7	2.6 ± 0.6	3.3 ± 0.6	4.2 ± 0.7	4.3 ± 0.7	4.5 ± 0.6	4.2 ± 0.6
	VEE (cc)	0.50 ± 0.4	0.52 ± 0.5	0.56 ± 0.5	0.60 ± 0.5	0.79 ± 0.7	0.74 ± 0.6	0.73 ± 0.7	0.73 ± 0.6
	VEE (%)	14.9 ± 5.2	15.3 ± 6.8	17.5 ± 6.3	17.8 ± 9.1	21.3 ± 9.4	23.3 ± 12	24.1 ± 9.4	24.7 ± 9.4
3D UNet + MSGA	DSC	89.7 ± 5	88.9 ± 5.3	85.6 ± 5.1	84.8 ± 6.2	84.1 ± 7.7	83.5 ± 8.3	82.4 ± 8.8	82.1 ± 9.6
	HD (mm)	2.3 ± 0.5	2.5 ± 0.7	2.6 ± 0.5	3.2 ± 0.8	4 ± 0.7	3.5 ± 0.7	3.6 ± 0.7	3.6 ± 0.7
	VEE (cc)	0.51 ± 0.4	0.52 ± 0.4	0.55 ± 0.4	0.59 ± 0.5	0.74 ± 0.7	0.74 ± 0.7	0.72 ± 0.7	0.71 ± 0.7
	VEE (%)	12.3 ± 4.3	13.6 ± 6	16.9 ± 8.5	18 + 8.8	19.3 ± 7.8	21.5 ± 6.8	22.5 ± 9	23.3 ± 12
Cascaded 2D & 3D UNets	DSC	90.1 ± 4.4	89.6 ± 4.6	86.2 ± 4.6	85.1 ± 5	84.3 ± 7.2	83.4 ± 7	82.9 ± 7	82.8 ± 7
	HD (mm)	2.3 ± 0.2	2.4 ± 0.4	2.6 ± 0.5	3.1 ± 0.8	3.8 ± 0.7	3.2 ± 0.6	3.5 ± 0.6	3.4 ± 0.6
	VEE (cc)	0.5 ± 0.3	0.51 ± 0.4	0.55 ± 0.4	0.57 ± 0.5	0.73 ± 0.6	0.71 ± 0.6	0.72 ± 0.6	0.71 ± 0.7
	VEE (%)	11.1 ± 4.2	12.5 ± 5.3	16.7 ± 8.3	18.2 ± 8	18.7 ± 7.5	20 ± 7.5	21.7 ± 9	22.6 ± 10
Cascaded 2D & 3D UNets + MSGA (with T1-weighted only)	DSC	91.1 ± 3.8	90.3 ± 4.1	87.1 ± 3.9	86.4 ± 4.9	85.3 ± 7.2	83.7 ± 8.4	83.2 ± 8.5	82.8 ± 9.2
	HD (mm)	2 ± 0.4	2.3 ± 0.5	2.43 ± 0.5	2.95 ± 0.7	3.7 ± 0.7	3.4 ± 0.7	3.5 ± 0.7	3.4 ± 0.6
	VEE (cc)	0.42 ± 0.3	0.49 ± 0.4	0.55 ± 0.4	0.58 ± 0.5	0.72 ± 0.6	0.68 ± 0.6	0.68 ± 0.6	0.69 ± 0.6
	VEE (%)	10.5 ± 4.8	11 ± 5	14.6 ± 5.4	16.4 ± 5.4	17.9 ± 6.8	18.7 ± 7	19.3 ± 8.2	23.8 ± 10.1
Cascaded 2D & 3D UNets + MSGA	DSC	**92.3 ± 3.1**	**91.5 ± 3.7**	**88.7 ± 3.7**	**87.4 ± 5.2**	**86.7 ± 5.5**	85.1 ± 6.1	**84.1 ± 6.8**	**84.5 ± 7**
	HD (mm)	**1.84 ± 0.4**	**2.1 ± 0.6**	2.21 ± 0.5	**2.84 ± 0.7**	**2.98 ± 0.7**	3 ± 0.6	**2.89 ± 0.6**	**2.9 ± 0.58**
	VEE (cc)	**0.39 ± 0.3**	**0.44 ± 0.4**	**0.52 ± 0.4**	**0.57 ± 0.5**	**0.59 ± 0.5**	**0.61 ± 0.5**	**0.62 ± 0.6**	**0.6 ± 0.56**
	VEE (%)	**9.2 ± 4.6**	**10.2 ± 5.3**	**12.5 ± 4**	**13.4 ± 5.1**	**15.7 ± 6.3**	16.5% ± 7	**17.3 ± 8.3**	**19.7 ± 8.5**
nnUNet	DSC	91.7 ± 3.3	90.9 ± 3.1	88.5 ± 3.2	87.1 ± 5.9	86.6 ± 6	**85.3 ± 6.5**	84 ± 7	84.2 ± 8.1
	HD (mm)	1.88 ± 0.4	2.3 ± 0.57	**2.18 ± 0.5**	2.95 ± 0.8	3 ± 0.8	**2.9 ± 0.7**	2.93 ± 0.61	2.9 ± 0.66
	VEE (cc)	0.43 ± 0.4	0.5 ± 0.3	**0.52 ± 0.4**	0.63 ± 0.5	0.67 ± 0.5	0.61 ± 0.6	0.63 ± 0.5	0.61 ± 0.6
	VEE (%)	9.8 ± 4.2	11 ± 5.5	12.5 ± 4.1	15.1 ± 5.7	16.3% ± 6.8	**16.4 ± 7**	17.5% ± 8.2	20.2 ± 8.8

Table 2.2- The 95% confidence intervals associated with the results in Table 2.1 The confidence intervals were estimated for the dice similarity coefficient (DSC), Hausdorff distance (HD), and the volume estimation error (VEE) based on the sample mean and standard error of the mean in each experiment using the t test.

Segmentation Model	Metric	Baseline		1st Follow-up		2nd Follow-up	3rd Follow-up	4th Follow-up	5th Follow-up
		Training Set	Test Set	Training Set	Test Set	Test Set	Test Set	Test Set	Test Set
Cascaded 2D UNets	DSC	85.5 - 87.5	82.3 - 88.5	81.8 - 83.8	78.7 - 83.9	73.8 - 85.6	70.7 - 83.7	68.5 - 83.7	60.4 - 88.2
	HD (mm)	2.7 - 2.9	2.7 - 3.3	3.1 - 3.3	3.5 - 3.9	4.2 - 5.0	4.0 - 4.8	3.8 - 4.8	3.4 - 5.6
	VEE (cc)	0.6 - 0.6	0.4 - 0.8	0.5 - 0.7	0.4 - 0.8	0.4 - 1.2	0.3 - 1.3	0.4 - 1.2	0.0 - 1.6
	VEE (%)	14.6 - 17.0	12.4 - 20.4	16.8 - 19.8	15.5 - 22.9	17.8 - 29.6	16.1 - 32.9	17.6 - 34.6	12.8 - 47.6
3D UNet	DSC	88.0 - 89.6	84.8 - 89.6	83.8 - 85.8	80.9 - 86.5	78.7 - 87.3	76.4 - 86.8	74.3 - 85.7	69.9 - 89.7
	HD (mm)	2.3 - 2.5	2.3 - 2.9	2.5 - 2.7	3.0 - 3.6	3.8 - 4.6	3.8 - 4.8	4.1 - 4.9	3.5 - 4.9
	VEE (cc)	0.4 - 0.6	0.3 - 0.7	0.6 - 0.6	0.4 - 0.8	0.4 - 1.2	0.3 - 1.1	0.2 - 1.2	0.0 - 1.6
	VEE (%)	14.0 - 15.8	12.3 - 18.3	16.4 - 18.6	13.8 - 21.8	16.3 - 26.3	15.2 - 31.4	17.4 - 30.8	13.0 - 36.4
3D UNet + MSGA	DSC	88.8 - 90.6	86.6 - 91.2	84.7 - 86.5	82.1 - 87.5	80.0 - 88.2	77.9 - 89.1	76.1 - 88.7	70.2 - 94.0
	HD (mm)	2.2 - 2.4	2.2 - 2.8	2.5 - 2.7	2.8 - 3.6	3.6 - 4.4	3.0 - 4.0	3.1 - 4.1	2.7 - 4.5
	VEE (cc)	0.4 - 0.6	0.3 - 0.7	0.6 - 0.6	0.4 - 0.8	0.4 - 1.2	0.4 - 1.2	0.2 - 1.2	0.2 - 1.6
	VEE (%)	11.6 - 13.0	10.9 - 16.3	15.4 - 18.4	14.1 - 21.9	15.1 - 23.5	16.9 - 26.1	16.1 - 28.9	8.4 - 38.2
Cascaded 2D & 3D UNets	DSC	89.3 - 90.9	87.6 - 91.6	85.4 - 87.0	82.9 - 87.3	80.5 - 88.1	78.7 - 88.1	77.9 - 87.9	74.1 - 91.5
	HD (mm)	2.3 - 2.3	2.2 - 2.6	2.5 - 2.7	2.7 - 3.5	3.4 - 4.2	2.8 - 3.6	3.1 - 3.9	2.7 - 4.1
	VEE (cc)	0.4 - 0.6	0.3 - 0.7	0.6 - 0.6	0.4 - 0.8	0.4 - 1.0	0.3 - 1.1	0.3 - 1.1	0.2 - 1.6
	VEE (%)	10.4 - 11.8	10.2 - 14.8	15.3 - 18.1	14.7 - 21.7	14.7 - 22.7	15.0 - 25.0	15.3 - 28.1	10.2 - 35.0
Cascaded 2D & 3D UNets + MSGA (with T1-weighted only)	DSC	90.4 - 91.8	88.5 - 92.1	86.4 - 87.8	84.2 - 88.6	81.5 - 89.1	78.1 - 89.3	77.1 - 89.3	71.4 - 94.2
	HD (mm)	1.9 - 2.1	2.1 - 2.5	2.3 - 2.5	2.7 - 3.3	3.3 - 4.1	2.9 - 3.9	3.0 - 4.0	2.7 - 4.1
	VEE (cc)	0.4 - 0.4	0.3 - 0.7	0.6 - 0.6	0.4 - 0.8	0.4 - 1.0	0.3 - 1.1	0.3 - 1.1	0.2 - 1.4
	VEE (%)	9.7 - 11.3	8.8 - 13.2	13.7 - 15.5	14.0 - 18.8	14.3 - 21.5	14.0 - 23.4	13.4 - 25.2	11.3 - 36.3
Cascaded 2D & 3D UNets + MSGA	DSC	**91.8 - 92.8**	**89.9 - 93.1**	**88.1 - 89.3**	**85.1 - 89.7**	**83.8 - 89.6**	**81.0 - 89.2**	**79.2 - 89.0**	75.8 - 93.2
	HD (mm)	**1.7 - 1.9**	**1.8 - 2.4**	**2.1 - 2.3**	**2.4 - 3.2**	**2.6 - 3.4**	2.6 - 3.4	**2.5 - 3.3**	2.2 - 3.6
	VEE (cc)	**0.3 - 0.5**	**0.2 - 0.6**	**0.6 - 0.6**	**0.4 - 0.8**	**0.3 - 0.9**	0.3 - 0.9	**0.2 - 1.0**	0.1 - 1.3
	VEE (%)	**8.4 - 10.0**	**7.9 - 12.5**	**11.8 - 13.2**	**11.1 - 15.7**	**12.3 - 19.1**	11.8 - 21.2	**11.4 - 23.2**	9.1 - 30.3
nnUNet	DSC	91.1 - 92.3	89.5 - 92.3	87.9 - 89.1	84.5 - 89.7	83.4 - 89.8	80.9 - 89.7	79.0 - 89.0	74.1 - 94.3
	HD (mm)	**1.7 - 1.9**	2.0 - 2.6	**2.1 - 2.3**	2.7 - 3.3	**2.6 - 3.4**	**2.4 - 3.4**	2.6 - 3.4	2.1 - 3.7
	VEE (cc)	85.5 - 87.5	82.3 - 88.5	81.8 - 83.8	78.7 - 83.9	73.8 - 85.6	70.7 - 83.7	68.5 - 83.7	60.4 - 88.2
	VEE (%)	2.7 - 2.9	2.7 - 3.3	3.1 - 3.3	3.5 - 3.9	4.2 - 5.0	4.0 - 4.8	3.8 - 4.8	3.4 - 5.6

Table 2.3- Experimental results of the proposed segmentation framework with different hyperparameters. DSC: dice similarity coefficient; CI: 95% confidence interval

Network	Learning Rate	Batch Size	Baseline Validation Set (DSC)	Baseline Test Set (DSC)	Baseline Test Set (DSC CI)
2D UNets, 3D UNet and MSGA	1e-4	4, 1, 1	**91 ± 3.8**	**91.5 ± 3.7**	**89.9 – 93.1**
3D UNet (2D UNets and MSGA fixed)	5e-5	2	90.2 ± 5.2	90.8 ± 4.2	88.9 – 92.7
MSGA (2D UNets and 3D UNet fixed)	5e-5	2	90.5 ± 4.4	91.4 ± 3.6	89.8 – 93
3D UNet (2D UNets and MSGA fixed)	5e-3	2	90 ± 5.5	89.7 ± 3.5	88.1 – 91.3
MSGA (2D UNets and 3D UNet fixed)	5e-3	2	90.1 ± 5.7	89.5 ± 4	87.7 – 91.3
3DUNet (2D UNets and MSGA fixed)	5e-5	4	89.9 ± 4.9	90.8 ± 4.1	89 – 92.6
MSGA (2D UNets and 3D UNet fixed)	5e-5	4	90.3 ± 4.1	91.3 ± 3.4	89.8 – 92.8

The segmentation results for separate categories of baseline tumour size are presented in Table 2.4. The results demonstrate that the proposed framework outperformed the other segmentation models in most cases, with few cases of performance in par with the nnUNet model. Table 2.5 reports the 95% confidence intervals associated with the results reported in Table 2.4. The confidence intervals were estimated based on the sample mean and standard error of the mean in each experiment using the t distribution.

Table 2.4- Dice similarity coefficient (DSC), Hausdorff Distance (HD), and Volume Estimation Error (VEE) for segmentation of brain metastasis various sizes (measured at baseline) using different network architectures. the results are associated with the patients of test set.

Segmentation Model	**Metric**	**Baseline**			**1st Follow-up**			**2nd Follow-up**			**3rd Follow-up**			**4th and 5th Follow-ups**		
		Size ≤ 1 cm	**1 cm < Size ≤ 2.5 cm**	**Size > 2.5 cm**	**Size ≤ 1 cm**	**1 cm < Size ≤ 2.5 cm**	**Size > 2.5 cm**	**Size ≤ 1 cm**	**1 cm < Size ≤ 2.5 cm**	**Size > 2.5 cm**	**Size ≤ 1 cm**	**1 cm < Size ≤ 2.5 cm**	**Size > 2.5 cm**	**Size ≤ 1 cm**	**1 cm < Size ≤ 2.5 cm**	**Size > 2.5 cm**
Cascaded 2D UNets	DSC	82 ± 11	85 ± 7	89 ± 3	79 ± 9	81 ± 6	84 ± 3	65 ± 6	80 ± 11	86 ± 4	59 ± 4	77 ± 9	80 ± 8	62 ± 5	79 ± 9	73 ± 8
	HD (mm)	3.1 ± 0.6	3.0 ± 0.6	2.9 ± 0.3	3.8 ± 0.5	3.7 ± 0.5	3.6 ± 0.3	5.4 ± 0.1	4.5 ± 0.7	4.4 ± 0.6	3.6 ± 0.1	4.6 ± 0.4	3.7 ± 0.2	5.1 ± 0.1	4.0 ± 0.7	4.7 ± 0.3
	VEE (cc)	0.06 ± 0.0	0.3 ± 0.1	1.3 ± 0.2	0.1 ± 0.0	0.4 ± 0.1	1.4 ± 0.2	0.1 ± 0.0	1.1 ± 0.7	0.5 ± 0.3	0.1 ± 0.0	0.7 ± 0.4	1.6 ± 0.6	0.1 ± 0.0	0.8 ± 0.6	1.0 ± 0.5
	VEE (%)	21 ± 10	20 ± 8	18 ± 7	20 ± 9	19 ± 9	18 ± 4	36 ± 1	22 ± 11	21 ± 10	39 ± 4	28 ± 8	14 ± 3	42 ± 1	24 ± 12	35 ± 6
3D UNet	DSC	85 ± 8	87 ± 5	89 ± 2	81 ± 10	83 ± 6	87 ± 3	72 ± 4	83 ± 8	88 ± 3	67 ± 2	82 ± 7	84 ± 6	70 ± 4	83 ± 7	78 ± 6
	HD (mm)	2.7 ± 0.7	2.6 ± 0.8	2.5 ± 0.4	3.4 ± 0.6	3.3 ± 0.6	3.2 ± 0.3	5.0 ± 0.1	4.1 ± 0.7	4.0 ± 0.6	5.2 ± 0.2	4.5 ± 0.4	3.2 ± 0.2	5.1 ± 0.1	4.2 ± 0.6	4.7 ± 0.3
	VEE (cc)	0.05 ± 0.0	0.3 ± 0.2	1.1 ± 0.4	0.08 ± 0.0	0.3 ± 0.1	1.3 ± 0.2	0.08 ± 0.0	1.1 ± 0.7	0.5 ± 0.3	1.0 ± 0.05	0.8 ± 0.5	1.0 ± 0.7	0.1 ± 0.0	0.8 ± 0.7	0.9 ± 0.5
	VEE (%)	16 ± 7	17 ± 7	14 ± 6	20 ± 11	16 ± 8	22 ± 7	32 ± 1	20 ± 10	19 ± 8	33 ± 2	26 ± 8	10 ± 3	35 ± 1	21 ± 9	29 ± 4
Cascaded 2D & 3D UNets	DSC	88 ± 7	89 ± 4	92 ± 2	83 ± 8	85 ± 5	87 ± 2	75 ± 4	85 ± 7	88 ± 3	70 ± 3	83 ± 7	85 ± 6	74 ± 4	85 ± 6	81 ± 6
	HD (mm)	2.5 ± 0.4	2.4 ± 0.4	2.3 ± 0.2	3.2 ± 0.8	3.1 ± 0.9	3.0 ± 0.4	4.6 ± 0.1	3.7 ± 0.7	3.6 ± 0.6	4.0 ± 0.3	3.3 ± 0.4	2.2 ± 0.2	4.2 ± 0.1	3.3 ± 0.6	3.8 ± 0.3
	VEE (cc)	0.04 ± 0.0	0.3 ± 0.1	1.1 ± 0.2	0.05 ± 0.0	0.3 ± 0.2	1.3 ± 0.2	0.1 ± 0.0	0.1 ± 0.6	0.4 ± 0.3	0.1 ± 0.0	0.8 ± 0.5	1.0 ± 0.7	0.08 ± 0.0	0.8 ± 0.6	1.0 ± 0.5
	VEE (%)	13 ± 6	13 ± 6	12 ± 3	19 ± 8	18 ± 9	17 ± 4	27 ± 1	18 ± 8	17 ± 7	30 ± 3	22 ± 5	8 ± 2	32 ± 1	18 ± 9	27 ± 4
Cascaded 2D & 3D UNets + MSGA	DSC	90 ± 6	91 ± 4	**93 ± 2**	**85 ± 8**	**87 ± 5**	**89 ± 2**	**79 ± 3**	**87 ± 5**	**90 ± 2**	**74 ± 2**	**85 ± 6**	**87 ± 5**	**75 ± 4**	**86 ± 6**	**83 ± 6**
	HD (mm)	**2.2 ± 0.6**	**2.1 ± 0.6**	**2.0 ± 0.3**	**2.9 ± 0.7**	**2.9 ± 0.8**	**2.7 ± 0.4**	**3.8 ± 0.1**	**2.9 ± 0.7**	**2.8 ± 0.6**	**3.9 ± 0.1**	3.2 ± 0.4	1.9 ± 0.2	**3.6 ± 0.1**	**2.7 ± 0.6**	**3.2 ± 0.3**
	VEE (cc)	**0.03 ± 0.0**	**0.2 ± 0.1**	1.0 ± 0.2	**0.05 ± 0.0**	**0.3 ± 0.1**	**1.3 ± 0.2**	**0.05 ± 0.0**	**0.8 ± 0.5**	**0.3 ± 0.2**	**0.06±0.05**	0.7 ± 0.5	**0.8 ± 0.6**	**0.05 ± 0.0**	**0.7 ± 0.6**	**0.7 ± 0.5**
	VEE (%)	**12 ± 8**	**9 ± 5**	13 ± 4	**14 ± 5**	**14 ± 5**	**12 ± 3**	**23 ± 1**	**15 ± 6**	**14 ± 5**	**26 ± 2**	**18 ± 4**	**5 ± 1**	**28 ± 1**	**15 ± 8**	**23 ± 4**
nnUNet	DSC	**90 ± 5**	**91 ± 3**	92 ± 1	**85 ± 8**	87 ± 6	89 ± 3	**79 ± 3**	87 ± 6	**90 ± 2**	73 ± 3	**85 ± 6**	**87 ± 5**	**75 ± 4**	**86 ± 6**	**83 ± 6**
	HD (mm)	2.4 ± 0.6	2.3 ± 0.6	2.2 ± 0.3	3.1 ± 0.8	3.0 ± 0.9	2.8 ± 0.4	3.9 ± 0.1	2.9 ± 0.8	2.8 ± 0.7	**3.9 ± 0.1**	**3.1 ± 0.4**	**1.7 ± 0.2**	**3.6 ± 0.1**	**2.7 ± 0.6**	3.3 ± 0.3
	VEE (cc)	0.04 ± 0.0	0.3 ± 0.2	**0.9 ± 0.1**	**0.05 ± 0.0**	**0.3 ± 0.1**	**1.3 ± 0.2**	0.06 ± 0.0	0.9 ± 0.6	0.4 ± 0.2	0.07±0.05	**0.6 ± 0.5**	1.0 ± 0.8	**0.05 ± 0.0**	**0.7 ± 0.6**	0.9 ± 0.4
	VEE (%)	13 ± 7	11 ± 6	**10 ± 3**	16 ± 6	15 ± 6	14 ± 3	24 ± 1	16 ± 7	15 ± 6	26 ± 3	**18 ± 4**	5 ± 2	**28 ± 1**	16 ± 8	24 ± 4

Results of statistical comparisons on percental tumour size changes at each follow-up session relative to the baseline are presented in Table 2.6. The table includes the results of Pearson correlation analyses and paired t-tests (two-sided, $\alpha = 0.05$) performed on the tumour size changes obtained using the automatically generated segmentation masks compared to those based on the

ground-truth masks. The results in this table demonstrate good correlations between the percentual tumour size changes estimated automatically compared to the ground-truth at different follow-ups, with no statistically significant difference between them, where the proposed segmentation framework and the nnUNet model demonstrate better results compared to the other models.

Table 2.5- The 95% confidence intervals associated with the results reported in Table 2.4. The confidence intervals were estimated for the dice similarity coefficient (DSC), Hausdorff distance (HD), and the volume estimation error (VEE) based on the sample mean and standard error of the mean in each experiment using the t distribution.

Segmentation Model	Metric	Baseline			1st Follow-up			2nd Follow-up			3rd Follow-up			4th and 5th Follow-ups		
		Size ≤ 1 cm	1 cm < Size ≤ 2.5 cm	Size > 2.5 cm	Size ≤ 1 cm	1 cm < Size ≤ 2.5 cm	Size > 2.5 cm	Size ≤ 1 cm	1 cm < Size ≤ 2.5 cm	Size > 2.5 cm	Size ≤ 1 cm	1 cm < Size ≤ 2.5 cm	Size > 2.5 cm	Size ≤ 1 cm	1 cm < Size ≤ 2.5 cm	Size > 2.5 cm
Cascaded 2D UNets	DSC	65.0-99.0	80.0-90.0	87.0-91.0	68.0-90.0	76.0-86.0	82.0-86.0	56.0-74.0	70.0-90.0	81.0-91.0	49.0-69.0	66.0-88.0	61.0-99.0	56.0-68.0	70.0-88.0	61.0-85.0
	HD (mm)	2.1 - 4.1	2.5 - 3.5	2.7 - 3.1	3.2 - 4.4	3.3 - 4.1	3.4 - 3.8	5.2 - 5.6	3.9 - 5.1	3.7 - 5.1	3.4 - 3.8	4.1 - 5.1	3.2 - 4.2	5.0 - 5.2	3.3 - 4.7	4.2 - 5.2
	VEE (cc)	0.06-0.06	0.2 - 0.4	1.1 - 1.5	0.1 - 0.1	0.3 - 0.5	1.2 - 1.6	0.1 - 0.1	0.5 - 1.7	0.1 - 0.9	0.1 - 0.1	0.2 - 1.2	0.1 - 3.1	0.1 - 0.1	0.2 - 1.4	0.2 - 1.8
	VEE (%)	6.0 - 36.0	14.0-26.0	13.0-23.0	9.0-31.0	12.0-26.0	15.0-21.0	35.0-37.0	12.0-32.0	9.0-33.0	30.0-48.0	19.0-37.0	7.0-21.0	40.8-43.2	11.4-36.6	25.5-44.5
3D UNet	DSC	73.0-97.0	84.0-90.0	88.0-90.0	69.0-93.0	78.0-88.0	85.0-89.0	66.0-78.0	76.0-90.0	85.0-91.0	62.0-72.0	74.0-90.0	70.0-98.0	65.0-75.0	76.0-90.0	69.0-87.0
	HD (mm)	1.6 - 3.8	2.0 - 3.2	2.2 - 2.8	2.7 - 4.1	2.8 - 3.8	3.0 - 3.4	4.8 - 5.2	3.5 - 4.7	3.3 - 4.7	4.7 - 5.7	4.0 - 5.0	2.7 - 3.7	5.0 - 5.2	3.6 - 4.8	4.2 - 5.2
	VEE (cc)	0.05-0.05	0.1-0.5	0.8-1.4	0.08-0.08	0.2-0.4	1.1-1.5	0.1-0.1	0.5-1.7	0.1-0.9	0.9-1.1	0.2-1.4	0.7-2.7	0.1-0.1	0.1-1.5	0.1-1.7
	VEE (%)	5.0-27.0	11.6-22.4	9.4-18.6	6.3-33.7	9.3-22.7	16.6-27.4	30.4-33.6	10.8-29.2	9.1-28.9	28.0-38.0	16.1-35.9	2.5-17.5	33.8-36.2	11.6-30.4	22.6-35.4
Cascaded 2D & 3D UNets	DSC	76.9-99.1	85.9-92.1	90.5-93.5	73.1-92.9	80.8-89.2	85.5-88.5	68.6-81.4	78.5-91.5	84.3-91.7	62.5-77.5	74.3-91.7	70.1-99.9	69.0-79.0	78.7-91.3	71.5-90.5
	HD (mm)	1.9 - 3.1	2.1 - 2.7	2.1 - 2.5	2.2 - 4.2	2.3 - 3.9	2.7 - 3.3	4.4 - 4.8	3.1 - 4.3	2.9 - 4.3	3.3 - 4.7	2.8 - 3.8	1.7 - 2.7	4.1 - 4.3	2.7 - 3.9	3.3 - 4.3
	VEE (cc)	0.0 - 0.0	0.2 - 0.4	0.9 - 1.3	0.0 - 0.0	0.1 - 0.5	1.1 - 1.5	0.1 - 0.1	0.5 - 0.7	0.0 - 0.8	0.1 - 0.1	0.2 - 1.4	0.7 - 2.7	0.1 - 0.1	0.2 - 1.4	0.2 - 1.8
	VEE (%)	3.5-22.5	8.4-17.6	9.7-14.3	9.1-28.9	10.5-25.5	13.9-20.1	25.4-28.6	10.6-25.4	8.3-25.7	22.5-37.5	15.8-28.2	3.0-13.0	30.8-33.2	8.6-27.4	20.6-33.4
Cascaded 2D & 3D UNets + MSGA	DSC	80.5-99.5	87.9-94.1	91.5-94.5	75.1-94.9	82.8-91.2	87.5-90.5	74.2-83.8	82.4-91.6	87.5-92.5	69.0-79.0	77.6-92.4	74.6-99.4	70.0-80.0	79.7-92.3	73.5-92.5
	HD (mm)	1.2 - 3.2	1.6 - 2.6	1.8 - 2.2	2.0 - 3.8	2.2 - 3.6	2.4 - 3.0	3.6 - 4.0	2.3 - 3.5	2.1 - 3.5	3.7 - 4.1	2.7 - 3.7	1.4 - 2.4	3.5 - 3.7	2.1 - 3.3	2.7 - 3.7
	VEE (cc)	0.03-0.03	0.1-0.3	0.8-1.2	0.05-0.05	0.2-0.4	1.1-1.5	0.05-0.05	0.3-1.3	0.1-0.5	0.14-0.2	0.1-1.3	0.7-2.3	0.05-0.05	0.1-1.3	0.1-1.5
	VEE (%)	0.7-24.7	5.2-12.8	9.9-16.1	7.8-20.2	9.8-18.2	9.7-14.3	21.4-24.6	9.5-20.5	7.8-20.2	21.0-31.0	13.0-23.0	2.5-7.5	26.8-29.2	6.6-23.4	16.6-29.4
nnUNet	DSC	82.0-98.0	88.7-93.3	91.2-92.8	75.1-94.9	82.0-92.0	86.7-91.3	74.2-83.8	81.5-92.5	87.5-92.5	65.5-80.5	77.6-92.4	74.6-99.4	70.0-80.0	79.7-92.3	73.5-92.5
	HD (mm)	1.4 - 3.4	1.8 - 2.8	2.0 - 2.4	2.1 - 4.1	2.2 - 3.8	2.5 - 3.1	3.7 - 4.1	2.2 - 3.6	1.9 - 3.7	3.7 - 4.1	2.6 - 3.6	1.2 - 2.2	3.5 - 3.7	2.1 - 3.3	2.8 - 3.8
	VEE (cc)	0.05-0.05	0.1-0.5	1.0-1.2	0.08-0.08	0.2-0.4	1.1-1.5	0.1-0.1	0.5-1.7	0.3-0.7	0.8-1.2	0.2-1.4	1-3	0.1-0.1	0.2-1.4	0.3-1.5
	VEE (%)	1.9-24.1	6.4-15.6	7.7-12.3	8.6-23.4	10.0-20.0	11.7-16.3	22.4-25.6	9.5-22.5	7.6-22.4	18.5-33.5	13.0-23.0	0.0-10.0	26.8-29.2	7.6-24.4	17.6-30.4

Table 2.7 presents the results of detecting tumour size status at the imaging follow-ups after SRT for patients of the test set using the five different segmentation models. The cascaded 2D & 3D UNets + MSGA architecture demonstrated the best performance with an accuracy of 85.9%, while the nnUNet model resulted in an accuracy of 84.4%.

Table 2.6- Results of statistical comparison on the percental tumour size changes at each follow-up session relative to the baseline obtained from the automatically generated segmentation masks compared to those from the ground-truth masks.

Segmentation Model	Statistical Test	1st Follow-up	2nd Follow-up	3rd Follow-up	4th Follow-up	5th Follow-up
Cascaded 2D UNets	Pearson Correlation (r)	0.82	0.83	0.81	0.79	0.77
	Pearson Correlation (p-value)	3e-6	7e-5	2e-3	6e-3	0.12
	Paired T-Test (p-value)	0.05	0.13	0.08	0.1	0.09
3D UNet	Pearson Correlation (r)	0.83	0.82	0.82	0.79	0.78
	Pearson Correlation (p-value)	2e-6	1e-4	2e-3	6e-3	0.12
	Paired T-Test (p-value)	0.08	0.12	0.15	0.12	0.09
Cascaded 2D & 3D UNets	Pearson Correlation (r)	0.86	0.83	0.83	0.81	0.8
	Pearson Correlation (p-value)	3e-7	7e-5	1e-3	4e-3	0.1
	Paired T-Test (p-value)	0.15	**0.24**	0.08	0.11	0.09
Cascaded 2D & 3D UNets + MSGA	Pearson Correlation (r)	**0.9**	**0.88**	0.85	0.84	**0.84**
	Pearson Correlation (p-value)	**1e-8**	**7e-6**	9e-4	**6e-7**	**0.07**
	Paired T-Test (p-value)	**0.22**	0.2	**0.35**	**0.16**	**0.11**
nnUNet	Pearson Correlation (r)	0.89	**0.88**	**0.86**	**0.85**	**0.84**
	Pearson Correlation (p-value)	3e-8	**7e-6**	**7e-4**	e-6	**0.07**
	Paired T-Test (p-value)	0.2	0.17	0.1	0.08	0.05

Table 2.7- Results of detecting tumour size status at follow-up sessions after SRT for the patients of test set using different segmentation models.

Segmentation Model	Tumour Size Status	Accuracy	Precision	Recall
Cascaded 2D UNets	Increase	71.8%	82.3%	70%
	Stable		57.6%	82.6%
	Decrease		92.8%	62.9%
3D UNet	Increase	79.6%	84%	80%
	Stable		67.8%	82.6%
	Decrease		94%	76.4%
Cascaded 2D & 3D UNets	Increase	82.8%	**90%**	**90%**
	Stable		70%	**91.3%**
	Decrease		**100%**	66.7%
Cascaded 2D & 3D UNets + MSGA	Increase	**85.9%**	**90%**	**90%**
	Stable		75%	**91.3%**
	Decrease		**100%**	76.2%
nnUNet	Increase	84.4%	81.8%	**90%**
	Stable		**76%**	82.6%
	Decrease		**100%**	**81%**

The results of detecting the shrinkage/steady/enlargement categories at individual follow-up scans are presented in Table 2.8. Here, the proposed framework demonstrated a similar performance to that of nnUNet in terms of accuracy when the tumour size changes were

categorized based on the longest diameter of tumour, but it outperformed the nnUNet model when changes in tumour volume were used as the measurement method.

Table 2.8- Results of detecting RANO-BM response categories at individual follow-up scans for the patients of test set using different segmentation models, validated based on the response categories identified from the ground-truth segmentation masks.

Segmentation Model	Tumour Size Status (Response Category)	Measurement Method					
		Longest Diameter of Tumour			Tumour Volume		
		Accuracy	Precision	Recall	Accuracy	Precision	Recall
Cascaded 2D UNets	Enlargement (PD)	72%	70%	70%	68.8%	68.4%	65%
	Steady (SD)		71.4%	74.1%		70.4%	70.4%
	Shrinkage (PR)		75%	70.6%		66.7%	**70.6%**
3D UNet	Enlargement (PD)	78.1%	77.8%	70%	71.9%	75%	75%
	Steady (SD)		81.5%	81.5%		79.2%	70.4%
	Shrinkage (PR)		73.7%	**82.4%**		60%	**70.6%**
Cascaded 2D & 3D UNets	Enlargement (PD)	81.3%	**83.4%**	75%	75.4%	75%	75%
	Steady (SD)		82.1%	**85.1%**		78.6%	78.6%
	Shrinkage (PR)		77.8%	**82.4%**		70.6%	**70.6%**
Cascaded 2D & 3D UNets + MSGA	Enlargement (PD)	**84.4%**	78.3%	**90%**	**81.3%**	**76.2%**	**80%**
	Steady (SD)		**91.7%**	81.5%		**85.7%**	**88.9%**
	Shrinkage (PR)		82.4%	**82.4%**		**80%**	**70.6%**
nnUNet	Enlargement (PD)	**84.4%**	78.3%	**90%**	79.7%	**76.2%**	**80%**
	Steady (SD)		88%	81.5%		85.2%	85.2%
	Shrinkage (PR)		**87.5%**	**82.4%**		75%	**70.6%**

Table 2.9 reports the results of automatic outcome assessment for the test set patients using five segmentation models. The results demonstrate that the proposed framework and the nnUNet model resulted in the best performance with a sensitivity and specificity of 88.9% and 92.3%, respectively, for detecting the LC/LF, and 100% and 89.2% for detecting the ARE outcome.

Table 2.9- Results of detecting the LC/LF and are outcomes for the patients of test set based on the RANO-BM criteria using different segmentation models. Sens: sensitivity; Spec: specificity; Acc: accuracy.

Segmentation Model	LC/LF Detection			ARE Detection		
	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.
Cascaded 2D UNets	72.7%	66.7%	76.9%	77.2%	66.7%	79%
3D UNet	81.9%	77.8%	84.6%	81.9%	66.7%	84.2%
Cascaded 2D & 3D UNets	86.3%	77.8%	**92.3%**	86.4%	**100%**	84.2%
Cascaded 2D & 3D UNets + MSGA	**90.9%**	**88.9%**	**92.3%**	**90.9%**	**100%**	**89.2%**
nnUNet	**90.9%**	**88.9%**	**92.3%**	**90.9%**	**100%**	**89.2%**

Kaplan-Meier analyses were conducted to compare the time to detected event for LF and ARE based on the clinical radiotherapy outcome assessment and the assessment performed by the

proposed automatic system. A log-rank test was applied to evaluate for any statistically significant difference between the curves for each event. Figure 2.5 demonstrates the Kaplan-Meier curves for the LF and ARE events. The curves obtained for the automatic system are similar to their clinically assessed counterparts. No significant difference was observed between the curves for the LF (p-value = 0.95) or ARE (p-value = 0.49) event.

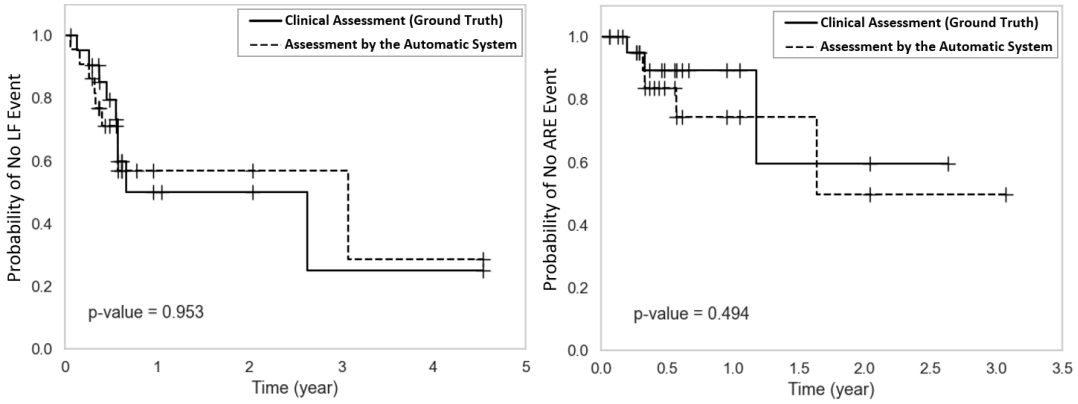

Figure 2.5- Kaplan-Meier curves for comparative time to event analysis between the clinical radiotherapy outcome assessment and the assessment performed by the proposed automatic system on the test set. The plots have been shown for the LF (left) and ARE (right) events. The time to event for each tumour was calculated from the date of radiotherapy to the date an LF/ARE was detected clinically or by the automatic system using the proposed segmentation framework.

2.4. Discussion and Conclusion

In this chapter, a novel system was proposed for automatic assessment of therapy outcome in brain metastasis patients treated with SRT. At the heart of the proposed system is a deep learning-based segmentation framework to delineate tumours longitudinally in serial MRI with high precision. Longitudinal segmentations of tumour before and at multiple follow-up sessions after the SRT permit monitoring changes in tumour size for automatic assessment of therapy outcome based on standard clinical criteria.

The segmentation framework was designed such that it can tackle the memory limitations associated with effective training of complex deep networks by cropping the volumetric images around the tumour. Two cascaded 2D UNets were trained to find the approximate position of the tumour. This position is later used to crop the MRI volume around the tumour. Experimental results show that the cascaded 2D & 3D UNet model could considerably improve the segmentation accuracy compared to the cascaded UNets and the 3D UNet alone. Further, the segmentation framework proposed in this study outperformed the cascaded 2D UNets, the 3D UNet, the cascaded 2D & 3D UNet, and the nnUNet models. By incorporating the MSGA network into the framework, the model benefits from both the cascading and ensembling mechanisms to improve the segmentation accuracy [112]. The MSGA network applies a multi-scale attention mechanism to focus on crucial regions of the images and discard redundancies in the extracted features while learning tumour segmentation. Also, complementary information is provided to the framework through MSGA by feeding T2-FLAIR images as an additional input channel to the MSGA network. As such, fusing the outcome of this network with 3D UNet potentially improves the overall performance of the segmentation framework, as observed in this study. In terms of segmentation time, the framework takes about 5 seconds to find the approximate position of tumour in an input MRI volume and 5 seconds to perform tumour segmentation with the 3D UNet and MSGA networks and generate the final mask, using an Nvidia GeForce RTX 2080 Ti (~10 seconds in total in comparison to ~8 seconds associated with the nnUNet model.)

The performance of the proposed system was subsequently evaluated in monitoring the tumour size status at several imaging follow-ups after SRT. Experimental results demonstrated an accuracy of 86% in detecting tumour size status (increase/stable/decrease), on the independent test.

It should be noted though, that these labels were manually determined at each follow-up by only one observer, and therefore labeling error is expectable due to measurement errors, especially for smaller tumours and those lying closer to the class boundaries. Such errors may affect the reported accuracies in automatic labeling of the tumour size status. Future studies may mitigate possible errors in ground-truth labeling of tumour size status using a multiple observer strategy.

The proposed system also demonstrated a promising performance in detecting tumour size status in terms of response categories at individual follow-up scans, and subsequently automatic assessment of SRT outcome (LC/LF and ARE) on the independent test set. The automatic outcome assessment system in this study evaluates the presence of ARE after radiotherapy based on the pattern of changes in tumour size on serial MRI, with acceptable accuracy. However, it should be noted that monitoring tumour size changes on serial imaging is not always enough to draw an accurate conclusion on whether an observed tumour size increase on imaging is associated with progressive disease or ARE. Along with other radiological insights such as those based on T1/T2 matching or use of perfusion MRI [93], [113], additional clinical evidence including histological confirmation is sometimes required to diagnose ARE. As such, standard serial MRI is usually used by oncologists in conjunction with other clinical criteria to detect pseudo-progression or radiation necrosis after radiotherapy. Considering the performance of the proposed system in accurate tumour segmentation, monitoring tumour size changes longitudinally, and detecting LC/LF and ARE outcomes, it can be applied as an effective decision support system for radiotherapy outcome assessment to triage complicated boundary cases that required further assessment by clinicians.

Previous studies have shown the potential of deep-learning-based methods in automatic brain tumour segmentation and assessment of tumour size changes in response to treatment. Xue

et al. [49] proposed a cascade of modified 3D UNet architecture for the detection and segmentation of brain metastases on 3D T1 MPRAGE images. They proposed the utility of automatically generated segmentation masks for facilitating radiotherapy treatment planning and post-treatment monitoring of tumour size, where they demonstrated example results for one case. Cho et al. [50] developed a CAD system for automated detection of brain metastasis on MRI using a UNet based cascaded model and applied it for categorizing tumour size changes at two follow-up sessions separately, where they achieved a moderate agreement with RANO-BM criteria. The study here features a novel deep-learning-based system for automatic assessment of radiotherapy outcome in brain metastasis using an attention-guided architecture for accurate tumour segmentation. The system was evaluated on multiple MRI scans for each patient to demonstrate its performance in precise tumour segmentation and monitoring tumour size status at individual follow-up sessions, and in detecting LC/LF and ARE outcomes after SRT using the pattern of tumour size changes on serial MRI. The system was also evaluated in terms of similarity of time to detected LF and ARE events compared to those identified clinically. To our knowledge, this is the first time a comprehensive study is performed to investigate the efficacy of deep-learning-based segmentation frameworks for automatic radiotherapy outcome assessment. The findings of this study are in agreement with observations of the previous papers where the potential of data-driven segmentation models was shown in monitoring tumour size changes after treatment, while it extends the preliminary investigations by developing a novel segmentation framework and demonstrating its promising performance for various tasks within a radiotherapy outcome assessment workflow.

Objective assessment of tumour response to therapy has been the basis for many investigations in cancer therapeutics during recent years [68]. The RANO-BM criteria and recommendations were proposed to establish a basis for standard response assessment in clinical trials for brain metastasis. The improved uniformity in response assessment following the RANO-BM criteria facilitates the interpretation of studies involving patients with brain metastasis. This is especially important as the new trend is away from automatically excluding patients with active brain metastasis from the clinical trials of novel therapies [114]. A number of previous studies have explored the RANO-BM criteria as a tool for objective response assessment. Douri et al. [115] evaluated RANO-BM criteria's current threshold in a cohort of 50 patients with brain metastasis treated by SRS. Their findings show that current RANO-BM thresholds are useful in assessing diameter increases caused by tumour progression and pseudo progression but may need adjustments to identify clinically relevant tumour progression reliably. Fischedick et al. [116] compared the 2D linear and 3D volumetric measurement methods for post-SRT monitoring of brain metastasis. The 2D and 3D measurements were categorized according to the RANO-BM criteria and Matthew J. et al. [117], respectively. They concluded that results obtained from the 2D and 3D measurements are highly comparable. While the criteria proposed for volumetric analysis in RANO-BM guideline are incomplete due to lack of research to support specific recommendations, Oft et al. [111] adopted the basic concept from the RANO-BM guideline to derive volumetric criteria and investigated the predictors for volumetric regression after SRT. Their result show that volumetric regression post-SRT does not occur at a constant rate, and a cut-off of ≥20% regression for the volumetric definition of response at 3 months post-SRT was predictive for subsequent control. Further research is required to validate specific threshold

recommendations for volumetric monitoring of brain metastasis after radiotherapy. The automatic system proposed in this paper can facilitate such investigations in the future and is a step forward toward a volumetric radiotherapy response assessment paradigm.

There is a huge interest in finding reliable clinical and/or imaging features that would assist in distinguishing ARE from tumour progression to limit the number of cases triaged for diagnostic biopsy or surgical resection [118]. Various methods such as those based on the qualitative [93] and quantitative [119] assessment of T1/T2 matching, and perfusion [120], [121], and CEST [108] MRI have shown relatively effective with different degrees of accuracy to differentiate ARE from tumour progression. Accurate segmentation of tumour on MRI is a prerequisite for all these methods. Wiggenraad et al. have investigated the use of cine-loops for monitoring tumour size changes in brain metastasis after SRT to identify pseudo-progression (ARE) [122]. They created the cine-loops for ten patients using the axial slice with the largest tumour diameter on pre-treatment contrast-enhanced T1-weighted MRI and the corresponding slices in the co-registered follow-up images. The cine-loops were evaluated by a group of radiation oncologists and neuroradiologists for interpretation of events after SRT, where it was concluded that the use of cine-loops was superior to the assessment of separate MRI scans. To our knowledge, no previous study has investigated the application of automatic brain tumour segmentation on serial MRI for monitoring the pattern of tumour size changes to detect ARE.

One potential limitation associated with this study is its relatively small cohort size. Here, several MRI datasets acquired at different imaging sessions for each patient were applied to evaluate the proposed framework. While the results presented are encouraging and pave the way for future studies, more investigations are required for further evaluation of the proposed

methodologies on larger patient populations and possibly multi-centre imaging data. The patients in this study had relatively large brain metastases treated with hypo-fractionated SRT. Although tumours with a size of 5 mm and above were included in this study, future studies focusing on tumours with size of less than 1 cm are required for further assessment of the performance of the framework on smaller brain metastases typically treated with SRS. ARE in this study was diagnosed clinicoradiologically based on serial imaging, and/or histological confirmation. Diagnosing ARE clinicoradiologically without histological confirmation, however, may be prone to errors due to misinterpretation of images in complicated cases. As such, future studies on imaging datasets with ground-truth histology for all ARE cases are necessary for further validation of the results of this study.

The proposed segmentation framework demonstrated good generalizability in longitudinal segmentation of brain tumours on serial MRI, while it was only trained on the baseline images of the training set. The generalizability of the proposed framework makes it an appropriate fit for the task of automatic therapy outcome assessment. Implementation of the proposed system in clinical settings can potentially accelerate longitudinal tumour size analyses, streamlines image-guided therapy outcome evaluation workflows, *e.g.*, for local response assessment and ARE detection, and facilitates precision oncology through regular and high-throughput response assessment. This is particularly important in the case of patients with multiple brain metastases where manually segmenting tumours on several follow-up scans put a substantial burden on clinical workflow. The system can possibly be coupled with PACS-based databases to perform online and/or offline tumour size analyses on serial imaging and act as an invaluable decision support tool in the clinic. Although a more comprehensive study (along with repeating the experiments through cross-

validation, given the availability of ground-truth masks for all patient follow-ups) is a prerequisite to further validate the results of this study and the clinical utility of the proposed system, the promising results obtained here, and the prospect of its real-world applications highlight the importance of the findings in this study.

Chapter 3

Impact of Tumour Segmentation Accuracy on Efficacy of Quantitative MRI Biomarkers of Radiotherapy Outcome in Brain Metastasis*

3.1. Introduction

While the exact incidence of brain metastasis is difficult to determine, about 70,000 to 400,000 new cases are annually diagnosed in the United States alone [123]. The incidence of brain metastasis appears to be ten times higher than the incidence of primary malignant brain tumours [124]. Because of the accompanying neurologic symptoms, psychological effects, and changes in oncologic treatment plans, the development of brain metastasis may significantly affect a patient's clinical course [125].

Radiation therapy, chemotherapy, immunotherapy, and surgery are the main treatment options for the management of metastatic brain tumours. Because systemic therapy often fails to penetrate the blood-brain barrier, local brain-directed therapies such as radiation or neurosurgical resection are commonly used [126], although there has been some progress in the use of systemic targeted therapy for the management of patients with brain metastasis [127]. Whole-brain radiation

*A version of the material presented in this chapter has been published in Cancers, 14(20): 5133, 2022. DOI: 10.3390/cancers14205133.

therapy (WBRT), hypo-fractionated stereotactic radiotherapy (SRT), and single-fraction stereotactic radiosurgery (SRS) are available options for radiotherapy. Apart from these options, recent studies have shown that a combination of immunotherapy with radiotherapy is associated with improved overall survival compared to radiotherapy alone in patients with brain metastasis who received surgery for their primary cancer [128]. Specially, immunotherapy has shown significant control of intracranial metastasis in patients with melanoma [128]. Moreover, a combination of immunotherapy, *e.g.*, immune checkpoint inhibitors, with radiation therapy has shown promise for the treatment of brain metastasis in non-small-cell lung cancer [129], [130]. Magnetic resonance imaging (MRI) is the primary imaging modality used for diagnosis, treatment planning, and treatment outcome assessment in brain metastasis. The conventional treatment planning and outcome evaluation process includes obtaining MRI scans before (baseline) and after radiation therapy, during several follow-up sessions. This procedure involves precise delineation of the tumour, which is often carried out by experienced radiation oncologists and neuroradiologists. The response assessment in neuro-oncology-brain metastasis (RANO-BM) group [28] has developed standard criteria for evaluating radiotherapy outcome in brain metastasis using serial MRI. The local response of brain metastasis to stereotactic radiation therapy is determined by changes in tumour size on follow-up serial imaging [28] and may be classified into two categories: local control (LC; shrinking or stable tumour) and local failure (LF; enlarging tumour excluding adverse radiation effect). In case of immunotherapy, the immunotherapy response assessment in neuro-oncology (iRANO), a modification of RANO criteria, is used for response assessment [131].

Since the conventional response assessment is based on the changes in tumour size following treatment, it may take months before a local response is apparent on standard follow-up images. This is not desirable, especially given the fact that patients diagnosed with brain metastasis suffer from a short median survival. To make things more complicated, lesion enlargement after treatment on MRI is not necessarily a sign of tumour progression, but also of a condition known as pseudo-progression due to adverse radiation effect [91].

To mitigate these complications and since the local response is highly varied even among the patients going through the same treatment regimen (because of the patient and/or tumour-related factors) there have been efforts to develop a tailored treatment strategy based on the patient's subgroup and predicted survival [32]. Early efforts involved stratifying patients based on factors such as age, performance status, control of primary tumour, and extent of extracranial disease using recursive partitioning analysis (RPA) [12], [30], [132], followed by more complicated stratification methods such as diagnosis-specific graded prognostic assessment (DS-GPA) [31], [32]. Significant prognostic factors such as the primary site of cancer, age, and Karnofsky performance status are used to define the DS-GPA prognostic index. A GPA of 4.0 correlates with the best prognosis, whereas a GPA of 0.0 corresponds to the worst prognosis.

The successful application of artificial intelligence (AI) for diagnostic purposes has led to the development of AI-based cancer imaging analysis, which is now being employed to meet more sophisticated therapeutic requirements, such as patient stratification and therapy outcome prediction. Neuroimaging data can be used to extract quantitative and semi-quantitative image features that are often beyond human vision. Given the abundance of MRI and CT data acquired as part of the standard of care for patients with brain metastasis, there is a huge amount of data

available for mining useful prognostic features to help with response prediction before or early after therapy. Radiomics has been introduced as a formal approach for extracting and discovering quantitative diagnostic and prognostic features from medical images [133]. The quantitative features extracted from the medical imaging data for biomarker discovery in radiomic analysis are usually categorized as [134] morphological features that quantify the geometry and shape of the region of interest such as sphericity, first-order statistical features that describe the voxel intensities without considering the spatial relationship between them such as intensity mean or standard deviation, second-order texture features that are obtained by calculating the statistical interrelationships between the intensity of neighboring voxels, such as those based on grey-level co-occurrence matrix (GLCM), and higher-order statistical features that are obtained after applying a transformation on the image, *e.g.*, features extracted from textural parametric images.

New studies have shown connections between tumour radiomic signatures and their phenotypes, genomic, and proteomic profiles [75], [135]–[137]. Inspired by such connections, several studies have explored the potential of radiomic features in conjunction with machine learning (ML) to develop an efficient and non-invasive method of characterizing metastatic brain tumours and predicting their treatment outcome. Karami et al. [77] have proposed an MRI-based radiomic framework for early prediction of treatment outcome in patients with brain metastasis treated with hypo-fractionated stereotactic radiation therapy (SRT). The proposed quantitative MRI (qMRI) biomarkers were developed through a multistep feature extraction/reduction/selection framework and fed to a support vector machine (SVM) classifier to predict the radiotherapy outcome in terms of LC/LF. A recent study by Mouraviev et al. has investigated whether MRI radiomic features provide any additional value to clinical variables for

predicting local control in brain metastasis following SRS [138]. The results show that the addition of radiomic features to the clinical variables increases the area under the ROC curve considerably.

In quantitative cancer imaging and radiomics, the entire tumour volume should be segmented to determine the region of interest (ROI) for analysis. The segmented tumour boundaries also determine other relevant regions such as the peri-tumoural areas for further analysis. Therefore, accurate tumour delineation is a fundamental step in oncologic radiomics. Manual segmentation of tumour on volumetric images acquired at several imaging sessions for each patient is a tedious and time-consuming job. Automatic segmentation of tumours facilitates radiomic analyses and streamlines the standard process of therapy outcome evaluation in the clinic considerably, possibly at the cost of less accuracy.

In this chapter, we evaluate the impact of tumour segmentation accuracy on efficacy of the quantitative MRI biomarkers for radiotherapy outcome prediction in brain metastasis. A cascaded attention-guided framework is proposed to accurately segment the tumour on the baseline and first follow-up automatically. Using the segmentation masks generated with this framework, we extracted radiomic biomarkers associated with the tumour and peri-tumoural areas from T1-weighted (T1w) and T2-weighted-fluid-attenuation-inversion-recovery (T2-FLAIR) MR images to develop predictive models of radiotherapy outcome. The study results show that to extract meaningful and distinguishing biomarkers, tumour masks should be reasonably accurate but not necessarily matching completely the ground-truth identified manually by expert clinicians. In particular, the results of outcome prediction using the ground-truth and the automatically generated masks are comparable with the top selected biomarkers shared between the two approaches.

3.2. Materials and Methods

3.2.1. Data Acquisition

This research was carried out in compliance with the institutional research ethics board approval from Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada. The imaging data were obtained from 124 patients who had been diagnosed with brain metastasis and were treated with hypo-fractionated SRT over five fractions. Gadolinium contrast-enhanced T1w and T2-FLAIR images were acquired before the treatment (baseline) and at the first follow-up after SRT. Both the T1w and the T2-FLAIR had an in-plane resolution of 0.5 mm. The slice thickness for the T1w and T2-FLAIR images was 1.5 mm, and 5 mm, respectively. T2-FLAIR images were co-registered on the T1w images, and both images were resampled to make the voxel size isotropic ($0.5 \times 0.5 \times 0.5\ mm^3$), rendering the size of both MRI volumes at $512 \times 512 \times 348$ voxels. The voxel intensities in each image were normalized between 0 and 1. The treatment-planning tumour outlines delineated by expert oncologists and neuroradiologists for each patient were included in the dataset. Among 124 patients (156 lesions), 99 patients (116 lesions) were randomly assigned for training and optimization of the predictive models (10 patients with 15 lesions as the validation set for optimizing the model hyperparameters) and 25 patients (40 lesions) were kept unseen as an independent test set.

After SRT, the patients underwent follow-up MRI scans every two to three months. The lesions were monitored using serial MRI, and a radiation oncologist and a neuroradiologist determined the local response for each lesion using the RANO-BM [28] criteria. The local outcome was categorized as either LC (complete response, partial response, or stable disease) or LF (progressive disease) based on the response determined in the last follow-up session. Using serial

imaging (including perfusion MRI) and/or histological confirmation, the adverse radiation effect (ARE) was diagnosed and distinguished from local progression [92], based on the report by Sneed et al. [91]. Following these criteria, a total of 93 and 63 lesions were classified as LC and LF, respectively.

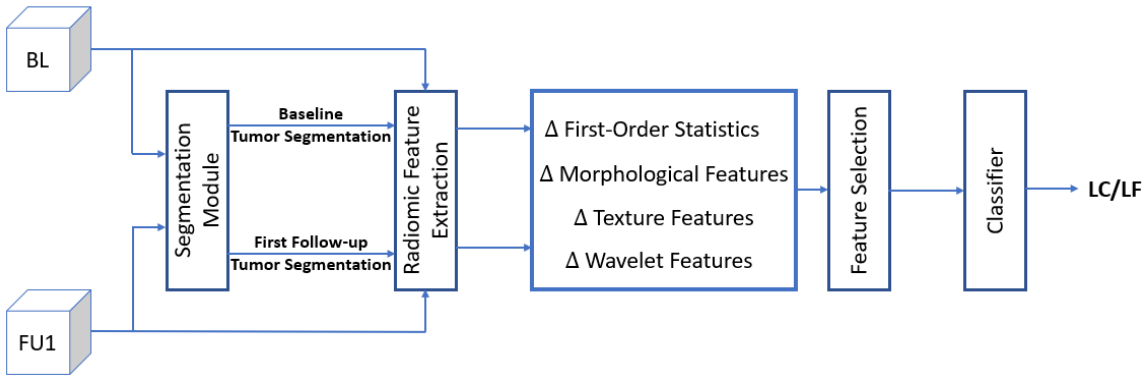

Figure 3.1- The overall system for SRT outcome prediction. The segmentation module segments tumours on the baseline (BL) and the first follow-up (FU1) MR images. Using the MR images and segmentation masks, different radiomic features are extracted. The features are then sorted based on their relevance/redundancy and a subset of them identified through feature selection are used to train a classifier for predicting the local outcome.

3.2.2. System Overview

Figure 3.1 shows the overall framework adapted for predicting the local outcome in patients diagnosed with brain metastasis. Since a tumour segmentation mask is necessary for extracting radiomic features, in this study we replaced the manual tumour masks with the ones generated automatically by our proposed segmentation framework. Using the MR images of both modalities (T1w and T2-FLAIR) and the automatically generated segmentation masks, the radiomic features are extracted from the tumour and peri-tumoural regions for both the baseline and first follow-up sessions. The change in each feature is then calculated at the first follow-up relative to the baseline. The number of features is reduced using feature selection techniques. The selected features are

then fed to train a classifier to predict the therapy outcome. More details on different components of the framework are provided in the following subsections.

3.2.3. Segmentation Module

The segmentation module applied in this chapter and its components have been described in detail in Section 2.2.2.

3.2.4. Radiomic Feature Extraction

Using the radiomic features, a quantitative description of each tumour can be derived from imaging data. A total of 3,436 radiomic features were extracted from the tumour region and its 5 mm margin [77] on T1w and T2-FLAIR MR volumes acquired at baseline and the first follow-up using the automatically generated masks. The extracted radiomic features included the first-order statistics (including energy, entropy, etc., a total of 19 features), 2D and 3D morphological features (including maximum 2D diameter, flatness, sphericity, a total of 26 features), and texture features derived based on the gray level co-occurrence matrix (GLCM) (total of 24 features), gray level run length matrix (GLRLM) (total of 16 features), gray level size zone matrix (GLSZM) (total of 16 features), neighbouring gray tone difference matrix (NGTDM) (total of 5 features), and gray level dependence matrix (GLDM) (total of 14 features). All features were extracted using the Pyradiomics package in Python [139]. The morphological features were derived from the generated binary masks. Other features were extracted from both the original MR images and the associated wavelet-filtered images. For the latter, low- and high-pass wavelet filters were applied in the x, y, and z directions of the 3D images resulting in eight filtered images of I_{HHH}, I_{HHL}, …, I_{LLL} for each original image. The delta radiomic features were calculated using the relative difference between the value of each feature at baseline and the first follow-up.

3.2.5. Feature Extraction

In machine learning, feature selection approaches allow to save computation time, improve prediction performance, and gain a deeper knowledge of the data [140]. The delta radiomic features were processed through a feature selection procedure using the minimum redundancy maximum relevance (mRMR) method [141] with a mutual information quotient (MIQ) criterion. The mRMR method tends to select a feature subset with high correlation to the target class (output) and low inter-feature correlations. The F-statistics were used to calculate the correlation with the target class (relevance), while the Pearson correlation coefficient was applied to calculate the inter-feature correlations (redundancy). The MIQ score represents the quotient of relevance and redundancy. Ultimately, among the 3,436 features, seven features were selected using the mRMR technique. For the feature selection process, only the training data was applied to prevent data leakage from the training set to the test sets.

3.2.6. Classifier

A support vector machine (SVM) with a Gaussian kernel was adapted as the classifier for outcome prediction because of its demonstrated performance, versatility coming from different kernels, and effectiveness in high dimensional space [142]. The selected delta radiomic features in the training set were normalized between 0 and 1 and used to train the SVM classifier with various hyperparameters. The best SVM model was selected based on the performance of the model on the validation set using the area under the receiver operating characteristics (ROC) curve (AUC) criterion. In this model, the penalty (C), gamma, and the tolerance parameters were set to 1, $\frac{1}{num_features}$, and $1e-3$, respectively. A class weight inversely proportional to the class frequencies was also assigned to each class. Finally, the model was tested on the independent test

set, where the corresponding features were normalized using the normalization coefficients obtained for the training set.

3.3. Results

Table 3.1 demonstrates the results obtained with the proposed framework on the training and test sets for tumour segmentation at the baseline and first follow-up in terms of dice similarity coefficient (DSC), Hausdorff distance (HD), and volume estimation error (VEE). For comparison, the segmentation results are also reported for the cascaded 2D UNets, 3D UNet, and the cascaded 2D and 3D UNets. The results of Table 3.1 demonstrate the superiority of the proposed framework compared to the other models in accurately segmenting tumours on the baseline and first follow-up which later demonstrates its importance in predicting the local response to radiotherapy.

Table 3.2 shows the most important features selected using the mRMR feature selection method for different segmentation models. A comparison between the selected features extracted using the ground-truth masks and the cascaded 2D & 3D UNets + MSGA model shows that six out of seven features are shared between the two models. This indicates that the tumour masks generated using the proposed segmentation framework may be sufficiently accurate for MRI radiomic analysis in this application. The number of shared features with the ground-truth gradually reduces when the accuracy of segmentation models decreases, with no shared feature for the cascaded 2D UNets.

Figure 3.2 demonstrates the parametric maps of the top four selected features based on the ground-truth (Table 3.2) overlayed on the baseline and follow-up images of a representative tumour with an LF outcome. The parametric maps are shown for the ground-truth mask and the

automatically generated segmentation masks obtained from different models. The figure demonstrates that the parametric maps and the relative changes of features from the baseline are in good agreement with the ground-truth for the models with acceptable segmentation accuracy.

Table 3.1- Dice similarity coefficient (DSC), the 95% confidence interval (CI) of DSC, Hausdorff distance (HD), and volume estimation error (VEE) for segmentation of brain metastasis using different network architectures.

Segmentation Model		**Baseline**		**First Follow-up**	
		Training Set	**Test Set**	**Training Set Patients**	**Test Set Patients**
Cascaded 2D UNets	DSC	0.85 ± 0.05	0.85 ± 0.06	0.82 ± 0.05	0.81 ± 0.07
	DSC CI	0.84 – 0.86	0.83 – 0.87	0.81 – 0.83	0.79 – 0.81
	HD	3 ± 0.6 mm	3.4 ± 0.7 mm	3.46 ± 0.6 mm	3.7 ± 0.5 mm
	VEE	0.63 ± 0.44 cc 17.2%± 6.1%	0.71 ± 0.47 cc 19.4% ± 8.3%	0.75 ± 0.5 cc 20.6% ± 8.5%	0.77 ± 0.51 cc 21.5% ± 9%
3D UNet	DSC	0.87 ± 0.06	0.85 ± 0.06	0.85 ± 0.05	0.82 ± 0.06
	DSC CI	0.86 – 0.88	0.83 – 0.87	0.84 – 0.86	0.80 – 0.84
	HD	2.9 ± 0.8 mm	3.1 ± 0.82 mm	3.2 ± 0.85 mm	3.5 ± 0.6 mm
	VEE	0.6 ± 0.42 cc 15.8% ± 5.4%	0.7 ± 0.45 cc 17% ± 7%	0.72 ± 0.47 cc 18% ± 7.3%	0.75 ± 0.53 cc 18.9% ± 9.5%
Cascaded 2D & 3D UNets	DSC	0.89 ± 0.05	0.88 ± 0.05	0.86 ± 0.05	0.83 ± 0.05
	DSC CI	0.88 – 0.90	0.86 – 0.90	0.85 – 0.87	0.81 – 0.85
	HD	2.45 ± 0.6 mm	2.65 ± 0.63 mm	2.8 ± 0.6 mm	3.1 ± 0.5 mm
	VEE	0.55 ± 0.35 cc 13.1%± 4.2%	0.61 ± 0.4 cc 15.8% ± 6.5%	0.64 ± 0.43 cc 16.6% ± 6.8%	0.68 ± 0.5 cc 17.9% ± 7.7%
Cascaded 2D & 3D UNets + MSGA	DSC	**0.91 ± 0.03**	**0.90 ± 0.04**	**0.89 ± 0.04**	**0.87 ± 0.05**
	DSC CI	**0.90 – 0.92**	**0.89 – 0.91**	**0.88 – 0.90**	**0.85 – 0.89**
	HD	**2.1 ± 0.45 mm**	**2.3 ± 0.55 mm**	**2.21 ± 0.5 mm**	**2.74 ± 0.49 mm**
	VEE	**0.42 ± 0.3 cc 11.2% ± 3.9%**	**0.53 ± 0.36 cc 12.8 % ± 5.1%**	**0.57 ± 0.38 cc 14.7% ± 4.7%**	**0.61 ± 0.48 cc 15.9% ± 5.1%**

Table 3.3 shows the results of outcome prediction on the independent test set in terms of accuracy, sensitivity, specificity, AUC, and F1-score, using the selected radiomic features associated with different tumour segmentation models.

Table 3.2- List of the top seven selected features using the mRMR feature selection method for different segmentation models. Bold features are the shared ones with those obtained using the ground-truth segmentation masks.

Segmentation Model	Selected Features
Cascaded 2D UNets	wavelet-LLH_glcm_Correlation_T2_Margin
	wavelet-LHH_glrlm_RunVariance_T2
	original_gldm_DependenceVariance_T2_Margin
	wavelet-HLH_glcm_Imc2_T2_Margin
	wavelet-LHH_glcm_Idm_T2_Margin
	wavelet-LHL_gldm_SmallDependenceHighGrayLevelEmphasis_T1
	wavelet-HHH_glszm_ZonePercentage_T1_Margin
3D UNet	**wavelet-HLH_gldm_LargeDependenceLowGrayLevelEmphasis_T2**
	original_gldm_DependenceEntropy_T2_Margin
	wavelet-LHL_gldm_SmallDependenceHighGrayLevelEmphasis_T1
	wavelet-HLH_gldm_SmallDependenceLowGrayLevelEmphasis_T2
	wavelet-HHL_ngtdm_Contrast_T2
	wavelet-HLH_glcm_Imc2_T2_Margin
	original_gldm_DependenceVariance_T2_Margin
Cascaded 2D & 3D UNets	**wavelet-HHL_firstorder_Minimum_T1_Margin**
	original_gldm_DependenceEntropy_T2_Margin
	original_glcm_Idn_T2_Margin
	wavelet-HLH_gldm_LargeDependenceLowGrayLevelEmphasis_T2
	wavelet-LHL_glcm_Contrast_T1
	wavelet-HHH_gldm_DependenceVariance_T1_Margin
	wavelet-LHL_gldm_SmallDependenceHighGrayLevelEmphasis_T1
Cascaded 2D & 3D UNets + MSGA	**wavelet-HHL_firstorder_Minimum_T1_Margin**
	original_gldm_DependenceEntropy_T2_Margin
	original_glcm_Idn_T2_Margin
	wavelet-HLH_gldm_LargeDependenceLowGrayLevelEmphasis_T2
	wavelet-LLL_ngtdm_Strength_T1_Margin
	wavelet-HLL_glcm_Idn_T1_Margin
	wavelet-HHL_firstorder_Skewness_T1
Ground-Truth	wavelet-HHL_firstorder_Minimum_T1_Margin
	original_gldm_DependenceEntropy_T2_Margin
	original_glcm_Idn_T2_Margin
	wavelet-HLH_gldm_LargeDependenceLowGrayLevelEmphasis_T2
	wavelet-LLL_ngtdm_Strength_T1_Margin
	wavelet-HLL_glcm_Idn_T1_Margin
	wavelet-LHH_glszm_SizeZoneNonUniformityNormalized_T1_Margin

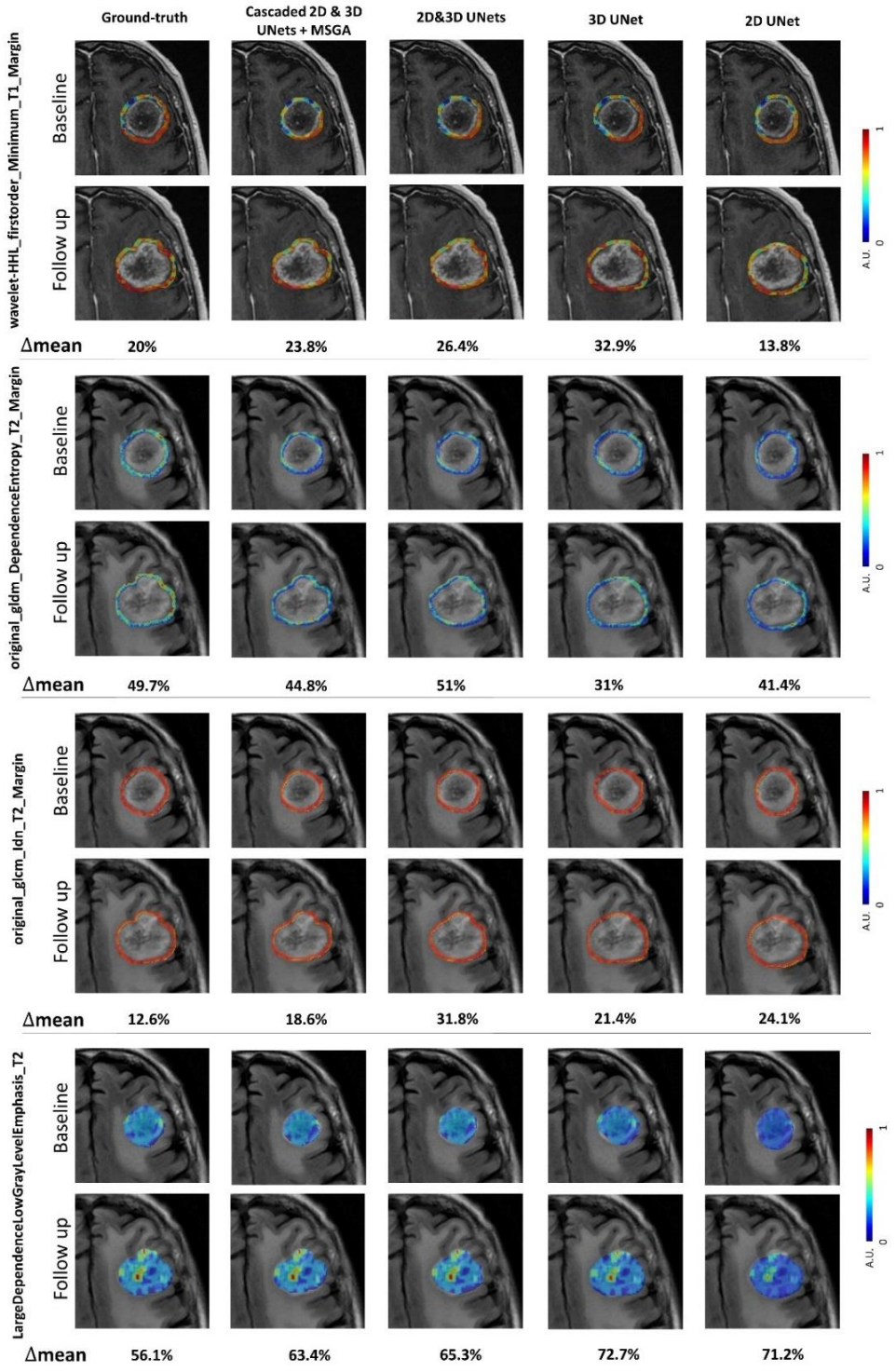

Figure 3.2- Representative parametric maps of the top four selected features based on the ground-truth (Table XVI) overlayed on the associated MR images acquired at the baseline and the first follow-up from a representative tumour with an LF outcome. The parametric maps are shown for the ground-truth mask and the segmentation masks generated automatically using different models. Δmean for each feature is the mean relative change from the baseline at the first follow-up that is calculated for different masks.

The results in Table 3.3 show that the prediction performance of the radiomic model associated with the tumour masks generated automatically using the cascaded 2D & 3D UNets + MSGA network is reasonably close to the ground-truth model (AUC of 0.78 versus 0.81). On the other hand, the tumour masks generated using the cascaded 2D UNets or 3D UNet have resulted in considerably lower performances of the corresponding radiomic models (AUC of 0.62 and 0.67, respectively) compared to the ground-truth. The sensitivity and specificity of the ground-truth model versus the model developed with the cascaded 2D & 3D UNets + MSGA network are 83% and 78% versus 77% and 83%. This also indicates reasonably close performance of the ground-truth masks and those generated automatically by the proposed segmentation framework in radiomic modeling for treatment outcome prediction. The sensitivity of the ground-truth model is slightly better which shows it can detect local failure cases more accurately. On the other hand, the model developed with the cascaded 2D & 3D UNets + MSGA network performs better in detecting the local control cases. These observations imply that although accurate segmentation of tumour region is important in developing radiomic-based predictive models of therapy response, such models can be reasonably robust to imperfections in their input tumour masks.

Table 3.3- Results of therapy outcome prediction using the radiomic models developed with different segmentation modules.

Segmentation Model	**Independent Test Set**				
	Accuracy	**Sensitivity**	**Specificity**	**AUC**	**F1-Score**
Cascaded 2D UNets	72.5%	70.6%	74%	0.62	68.5%
3D UNet	72.5%	70.6%	74.%	0.67	68.5%
Cascaded 2D & 3D UNets	77.5%	76.5%	78.2%	0.72	74.3%
Cascaded 2D & 3D UNets + MSGA	80%	76.5%	82.6%	0.78	76.5%
Ground-Truth	80%	82.5%	78.2%	0.81	77.8%

Figure 3.3 demonstrates the results of survival analysis on the test set. The Kaplan–Meier progression-free survival curves are presented for two patient cohorts classified based on their predicted outcome by the radiomic models developed with different segmentation modules. A log-rank test applied on the survival curves of the two cohorts demonstrates no statistically significant difference for the models developed with cascaded 2D UNets or 3D UNet. However, a significant difference is observed between the survival curves of the cohorts stratified by the models developed with cascaded 2D & 3D UNets, cascaded 2D & 3D UNets + MSGA, and the ground-truth masks.

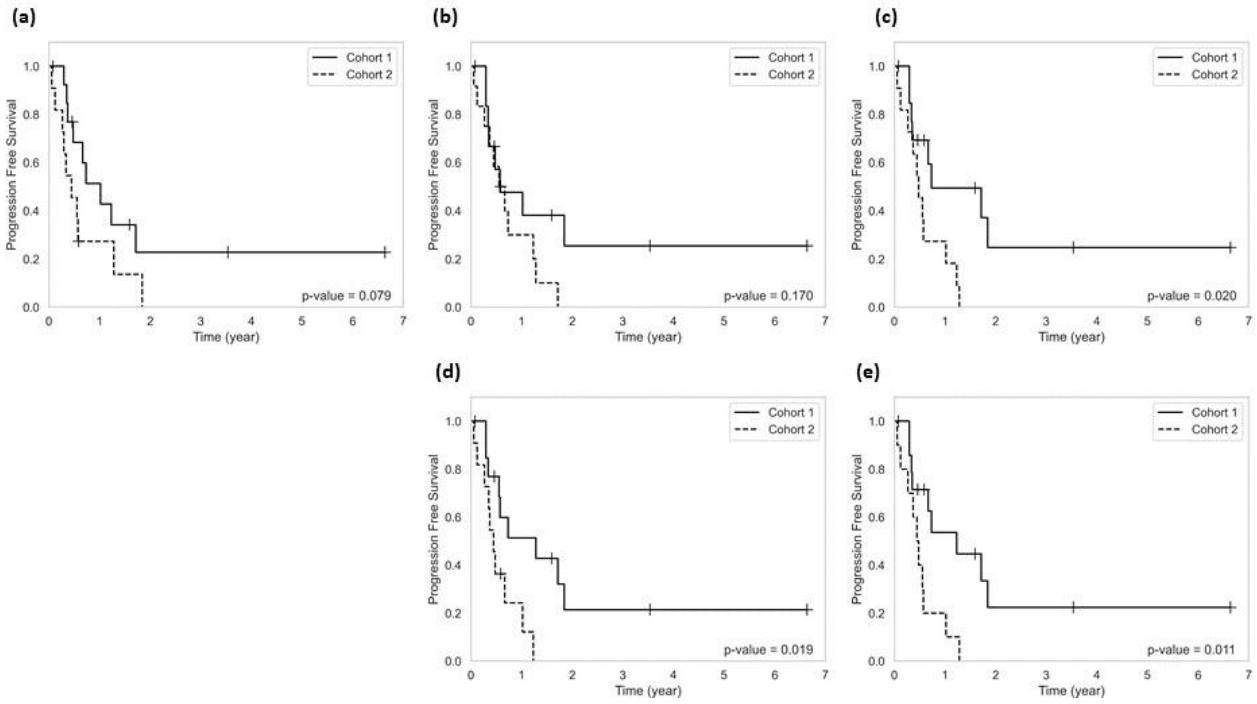

Figure 3.3- Kaplan-Meier progression-free survival curves for two cohorts of patients stratified based on the outcome prediction by the radiomic models developed with different segmentation modules: (a) Cascaded 2D UNets, (b) 3D UNets, (c) Cascaded 2D & 3D UNets, (d) Cascaded 2D & 3D UNets + MSGA, and (e) Ground-truth masks. Cohort 2 includes the patients in the independent test set who had at least one lesion with a predicted outcome of local failure, and cohort 1 includes all other patients in the independent test set.

3.4. Discussion and Conclusion

Accurate segmentation of brain tumours on the baseline and follow-up images is an essential yet laborious task. It is required for therapy response assessment as well as for extracting imaging biomarkers in various diagnostic and prognostic applications. An automatic segmentation method streamlines the therapy outcome prediction and evaluation workflow (which requires ROI masks for analysis), potentially at the cost of less accuracy. In this study, we investigated the effect of tumour segmentation accuracy on efficacy of the MRI radiomic biomarkers extracted from the tumour and tumour margin for radiotherapy outcome prediction in brain metastasis. The results of our study show that while the impact of tumour delineation accuracy is considerable for less accurate segmentation models (dice score ≤ 0.85), the radiomic features and prediction models are relatively robust to imperfections in the generated tumour masks. In particular, the list of selected features associated with the ground-truth tumour masks and those generated using the cascaded 2D & 3D UNets + MSGA model (with an average dice score of about 0.90) shared six out of the seven features selected based on the mRMR method. Furthermore, the features extracted using the tumour masks generated by this segmentation model resulted in an outcome prediction model with fairly close performance metrics to those obtained by the ground-truth model, including the accuracy, sensitivity, specificity, AUC, and F1-score of %80 versus 80%, 77% versus 83%, 83% versus 78%, 0.78 versus 0.81, and 77% versus 78%, respectively. Progression-free survival analyses demonstrated that both the outcome prediction models could stratify the patients into two cohorts (low-risk versus high-risk) with statistically significantly different survival curves.

The effect of segmentation accuracy on the performance of radiomic features for different cancer sites and imaging modalities is a subject yet to be explored in the literature. Among research

done so far, Jin et al. studied the effect of automatic segmentation using multiple UNet-based architectures on the accuracy of radiomic features for transvaginal ultrasound images of cervical cancer [143]. The results of that study show the feasibility and reliability of automatic segmentation, especially with UNet-based models, for relevant radiomic studies. This is in agreement with the observation of this study where fairly accurate tumour segmentation was found to be sufficient for radiomic models of cancer therapy response. Teng et al. studied the effect of automatic segmentation on preoperative lymph node status prediction models with radiomic features extracted from ultrasound for patients with early-stage cervical cancer. Their study shows that in some cases, automatic segmentation improves the prediction accuracy as human-derived segmentation methods introduce human bias into the radiomic process [144]. This is an interesting observation that may be valid only for smaller ROIs such as lymph nodes and/or low-SNR imaging modalities such as ultrasound but should be rigorously investigated in future works on large sample sizes for different cancer sites and imaging modalities.

The findings of this study show that radiomics in conjunction with machine learning can be used to predict radiotherapy outcome in brain metastasis early after treatment, where automatic tumour segmentation could potentially be utilized instead of manual segmentation to facilitate prediction model development and investigation. In the clinic, these implications could be significant as an early prediction of treatment response may lead to therapy adjustments which in turn, enhance patients' survival and quality of life [8]. The observation of this study that radiotherapy outcome prediction in brain metastasis is not very sensitive to small inaccuracies in tumour segmentation permits high-throughput implementation and exploration of new computational prediction models to develop robust systems for clinical decision support.

The promising result of this study opens the avenue to applying automatically generated segmentation masks for discovering diagnostic and prognostic biomarkers in brain metastasis without sacrificing their accuracy. This is a significant contribution considering the heavy burden that manual segmentation imposes on image-guided therapeutic systems in neuro-oncology, including the models for therapy outcome prediction and the platforms for therapy outcome evaluation. One potential future direction for this work is to report the performance of the radiomic models with respect to the variabilities in the tumour size, tumour origin, age, sex, etc., given the availability of a larger dataset. Also, an ablation study on the selected radiomic features can explore the contribution of individual features to the model. Investigating the effects of controlled/consistent over- and under-segmentation of tumours on the prediction performance of the radiomic models can also be the subject of future studies. Finally, while the findings of this paper are promising and pave the way for future research, future investigations are required to further assess the conclusions of this study on a larger scale when imaging data are available from larger patient cohorts and possibly multi-center studies.

Chapter 4

Predicting the Outcome of Radiotherapy in Brain Metastasis by Integrating the Clinical and MRI-Based Deep Learning Features*

4.1. Introduction

Brain metastases are the most prevalent malignancy of the central nervous system, with an incidence rate of 10% to 30% among cancer patients [3]. Because of the overall improvement in healthcare and the longer survival of cancer patients, the incidence of brain metastasis is projected to rise [2].

Timely diagnosis and precise treatment are critical in the survival of patients suffering from brain metastasis. The origin of cancer, size/number of metastases, and associated symptoms are important factors in planning a treatment strategy for brain metastasis. Surgery, radiation therapy, and chemotherapy are the main treatment options for the management of metastatic brain tumours.

Local response to radiotherapy is highly varied among patients despite administrating standardized dose/fraction regimens due to many tumour-related factors such as histology, tumour

*A version of the material presented in this chapter has been published in Medical Physics, 49(11): 7167-7178, 2022. DOI: 10.1002/mp.15814.

size, and location. Additionally, patient-related factors such as genetics, age, and performance status are also predictors of tumour response [29]. Local response of brain metastasis to stereotactic radiation therapy is assessed based on the changes in the tumour size on follow-up serial images [28], and can be categorized into local control (LC; shrinking or stable tumour) versus local failure (LF; enlarging tumour). It may take months, however, before a local response is evident on follow-up images, let alone the fact that early changes in tumour size are not always correlated with long-term local response. In particular, post-treatment lesion enlargement on imaging may not always be a sign of tumour progression, but rather of a condition known as pseudo-progression due to adverse radiation effect [122]. Given the median survival of as short as 5 months and up to 4 years [7], [34], early prediction of LF after radiotherapy can facilitate effective treatment adjustments, potentially resulting in improved treatment outcomes, survival, and quality of life.

The therapeutic paradigm for brain metastasis has steadily shifted to focus more on tailored treatments based on the subgroups and predicted survival [32]. Recursive partitioning analysis (RPA) was one of the first methods to classify patient prognosis based on age, performance status, control of primary tumour, and extent of extracranial disease [12], [30], [132]. A success at the time but overly simplistic, RPA is now replaced by more sophisticated stratification methods such as diagnosis-specific graded prognostic assessment (DS-GPA) [31], [32]. In this method, a GPA of 4.00 and 0.00 is associated with the best and the worst prognosis, respectively. The DS-GPA is calculated using prognosis factors based on the primary site of cancer, age, and Karnofsky performance status (KPS) [145]. The KPS is a clinical metric to quantify the ability of cancer patients to perform everyday activities, with a score in the range of 0-100.

Prognostic features could also be mined from textural information on medical images. A large body of research shows that imaging modalities, such as magnetic resonance imaging (MRI), potentially provide relevant prognostic information that, if appropriately retrieved, can be utilized to predict therapy outcome [134], [146], [147]. Radiomics is an emerging translational field of research concerned with the high-throughput mining of high-dimensional medical imaging data to discover quantitative diagnostic and prognostic features [44]. Studies show the effectiveness of radiomic features as prognostic factors. Karami et al. have proposed an MRI-based radiomic framework for early prediction of treatment outcome in brain metastasis patients treated with SRT [148]. Liao et al. [142] explored the use of radiomics and clinical features in conjunction with support vector machines to predict survival and local response of the tumour for patients diagnosed with brain metastasis and treated with Gamma Knife radiosurgery. Their study shows that combining clinical and radiomic features improve the capability of the model to predict both tumour's local response and overall survival. Mouraviev et al. [138] proposed extracting radiomic features from the tumour core and the peritumoural regions and trained a random forest classifier on these features to predict local control in brain metastasis treated with stereotactic radiosurgery. Their result shows that an optimized combination of radiomic and clinical features resulted in a 19% increase in the AUC compared to the clinical features alone. Some studies also demonstrate correlations between the radiomic signature of tumours and their phenotypes and genomic and proteomic profiles [75].

While radiomic features are handcrafted, deep learning algorithms could be used to extract distinguishing relevant imaging features automatically. Deep learning models have proven to be quite effective at identifying important and distinct characteristics, particularly in image data [79],

[80]. Deep models have the capability to outline regions of interest automatically, capture textural changes within a lesion, discriminate between cancerous and non-cancerous cells, and potentially extract distinctive information from lesions to be later used for the task of outcome prediction [54], [55], [81]–[88]. Diamant et al. [53] hypothesized that convolutional neural networks could enhance the performance of traditional radiomics, by detecting image patterns that may not be covered by a traditional radiomic framework. They tested their hypothesis for the task of head & neck cancer therapy outcome prediction and their results show that deep models can explicitly recognize traditional radiomic features and perform accurate outcome prediction. A recent study by Cho et al. [89] suggests that using deep learning methods instead of classic machine learning results in brain metastasis detection with a lower false-positive rate.

Many scenarios in medical imaging analysis require dealing with a sequence of spatially connected pictures, *i.e.*, 3D volumes, such as MRI volumes. Recurrent neural networks (RNN) are a suitable fit for dealing with spatial dependency. Recurrent neural networks allow quantifying the information persisting between image slices, whereas typical 2D convolutional neural networks (CNN) alone do not [149]. Long short-term memory (LSTM) networks are a special kind of RNNs capable of learning long-term dependencies [150].

In this chapter, a novel deep learning framework is proposed and investigated, for the first time, to predict local failure in brain metastasis treated with SRT using the treatment planning MRI and clinical information available at pre-treatment. The framework consists of a CNN (InceptionResNetV2) to extract textural features from single 2D slices in an MRI volume, followed by an LSTM network to account for the spatial dependency between the 2D slices. The framework is capable of integrating the conventional clinical factors such as histology, tumour location, size,

and the number of brain metastases, with the deep learning features of MRI in a comprehensive data-driven model for therapy outcome prediction. The results of the study show that coupling the clinical factors and deep-learning-based MRI features associated with the entire tumour volume improves the performance of therapy outcome prediction model considerably. For further comparison, two other models with state-of-the-art architectures, namely, sequence to sequence (Seq2Seq) and transformer networks, were investigated to incorporate inter-slice dependencies. The results demonstrate that the proposed model with the LSTM network could outperform the other two models in therapy outcome prediction. The results obtained with advanced methods of visualizing the network's decision basis highlight the importance of tumour/lesion margin characteristics on MRI in therapy response prediction.

4.2. Materials and Methods

4.2.1. Study Protocol and Data Acquisition

This study was conducted in accordance with intuitional research ethics approval from Sunnybrook Health Sciences Centre (SHSC), Toronto, Canada. The imaging and clinical data were collected from 124 patients diagnosed with brain metastasis and treated with SRT over five fractions. The imaging data applied in this study for therapy outcome prediction included gadolinium contrast-enhanced T1-weighted (T1w), and T2-weighted-fluid-attenuation-inversion-recovery (T2-FLAIR) images acquired before the treatment (baseline). The in-plane image resolution and the slice thickness were 0.5 and 1.5 mm for the T1w and 0.5 and 5 mm for T2-FLAIR images, respectively. The dataset also included the treatment-planning tumour contours for each patient delineated by expert oncologists and the edema contours outlined under their supervision. Among the 124 patients (156 lesions), 99 patients (116 lesions) were randomly

selected for training and optimization of the predictive models (10 patients with 15 lesions as the validation set for optimizing the model hyperparameters) and 25 patients (40 lesions) were kept unseen as an independent test set. The distribution of samples in terms of clinical attributes including age, gender, tumour size, histology, and outcome were inspected in the training and test sets to ensure statistical similarity.

The patients were followed up with MRI after radiotherapy on a 2 to 3-month schedule. The median imaging follow-up for all patients was 8 months. The lesions were monitored on serial MRI and the local response was determined for each lesion by a radiation oncologist and neuroradiologist based on the RANO-BM criteria [28]. The local outcome was defined as LC (complete response, partial response, or stable disease) or LF (progressive disease) identified in the last patient follow-up. Adverse radiation effect (ARE) was diagnosed and differentiated from local progression using serial imaging (including perfusion MRI) and/or histological confirmation [92] based on the report by Sneed et al. [91]. In keeping with these, 93 and 63 lesions were categorized as an LC and LF outcome, respectively.

4.2.2. Preprocessing

The baseline T1w and T2-FLAIR images were resampled with a voxel size of $0.5 \times 0.5 \times 1$ mm^3. The resampled MRI volumes had a size of $512 \times 512 \times 174$ voxel. The T2-FLAIR images were co-registered on T1w images using an affine registration. To ensure a local outcome prediction on separated lesions, the size of the smallest sub-volume that encompasses the entire region of interest (ROI), including the tumour and edema (lesion) and their 5-mm outer margin [51], was determined for the individual lesions. Observations of the study presented in [51] demonstrate that MRI

radiomic features derived from the lesion margin (3-5 mm) can contribute to radiotherapy outcome prediction models in brain metastasis. A sub-volume of $128 \times 128 \times 45$ voxels was determined as a fit standard for all lesions. Subsequently, the sub-volumes associated with individual lesions were cropped from the T1w and T2-FLAIR images. The tumour and edema contours were used to generate ROI masks (tumour + 5-mm margin for T1w; tumour + edema + 5-mm margin for T2-FLAIR) for each lesion that were used to mask out the areas outside the ROI within the cropped sub-volumes. The voxel intensities in each image were normalized to the range of 0 and 1.

4.2.3. Clinical Features

Standard clinical features at the baseline including the histology (primary cancer), tumour location (infratentorial/ supratentorial), tumour size (longest diameter in mm), number of brain metastases, total dose (Gy), previous WBRT (yes/no), prior SRT/SRS (yes/no), GPA (from 0 to 4) along with age (year) and gender (male/female) were collected for each patient and their performance was investigated in therapy outcome prediction with and without the deep learning features of MRI. The categorical features were converted to vectors using one-hot encoding, while the continuous features (*e.g.*, tumour size) were first discretized to categories and then converted to one-hot encoding format. A 3-layer fully-connected multi-layer perceptron (MLP) was trained and optimized using the training and validation data to predict LC/LF for each lesion solely with the clinical features. The MLP model included an input layer with 2 to 48 neurons (depending on the input features), one hidden layer with 10 neurons, and an output layer with two neurons (LC/LF). Feature selection was performed through an exhaustive search among all possible combinations of the features to obtain the best feature set based on the accuracy of the model on the validation set.

The selected features were also coupled with the deep learning features in the comprehensive model developed for therapy outcome prediction (described further below).

4.2.4. Deep Learning and Visualization Framework

4.2.4.1. System Overview

Figure 4.1 demonstrates an overview of the deep learning framework developed and investigated for LC/LF outcome prediction using the baseline MRI data and clinical features. At the heart of the proposed system, an InceptionResNetV2 is trained to classify the local response of lesions using associated single slices of the T1w and T2-FLAIR images as two parallel input channels of the network. Using the trained network, for each MRI slice of the lesion, 256 features are extracted from the last fully-connected layer of the network (Figure 4.1(a)). The clinical features are then fused with the extracted features from the 2D MRI slices through concatenation and fed to either an LSTM (Figure 4.1(b)), Seq2Seq (Figure 4.1(c)), or transformer network (Figure 4.1(d)) to incorporate 3D spatial dependencies exist within volumetric MRI in predicting the therapy outcome of each lesion. The clinical features are fused with the deep-learning features of each MRI slice before feeding them to the LSTM, Seq2Seq, or transformer network to let this information propagate through the network during the training phase. The LSTM network consists of two layers with N = 45 LSTM cells in each layer. Each cell is connected to its adjacent cell in the same layer, propagating the current cell state to the next one. The first layer of LSTM provides the input to the second layer and is connected to it in the same fashion that the original input is fed to the LSTM network. At the last cell of the second layer of the network, a dense layer with two output units and a *softmax* activation classifies the outcome of each lesion into either LC or LF.

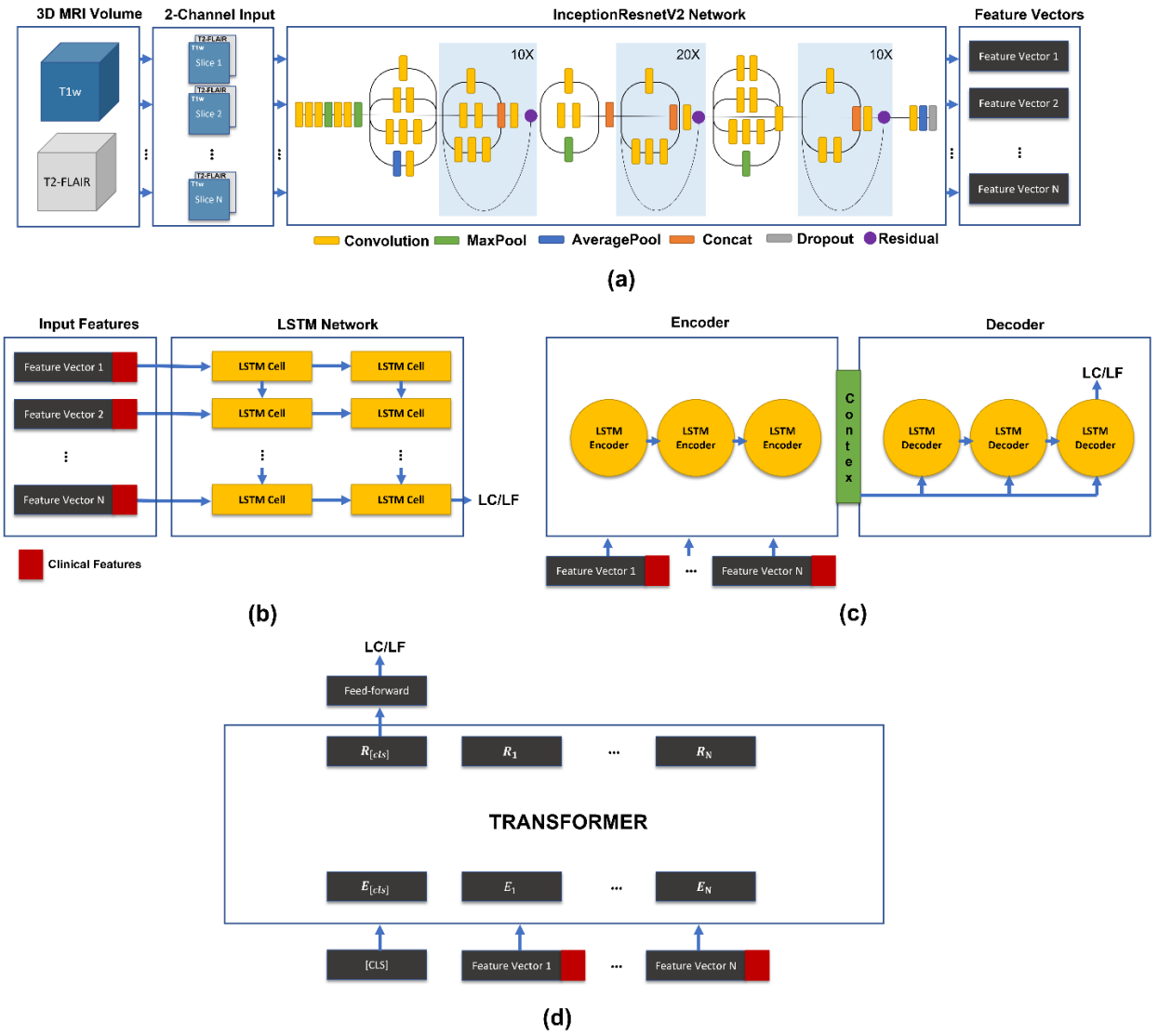

Figure 4.1- System overview. (a) Slice by slice feature extraction from MRI volumes using the InceptionResnetV2, (b) a two-layer LSTM, (c) a Seq2Seq network, and (d) a transformer network. The input feature vectors for all three networks are a concatenation of the features coming from the InceptionResNet and the clinical features.

The Seq2Seq network has an encoder-decoder style LSTM architecture. Each of the encoder and decoder components consists of three LSTM layers (each layer with N = 45 cells). Each LSTM encoder/decoder outputs a cell state and a hidden state based on the previous cell state, previous hidden state, and the current input. A weighted sum of the hidden states in the encoder generates the context vector that is fed to the decoder to set the initial cell state and hidden state of the

decoders. Each LSTM decoder produces a prediction, and the prediction of the last LSTM layer is applied as the overall outcome prediction of the Seq2Seq network for each lesion. The transformer input consists of 45 fixed-length vectors along with a trainable class (*cls)* token with the number of transformer blocks being 1, the number of attention heads being 2, and the hidden layer size of feed-forward network for input classification being 32. Through matrix multiplication, the transformer architecture allows the information to propagate from one slice to another in a way that all slices are considered when making a classification, with emphasis on more important slices. Further details on different components in the proposed framework, and the InceptionResNetV2, LSTM, Seq2Seq, and transformer architectures have been provided in Appendix A.

4.2.4.2. Model Training

The IncpetionResNetV2 network was trained for the task of therapy outcome prediction using all single two-channel images of T1w and T2-FLAIR (128×128 pixel $\times$ 2 channels of T1w and T2-FLAIR) associated with each lesion in the training set. The two-channel images included a standardized ROI encompassing the entire lesion (tumour + edema) and its 5 mm margin on the 2D imaging plane. As mentioned before, this network was developed to be used solely as a feature extractor in the framework. To serve this purpose, the prediction accuracy of the network on the validation set must improve as the ability of the network to predict therapy outcome is closely related to how distinctive the derived features in the last layer of the network are. In order to optimize the prediction accuracy of the IncpetionResNet, the network weights were initialized through pretraining, and the training was performed using a decreased batch size and learning rate through a curriculum learning strategy as described below. The network pretraining was performed using ImageNet, and then the BraTS dataset [48] on the task of brain tumour type classification.

Small batches can offer a regularization effect [151]. To obtain maximum generalization and prevent overfitting, a batch size of one was used during the network training, while to maintain stability due to the high variance of the gradient, the learning rate was set to a very small value of 10^{-6}. While this strategy was successful in improving the network performance on the validation set, it resulted in a longer training process because of the small learning rate and also very small batch size [152]. The idea of curriculum learning is to start the network training on easier subtasks and gradually increase the task difficulty [153]. Since outcome prediction using single 2D images can be extremely difficult in some cases, for a smoother network training, the network was initially trained only on the MRI slices of each lesion with a larger tumour cross-section. This is usually the middle slices of a lesion in the axial plane of an MRI volume. Gradually, more challenging images were added to the training set. Practically, this method of feeding data to the network resulted in smoother training and better validation accuracy and loss.

The LSTM and Seq2Seq networks were trained with a learning rate of 0.0001 and batch size of 2. An *RMSprop* optimizer was used for optimizing the categorical cross-entropy loss function in both networks. A *tanh* was used as the activation function in all layers of the LSTM and Seq2Seq networks. Both networks were trained for a total of 500 epochs. The transformer was trained for 200 epochs with a batch size of 8 and a learning rate of 0.0001. An Adam optimizer was used for optimizing the sparse categorical cross-entropy function. The *RELU* activation function was utilized for hidden layers of the transformer while the activation function for the last feed-forward layer (classifier) was *softmax*.

All the experiments were run in Python and models were trained and tested using Keras [154] with TensorFlow [155] backend. Additionally, we used the scikit-learn package [156] to

calculate performance metrics and the matplotlib package [157] for visualization. All model trainings were performed on a single GeForce RTX TI 2080 graphic card. The training process took about 18 hours (~54m parameters), 2 hours (60k parameters), 3 hours (6m parameters), and 5 hours (7m parameters), for the InceptionResNet, LSTM, Seq2Seq, and the transformer networks respectively. The total inference time for a single input was 27ms, 29ms, and 31ms for the LSTM, Seq2Seq, and transformer networks.

4.2.5. Visualization of Network Decision Basis

The outcome prediction framework was supplemented by a visualization algorithm to illustrate the contribution of different areas of ROI on MRI to the network's prediction for each lesion. Specifically, the visualization algorithm generates a heatmap that color-codes the importance of different regions on the input images for the network's conclusion and can be used to interpret the rationale behind its decision for each case. A modified version of the prediction difference analysis (PDA) technique was adopted in the applied visualization method in conjunction with a sliding window analysis [64]. A 2×2 pixel sliding black square was used to occlude a small region of the input image iteratively before feeding it to the trained network for outcome prediction. The absolute difference in the output probability of the network (*i.e.*, $|p_{input} - p_{occluded_input}|$) was recorded in each iteration as a metric to measure the contribution of the occluded region and applied to generate the heatmap. A higher difference between the obtained probabilities typically demonstrates more important regions of the image with more-telling information for and a higher impact on the network's prediction.

4.3. Results

The clinical characteristics of the patients in this study have been summarized in Table 4.1. Among the 124 patients, 40% were male and 60% were female. The patients had an average age of 62 ± 15 years and an average tumour size of 2 ± 1.03 cm. The average GPA for the patients was 2.2.

The exhaustive search process to select the best clinical features for outcome prediction resulted in a set of four features including the histology, tumour location, tumour size, and number of brain metastases. Using these features, the optimized MLP network could predict the therapy outcome with an accuracy of 68%, a sensitivity of 65%, and a specificity of 70% on the independent test set. These results set the ground to investigate in the next step whether integrating clinical variables with the deep learning features extracted from MRI can improve the accuracy of radiotherapy outcome prediction. Table 4.2 summarizes the performance of the deep learning networks on the validation and test set before and after integrating the clinical features. The table presents the results in terms of accuracy, sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve (AUC). Since the InceptionResNet processes MRI slices individually, the output probability of the network was averaged over all slices associated with the entire lesion volume before thresholding to obtain the overall prediction.

The results presented in Table 4.2 suggest that using the LSTM, Seq2Seq, or transformer network with the InceptionResNet to incorporate inter-slice dependencies outperformed a simple averaging over slices. Specifically, coupling the Seq2Seq, transformer, and LSTM with the InceptionResNet improved the sensitivity of the predictive model from 65% to 77% and its specificity from 74% to 78%, 78%, and 83%, respectively, on the independent test set. Further,

integrating the clinical variables and deep learning features of MRI in the framework improved the performance of the predictive model.

Table 4.1- Patient Characteristics

Clinical Features / Outcome	**Training Set (99 Patients and 116 lesions)**	**Test Set (25 patients and 40 lesions)**
Tumour Size (Longest Diameter)	Range: 0.4 – 7 cm Mean: 1.99 cm	Range: 0.6 – 6.6 cm Mean: 2.06 cm
Age	62 ± 15 years	63 ± 17 years
Gender		
Male	39 patients (39%)	11 patients (44%)
Female	60 patients (61%)	14 patients (56%)
Tumour Location		
Supratentorium	87 lesions (75%)	29 lesions (72.5%)
Infratentorium	29 lesions (25%)	11 lesions (27.5%)
Histology		
Lung cancer	58 lesions (50%)	23 lesions (57.5%)
Breast cancer	26 lesions (22%)	9 lesions (22.5%)
Melanoma cancer	9 lesions (8%)	3 lesions (7.5%)
Colorectal cancer	7 lesions (6%)	0 lesions (0%)
RCC cancer	8 lesions (7%)	1 lesion (2.5%)
Other	8 lesions (7%)	4 lesions (10%)
Total Dose (Over 5 Fractions)		
22.5 Gy	1 lesion (1%)	0 lesions (0%)
25 Gy	20 lesions (17%)	8 lesions (20%)
27.5 Gy	6 lesions (5%)	2 lesions (5%)
30 Gy	73 lesions (63%)	20 lesions (50%)
32.5 Gy	7 lesions (6%)	6 lesions (15%)
35 Gy	9 lesions (8%)	4 lesions (10%)
Previous WBRT		
Yes	45 lesions (39%)	9 lesions (22.5%)
No	71 lesions (61%)	31 lesions (77.5%)
Prior SRT/SRS		
Yes	1 lesion (1%)	0 lesions (0%)
No	115 lesions (99%)	40 lesions (100%)
Number of Brain Metastases		
One lesion	34 patients (34%)	9 patients (36%)
Two lesions	35 patients (35%)	7 patients (28%)
Three lesions	11 patients (11%)	4 patients (16%)
More than three lesions	19 patients (19%)	5 patients (20%)
Graded Prognostic Assessment (GPA)		
0.00 –1.00	15 patients (15%)	3 patients (12%)
1.01–2.00	39 patients (39%)	14 patients (56%)
2.01–3.00	36 patients (36%)	3 patients (12%)
3.01– 4.00	9 patients (9%)	5 patients (20%)
SRT Outcome		
LC	70 lesions (60%)	23 lesions (57.5%)
LF	46 lesions (40%)	17 lesions (42.5%)

Table 4.2- Results of therapy outcome prediction in terms of accuracy, sensitivity, specificity, and AUC for different models

Network	Validation Set				Independent Test Set			
	Acc.	**Sens.**	**Spec.**	**AUC**	**Acc.**	**Sens.**	**Spec.**	**AUC**
MLP / Clinical Features	60%	50%	66.7%	0.65	67.5%	65%	70%	0.68
InceptionResNet	66.7%	66.7%	66.7%	0.69	70%	65%	74%	0.72
InceptionResNet + Seq2Seq	73.3%	66.7%	77.8%	0.76	77.5%	76.5%	78.2%	0.81
InceptionResNet + Transformer	73.3%	66.7%	77.8%	0.75	77.5%	76.5%	78.2%	0.81
InceptionResNet + LSTM	80%	83.3%	77.8%	0.83	80%	76.5%	82.6%	0.83
InceptionResNet + Seq2Seq with Clinical Feature Fusion	80%	83.3%	77.8%	0.81	77.5%	70.6%	82.6%	0.8
InceptionResNet + Transformer with Clinical Feature Fusion	80%	83.3%	77.8%	0.82	77.5%	70.6%	82.6%	0.82
InceptionResNet + LSTM with Clinical Feature Fusion	**86.7%**	83.3%	**88.9%**	**0.88**	**82.5%**	76.5%	**87%**	**0.86**

Whereas the predictive models with only the clinical variables and with the MRI deep learning features solely demonstrated an outcome prediction accuracy of 68% and 78-80%, respectively, the models with the combined deep learning and clinical features demonstrated an accuracy of 78-83% on the test set. The Seq2Seq and transformer networks could not benefit from incorporating the clinical features that can be due to their overcomplex architecture for this application. The results of the ROC analysis (Figure 4.2) also support these observations and demonstrate the benefits of incorporating the 3D spatial dependencies in MRI deep learning features and fusing the clinical information to improve the performance of the outcome prediction system. The best results were obtained by the InceptionResNet + LSTM model with clinical feature fusion where the predictive model demonstrated an AUC of 0.86 on the independent test set, compared to an AUC of 0.72 and 0.8 for the InceptionResNet, and InceptionResNet + LSTM without the clinical

features, respectively. Table 4.3 summarizes experimental results of the proposed framework with different hyperparameters on validation set.

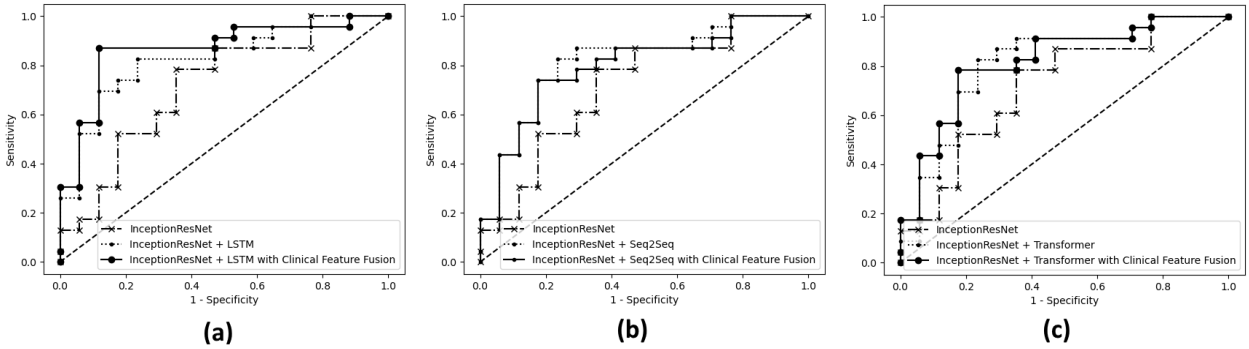

Figure 4.2- The ROC curves on the independent test set for different predictive models.

Table 4.3- Experimental results of the proposed framework on Validation Set. For the input of LSTM, Seq2Seq, and Transformer networks, the extracted features from the best InceptionResNet were concatenated with clinical features and then fed to these networks.

Model	Pre-train (ImageNet)	Pretrain (ImageNet + BraTS)	Learning Rate	Batch Size	Epochs	Validation Accuracy	Validation Sensitivity	Validation Specificity
InceptionResNet	✓	✓	**1e-6**	**1**	**500**	**66.7%**	**66.7%**	**66.7%**
InceptionResNet	✓	×	1e-6	1	500	60%	66.7%	56%
InceptionResNet	×	×	1e-6	1	500	53.4%	50%	56%
InceptionResNet	✓	✓	1e-3	1	500	46.7%	50%	44.4%
InceptionResNet	✓	✓	1e-3	32	500	53.4%	50%	56%
InceptionResNet	✓	✓	1e-3	64	500	53.4%	50%	56%
InceptionResNet	✓	✓	1e-6	32	500	60%	66.7%	56%
InceptionResNet	✓	✓	1e-6	1	200	53.4%	50%	56%
LSTM	×	×	**1e-4**	**2**	**500**	**86.7%**	**83.3%**	**88.9%**
LSTM	×	×	1e-4	16	500	73.3%	66.7%	77.8%
LSTM	×	×	1e-2	2	500	73.3%	83.3%	66.7%
LSTM	×	×	1e-4	2	100	80%	66.7%	88.9%
Seq2Seq	×	×	**1e-4**	**2**	**500**	**80%**	**83.3%**	**77.8%**
Seq2Seq	×	×	1e-4	16	500	73.3%	83.3%	66.7%
Seq2Seq	×	×	1e-2	2	500	73.3%	66.7%	77.8%
Seq2Seq	×	×	1e-4	2	100	53.3%	33.3%	66.7%
Transformer	×	×	**1e-4**	**8**	**200**	**80%**	**83.3%**	**77.8%**
Transformer	×	×	1e-4	16	200	80%	66.67%	88.9%
Transformer	×	×	1e-2	8	200	73.3%	83.3%	66.7%
Transformer	×	×	1e-4	8	100	80%	66.67%	88.9%

Figure 4.3 demonstrates the results of survival analysis on the two cohorts of patients in the test set stratified based on the outcome prediction at the baseline using different models. The Kaplan-Meier progression-free survival curves are presented for the patients in cohort 1 (with their all lesions having a predicted outcome of LC) and cohort 2 (with at least one lesion with a predicted

outcome of LF). A log-rank test applied on the survival curves of the two cohorts for each model demonstrated no statistically significant difference for the InceptionResNet or the clinical model, an approaching significance (p = 0.05) for the InceptionResNet + LSTM model, and a significant difference for the InceptionResNet + LSTM model with clinical feature fusion.

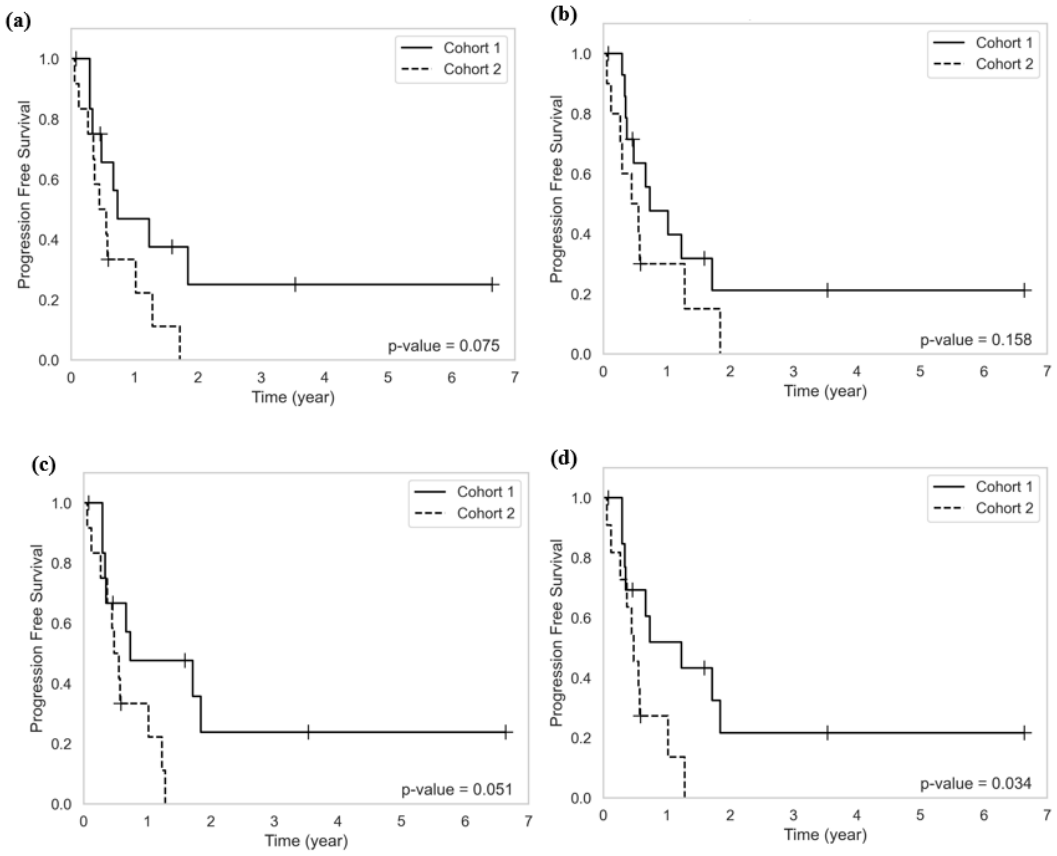

Figure 4.3- Kaplan-Meier progression-free survival curves for two cohorts of patients stratified at the baseline based on the outcome prediction by different models: (a) MLP with clinical features only, (b) InceptionResNet, (c) InceptionResNet + LSTM, and (d) InceptionResNet + LSTM with clinical feature fusion. Cohort 2 includes the patients in the independent test set who had at least one lesion with a predicted outcome of local failure, and cohort 1 includes all other patients in the independent test set.

Figure 4.4 depicts the visualization heatmaps associated with the T1w and T2-FLAIR images of four representative lesions obtained through the prediction difference analysis. The heatmaps demonstrate the impact level of different regions on MRI on the decision of the InceptionResNet in predicting the LC/LF outcome for each lesion. The visualization results imply

that the tumour/lesion margin areas are particularly among the high-impact regions on both T1w and T2-FLAIR images with higher attention gained from the network for outcome prediction.

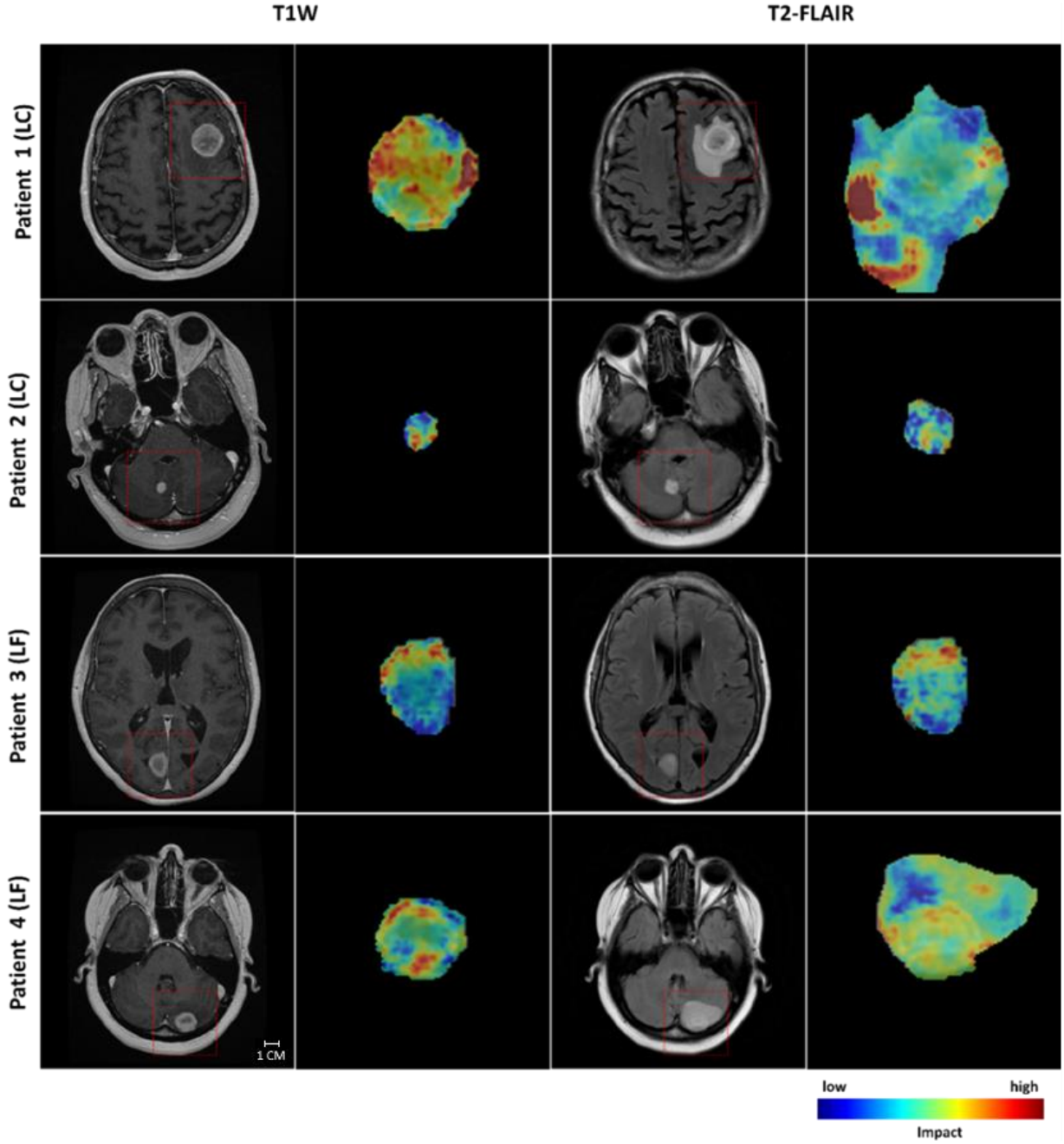

Figure 4.4- Treatment-Planning MRI and deep learning visualization heatmaps acquired for four representative patients with metastatic lesions treated with SRT, two with an LC and the other two with an LF outcome. The heatmaps are overlaid on the ROIs input to the InceptionResNet and demonstrate the impact level of different areas on each image for the network's prediction, as calculated through a modified PDA. The network predicted the outcome of all four lesions correctly.

4.4. Discussion and Conclusion

In this chapter, the possibility of early prediction of local outcome for brain metastasis patients treated with SRT was investigated using deep learning of the treatment-planning MRI and clinical variables. A comprehensive outcome prediction framework was developed to derive optimal deep learning-based MRI features, integrate them with standard clinical variables, and incorporate 3D spatial dependencies within volumetric MRI of each lesion to predict the therapy outcome. To extract features from MRI, an InceptionResNet architecture was adapted, mainly because this architecture proved to excel at learning distinctive features from images in many tasks by applying residual connections and Inception blocks. To incorporate spatial dependency between slices in each MRI volume, three different neural network architectures, namely, LSTM, Seq2Seq, and transformer networks were utilized and investigated. The results demonstrated a notable improvement in prediction performance of the model after integrating the recurrent or attention-based neural networks and clinical features. While using only the clinical variables or MRI features from single slices resulted in an AUC of 0.68 and 0.72, respectively, coupling the LSTM, Seq2Seq, and transformer network with the InceptionResNet improved the AUC to 0.83, 0.81, and 0.81, respectively. The clinical feature fusion with the MRI deep learning features improved the results further, with the best result of 83%, 77%, 87%, and 0.86 for the accuracy, sensitivity, specificity, and AUC obtained for the InceptionResNet + LSTM model with clinical feature fusion.

Results of risk stratification through survival analysis further highlighted the benefits of incorporating the 3D spatial dependencies of MRI features as well as the clinical feature fusion. Among different models explored, only the InceptionResNet + LSTM model with clinical feature fusion could stratify the patients into two cohorts with a statistically significant difference in

progression-free survival, with the cohorts identified by the InceptionResNet + LSTM model without any clinical features approached significance.

A visualization module was integrated with the outcome prediction framework to provide insights on the contribution level of different areas of lesion on MRI to the network's decisions. The results highlighted the importance of the tumour/lesion margin areas in radiotherapy outcome prediction. This observation is in agreement with the findings of previous studies on MRI radiomics for treatment outcome prediction in brain metastasis [51], [138]. Nests of tumour cells may exist for several millimeters outside the confines of the distinct metastatic brain lesion [158]. The information provided through the visualization modules, such as the one presented in this work, regarding the contribution of margin areas of a lesion to its predicted outcome can potentially be beneficial during radiation treatment planning to reduce the chance of local failure in brain metastasis [158].

The findings of this study show the possibility of early prediction of therapy outcome using a combination of quantitative MRI features and clinical attributes. This is in agreement with observations of previous studies that suggest the diagnostic and prognostic power of hand-crafted textural features in various imaging modalities such as CT [159], MRI [160], and Ultrasound [161] in different cancer sites. The study here highlights the advantage of using deep learning architectures in combination with recurrent neural networks for automated extraction of optimal quantitative features from volumetric MRI that can be effectively coupled with standard clinical variables for accurate radiotherapy outcome prediction. In conclusion, the promising results obtained in this study encourage future investigations on larger cohorts of patients. The results reported in this study were obtained on an independent test set. However, to evaluate the efficacy

and robustness of the framework in the clinic more rigorously, further investigations are required preferably on multi-institutional data.

Chapter 5

A Self-Attention-Guided 3D Deep Residual Network with Big Transfer to Predict Local Failure in Brain Metastasis after Radiotherapy using Multi-Channel MRI*

5.1. Introduction

A considerable portion of patients with extracranial malignancies develop brain metastases [3]. Because of increasing access to neuroimaging and developments in systemic therapies for patients with metastatic disease, as well as increased physician and patient awareness of brain metastasis, its incidence is expected to increase among cancer patients [3]. The survival of patients with brain metastasis depends on timely diagnosis and effective therapy. The major therapeutic options for metastatic brain tumours include surgery, radiation therapy (RT), and/or chemotherapy.

Because of various tumour and/or patient-related characteristics such as tumour size, location, and histology as well as the patient's genetics, age and performance status, local response of brain metastasis tumours to radiation varies among patients. This is true even when standardized dose/fraction regimens are administrated [29]. The local response to RT is classified as either local control (LC; stable or shrinking tumour that is indicative of a stable disease, partial response, or

*A version of the material presented in this chapter has been published in IEEE Journal of Translational Engineering in Health and Medicine, 11: 13-22, 2023. DOI: 10.1109/JTEHM.2022.3219625

complete response,) or local failure (LF; enlarging tumour associated with a progressive disease) based on tumour size changes on follow-up structural serial imaging [28]. However, it could take months for a local response to be visible on follow-up scans. Given that the median survival of patients with brain metastasis following RT can be between 5 months and 4 years [7], [34], early detection of LF after RT potentially permit effective adjustments in treatment that lead to enhanced therapy outcomes, patients' survival, and their quality of life.

Following the successful application of artificial intelligence (AI) methods in diagnostic imaging [162], [163], AI-based cancer imaging analysis is now being used to meet other and more complex clinical challenges [95], [164]. These methodologies have the capacity to uncover previously unknown features from routinely acquired medical images. Quantitative and semi-quantitative features, which are often beyond human perception, can be derived from obtained neuroimaging data. These features can potentially be applied to develop machine learning models to address crucial clinical challenges such as therapy outcome assessment or treatment response prediction. Radiomics is a relatively new transformational research domain that adapts high-throughput approaches for mining of large-scale medical imaging datasets to identify quantitative features (biomarkers) for different diagnostic and prognostic applications [44]. Multiple studies have shown links between radiomic signatures of tumours and their phenotypic, genomic, and proteomic profiles [75]. Several studies have also demonstrated the efficacy of radiomic-based machine learning models in therapy outcome prediction [76], [134], including local response of brain metastasis to radiotherapy [142], [148], [165].

Compared to hand-crafted radiomic features, the application of deep learning in medical imaging could possibly address more complicated challenges, particularly when large relevant

datasets are available. Deep learning models have shown great promise in recognizing important and distinctive aspects of medical image data in various applications including cancer therapeutics [166]–[168]. Deep models, and especially convolutional neural networks (CNNs), can detect complex textural patterns in tissue, distinguish between malignant and benign cells, and possibly derive information from tumour images for therapy outcome prediction [55], [86], [87]. Accordingly, the CNNs can potentially outperform the traditional radiomic models in diagnostic and prognostic applications for precision oncology by detecting patterns in medical images that are not captured by closed-form mathematical definitions of hand-crafted radiomic features [53], [89], [169]. A recent publication from our group shows that the deep-learning features derived from 2D MRI slices outperform the standard clinical variables in predicting radiotherapy outcome in brain metastasis [170].

Attention mechanisms in deep learning were introduced in the field of computer vision with the goal of imitating the human visual system's ability to naturally and effectively discover prominent regions in complex scenes [171]. An attention mechanism in a vision system can be thought of as a dynamic selection process that is implemented by adaptively weighing features based on the relevance of the input. Over the past few years, attention mechanism has played an increasingly important role in different computer tasks, including image classification [172], object detection [173], semantic segmentation [174], and 3D vision [175]. The attention mechanism has shown promise in medical imaging analysis, especially when the problem is not as straightforward as generic image classification, where the well-defined object of interest is usually in the image center [176]. A number of previous studies have applied attention mechanisms to provide more powerful architectures capable of catching subtle features covered in medical images. Guan et al.

[177] proposed a three-branch attention guided convolution neural network (AG-CNN) which learns from disease-specific regions through a local branch to reduce noise and improve alignment, with a global branch to compensate for the lost discriminative cues by the local branch. Using a fusion branch to combine the local and global cues, their model could achieve a new state-of-the-art performance in classifying images of the ChestX-ray14 dataset [177]. Rao et al. [178] conducted an experimental research to investigate the contribution of various attention mechanisms including squeeze-and-excitation (SE) [172], global context (CG) [179], and convolutional block attention module (CBAM) [180] to the performance of deep classification models for different imaging modalities including x-ray, MRI, and CT. The experimental results show that the attention mechanisms enable standard CNN models to focus more on semantically important and relevant content within features, with improved area under the receiver operating characteristic (ROC) curve (AUC) for all classification models investigated [178]. Furthermore, the CBAM outperformed the other two attention mechanisms in several experiments on different imaging datasets. Shaik et al. [181] proposed a multi-level attention mechanism for the task of brain tumour classification. The proposed multi-level attention network (MANet) combines spatial and cross-channel attention, focusing on tumour region prioritization while also preserving cross-channel temporal connections found in the Xception backbone's semantic feature sequence [181]. They benchmarked their framework on BraTS [182] and Figshare [183] datasets where their model outperformed several models proposed previously for the brain tumour classification task [181].

This chapter introduces an innovative transformer-convolutional deep learning model for predicting the LC/LF outcome in brain metastasis treated with SRT using two-channel MRI acquired at pre-treatment. A novel attention-guided 3D residual network architecture was

developed with embedded self-attention modules [184]–[186] and compared with another residual network with 3D CBAM as the attention mechanism. A training recipe was adapted for the therapy outcome prediction models during pretraining and training the down-stream task based on the recently-proposed big transfer (BiT) principles [187]. A new 3D visualization method has been introduced to illustrate the impact of different regions throughout the lesion volume upon the network's prediction of the therapy outcome. The results demonstrate that incorporating the attention mechanisms into the vanilla 3D residual network improves its performance in outcome prediction considerably, with the self-attention mechanism outperforming the CBAM in terms of accuracy, AUC, and F1-score. Further, the adapted BiT-based recipe for pretraining and hyperparameter tuning improves the deep models' performance in therapy outcome prediction.

5.2. Methods and Procedures

5.2.1. Data Acquisition

This study was carried out in compliance with the institutional research ethics board approval from Sunnybrook Health Sciences Centre (SHSC), Toronto, Canada. Data were obtained from 124 patients with brain metastasis treated with hypo-fractionated SRT (5 fractions). In this study, the baseline treatment-planning MRI including contrast-enhanced T1-weighted (T1w), and T2-weighted-fluid-attenuation-inversion-recovery (T2-FLAIR) images were applied for therapy outcome prediction. The MRI scans were acquired using a 1.5 T Ingenia system (Philips Healthcare, Best, Netherlands) and a 1.5 T Signa HDxt system (GE Healthcare, Milwaukee, WI, USA). The T1w and T2-FLAIR images had an in-plane image resolution of 0.5 mm and a slice thickness of 1.5 mm and 5 mm, respectively. The treatment-planning tumour contours delineated by expert oncologists as well as the edema contours outlined under their supervision were also

included in the dataset. The dataset (124 patients with 156 lesions) was randomly partitioned at patient level into a training set (99 patients with 116 lesions) that was used for model development and optimization, and an unseen test set (25 patients with 40 lesions) that was applied for independent evaluation of the models. From the training set, 10 patients with 15 lesions were randomly selected as the validation set for optimizing the model hyperparameters.

The patients were scanned with MRI after SRT on a two to three-month follow-up schedule. A radiation oncologist and a neuroradiologist determined the local response for each lesion separately after monitoring it on serial MRI using the RANO-BM [28] criteria. The outcome (LC or LF) was determined for each lesion in the last patient follow-up. Serial imaging (including perfusion MRI) and/or histological confirmation were used to diagnose adverse radiation effect (ARE) and distinguish it from progressive disease [92], in accordance with the report by Sneed et al. [91]. Following these criteria, a total of 93 lesions were categorized as LC while 63 lesions were labeled as LF.

5.2.2. Preprocessing

All MR images were resampled to a size of 512 × 512 × 174 voxels (voxel size: 0.5 × 0.5 × 1 mm^3). An affine registration method was used to co-register the T1w and T2-FLAIR images. Skull stripping was performed on all MR images. The voxel intensities in each skull-stripped MRI volume were normalized between 0 and 1. To ensure a lesion-level local outcome prediction the size of the smallest sub-volume enclosing the tumour and edema (lesion) and their 5-mm outer margin [94], was identified for all lesions. A sub-volume of 128 × 128 × 83 voxels was determined as a fit standard to encompass the entire region of interest (ROI) described above for all individual lesions. The standardized sub-volumes were then cropped from the T1w and T2-FLAIR images

and concatenated for each lesion as two channels of data, generating the input to the neural networks with a size of $128 \times 128 \times 83 \times 2$ voxels. The ROI masks (tumour + 5-mm margin for T1w; tumour + edema + 5-mm margin for T2-FLAIR) were generated using the tumour and edema contours and applied to mask out the areas outside the ROI for each lesion.

5.2.3. Network Overview

The backbone of the proposed network architecture is a vanilla 3D extension of deep residual networks (ResNets), first introduced by He et al. [188], [189]. Instead of learning unreferenced functions, ResNets learn residual functions with reference to the layer inputs. Also, rather than expecting each few stacked layers directly fit a desired underlying mapping, ResNets let these layers fit a residual mapping. Formally, instead of directly mapping the desired underlying function $H(x)$, the stacked nonlinear layers fit another mapping of $F(x) \coloneqq H(x) - x$. This way, the original mapping recast into $F(x) + x$. Residual connections allow for increased depth while addressing the vanishing gradient problem and are also easier to optimize [188].

Our vanilla 3D residual network (Figure 5.1(a)) is inspired by the architecture of ResNet-18, but instead of 2D convolution layers, our network employs 3D convolution with kernel size of $7 \times 7 \times 7$ and 3D pooling layers to handle the 3D nature of MRI volume.

To improve the performance of our 3D residual network in processing multi-channel MRI volumes, we explored incorporating two different attention mechanisms into the architecture, CBAM, and self-attention [185]. CBAM is a simple yet effective attention module which infers attention maps along two separate dimensions (channel and spatial) sequentially [180]. The attention maps are then multiplied by the feature tensors to produce the refined feature tensors.

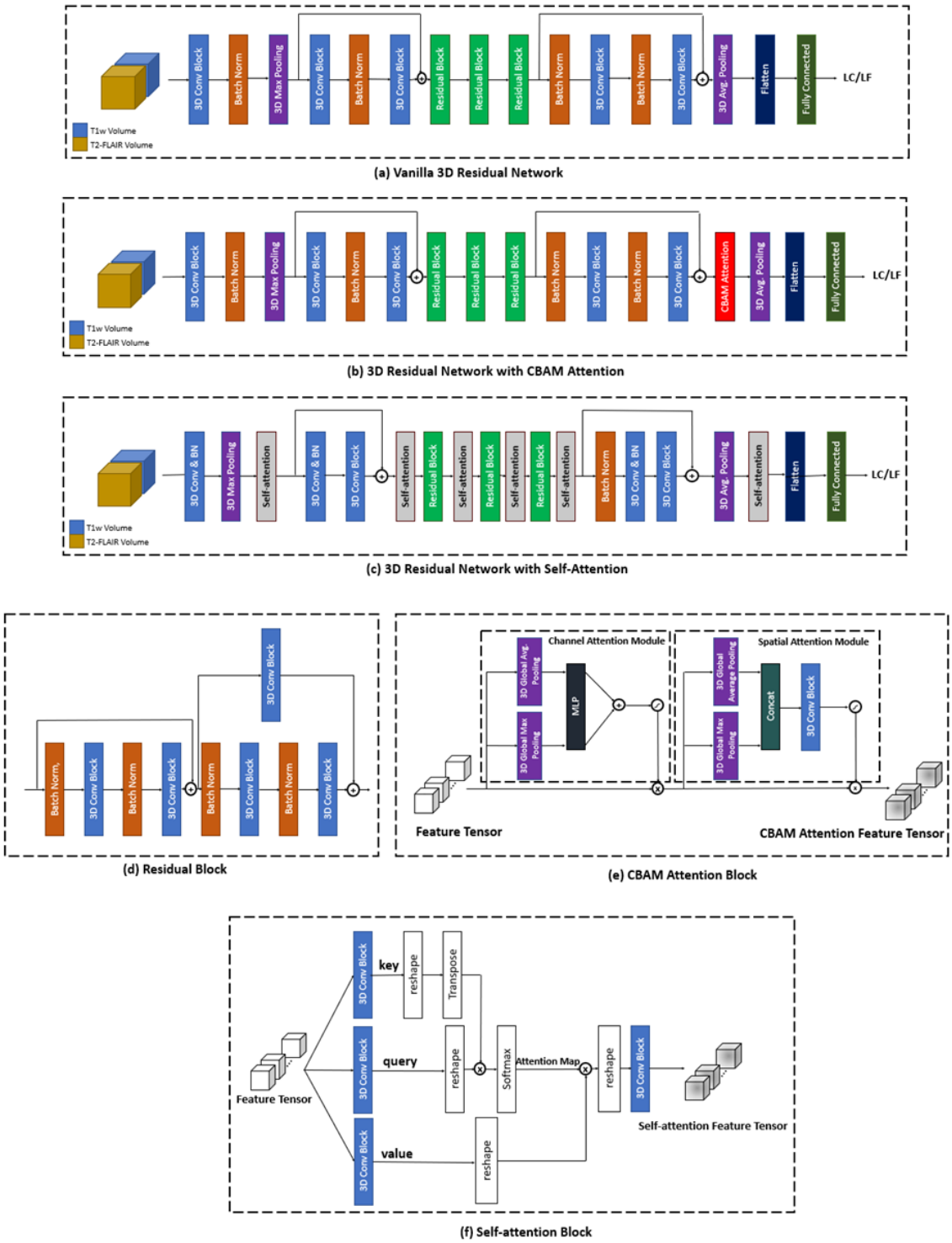

Figure 5.1- Architecture of (a) vanilla 3D residual network (baseline), (b) the 3D residual network with CBAM attention, (c) the proposed self-attention-guided 3D residual network, (d) the residual block in 3D residual network, consisting of residual connections, (e) the CBAM attention block consisting of the channel and spatial attention modules, and (f) the self-attention block consisting of the key, query, and value tensors that generates the final self-attention feature tensor.

Formally, CBAM has two sequential submodules and, given an input feature tensor $x \in \mathbb{R}^{X\times Y\times Z\times C}$, it sequentially infers 1D channel attention vector $M_c \in \mathbb{R}^C$ and a 3D spatial attention map $M_s \in \mathbb{R}^{X\times Y\times Z}$. The developed 3D residual network with CBAM attention is depicted in Figure 5.1(b). The CBAM attention module (Figure 5.1(e)) is embedded right before the average pooling and fully-connected layers to refine features before classification. The refined features are then flattened and fed to the fully-connected layer for classification.

Additionally, we introduced a novel transformer-convolutional network architecture by incorporating self- attention modules into the 3D residual network (Figure 5.1(c)). The convolution operator in CNNs only conducts local operations and has a local receptive field, but the self-attention mechanism can perform non-local operations and capture long-range dependencies and global information within the input images [190]. The self-attention method is based on the covariance between the elements of feature tensors [191]. Formally, a self-attention function can be described through mapping the input feature tensor to a query, a key, and a value tensor. The tensor mappings are performed using 3D 1 × 1 × 1 convolutions. Each element of the output self-attention feature tensor is a linear weighted sum of the elements of the value tensor. The query tensor defines which "values" to focus on for the learning process, while the key and value tensors carry the transformed features extracted from MRI volume. Given that key is $k(x) = W_k x$, query is $q(x) = W_q x$, and value is $v(x) = W_v x$ where W_k, W_q, W_v are learnable weights of the 1 × 1 × 1 convolution filters and x is the feature tensor from the previous layer, the self-attention map α could be calculated using Equation (5.1).

$$\alpha_{i,j} = \frac{\exp(q(x_i)k(x_j)^T)}{\sum_{i=1}^{n}\exp(q(x_i)k(x_j)^T)} \tag{5.1}$$

Figure 5.1.f shows the architecture of the 3D self-attention block incorporated into the proposed 3D residual network with self-attention. For performing the matrix multiplications in this block, the query, key, and value tensors ($\in \mathbb{R}^{X\times Y\times Z\times C}$) are reshaped into matrices ($\in \mathbb{R}^{XYZ\times C}$) and, at the end, reshaped back into tensors of the initial size. The final 1 × 1 × 1 convolution block ensures that the number of channels of the input and output feature tensors stays the same. The 3D self-attention module facilitates capturing long-range inter/intra slice dependencies, hence is added to the architecture after each residual block to ensure deriving such dependencies along with the convolution layers that mostly capture local features and dependencies. More details on the network architectures have been provided in the Appendix A.

5.2.4. Big Transfer and Training Details

Transfer of pretrained models on the target task improves sample efficiency and simplifies hyperparameter tuning when training deep neural networks [187]. Inspired by the work of Kolesnikov et al. [187], we followed the subsequent scheme for pretraining/training the outcome prediction models:

1. The network was first pretrained on the UCF101 dataset [192] for the task of activity recognition and subsequently on the BraTS dataset [48], [182], [193] for the task of classifying brain tumour types using MRI.
2. During pretraining, all batch normalization [194] layers were replaced with group normalization [195] and weight standardization [196] was used in all convolutional layers.

The combination of group normalization and weight standardization with large batches has a significant impact on transfer learning [197]. Also, due to the requirement to update running statistics, batch normalization is detrimental for the transfer [187].

3. During fine-tuning on the main dataset, we used BiT-HyperRule, a heuristic method for hyperparameter selection based on image resolution and number of datapoints as presented by [187]. The models were trained using the stochastic gradient descent (SGD) optimization algorithm with an initial learning rate of 0.003, momentum of 0.9, batch size of 4, and an early stopping based on the validation loss. Data augmentation was performed using horizontal flipping. During fine-tuning, the learning rate was decayed by a factor of 10 at 40%, 60% and 80% of the training steps.

All experiments were performed in Python. The models were developed and evaluated using Keras [154] with TensorFlow [155] backend. The performance metrices were calculated using scikit-learn package [156]. the matplotlib library was used [157] for visualization. The models were trained using four GeForce RTX TI 2080 graphic cards. The training process took 5 hours (~33M parameters), 6 hours (~33M parameters) and 10 hours (~42M parameters) for 3D residual network, 3D residual network + CBAM attention, and 3D residual network + self-attention respectively. The total inference time for a single input is 6ms, 7ms, and 12ms for 3D residual network, 3D residual network + CBAM attention, and 3D residual network + self-attention respectively.

5.2.5. Visualization of Network Decision Basis

A new 3D visualization algorithm was implemented to accompany the outcome prediction framework and show how different areas within the volumetric region of interest on the input images contribute to the prediction of network for each lesion. The visualization module provides a 3D heatmap color-coding the relevance of distinct peri-/intra-lesion areas on multi-channel volumetric MRI to the decision of network and may be applied to analyze the reasoning behind the predicted outcome for each case. The applied visualization method combined a modified version of the prediction difference analysis (PDA) with a sliding window analysis approach [198]. A 2 × 2 × 1 voxel sliding black cube (1 × 1 × 1 mm) was iteratively applied to block a tiny area of the input image. The occluded input was fed to the trained network to predict the associated therapy outcome. In each iteration, the absolute difference in the network's output probability (*i.e.*, $|p_{input} - p_{occluded_input}|$) was calculated and applied as a measure of contribution of the occluded cube to generate the volumetric heatmap. This method generates a point cloud where each point in the cloud maps to a region within the MRI volume. For 3D visualization of generated heatmap, the heatmap voxels were considered as a point cloud with each point maps to a region within the MRI volume. A surface reconstruction technique was adapted to create a 3D heatmap out of the point cloud on any desired surface within the volumetric ROI. Specifically, the cloud points located on the ROI surface were identified and the normal orientation of the point cloud was calculated at each surface point using a minimum spanning tree with the number of neighbours set to 3 for building the tree [199]. The estimated normal orientations were applied in conjunction with, the Poisson reconstruction technique [200] to build a smooth surface mesh from the point cloud. The Poisson surface reconstruction technique creates a 3D mesh from a dense point cloud

by reducing the difference between the surface normal directions of the reconstructed surface and the 3D points in the point cloud [201]. The following steps summarize the procedure to generate a 3D heatmap of importance:

1. For each point (voxel center) in the MRI volume a vector $[x, y, z, r, g, b]$ is assigned, where x, y, z refer to the position of the point and r, g, b refer to the color of the point which shows its impact (or intensity) according to a pre-defined color-coding scheme (color map).
2. The generated point cloud is then normalized (each x, y, z are normalized between 0 and 1)
3. A desired surface within the MRI volume is specified. Since the impact of all voxels within the MRI volume is estimated and color-coded, it is possible to explore and visualize any desired areas throughout the MRI (or the volumetric ROI), for a comprehensive understanding of how different intra- and peri-lesional regions contribute to the network's decisions in therapy outcome prediction.
4. Since the number of points in the point cloud is limited, in order to improve the quality of the final 3D heatmap, interpolation is performed to generate new random points with the constraint of being on the specified surface.
5. Using a k-nearest neighbor algorithm the 10 closest points to each newly generated point are identified and their average color code is assigned it.
6. The interpolated point cloud is used to calculate and generate surface normal orientations at each point required for surface reconstruction. Calculating normal orientations was done using a minimum spanning tree.

7. Once the normal orientations are calculated, using the Poisson surface reconstruction techniques a smooth surface is generated showing important regions contributing to network decisions.

Figure 5.2 shows the overall procedure for generating a desired 3D surface from the initial point cloud. The proposed 3D visualization framework can assist clinicians to get insight into how the network has reached its decision and help to validate the network's decisions by generating meaningful heatmaps.

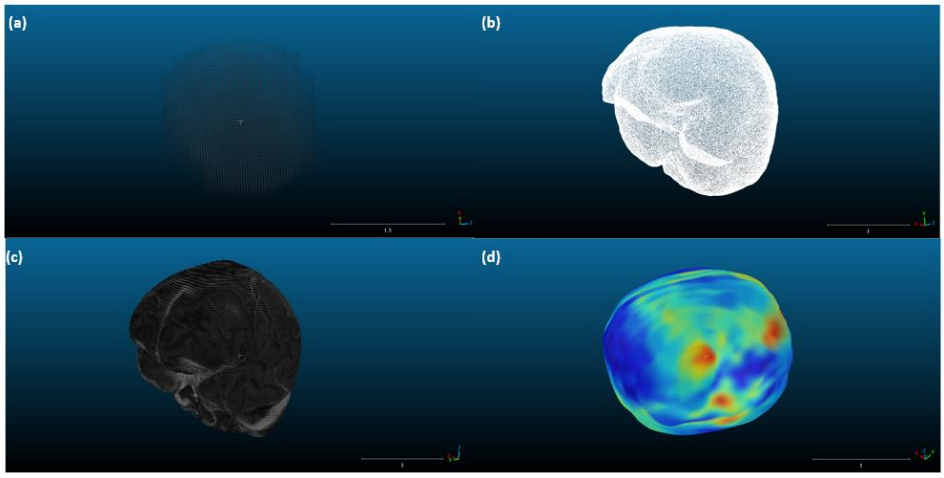

Figure 5.2- The procedure for creating explorable 3D brain models and visualization heatmaps from a set of individual slices. (a) initially, the coordinates (x, y, z) of each voxel center in the MRI volume is determined. Since the number of slices is often limited, the resulting point cloud consists of multiple clusters with the same z and different x and y which is visually undesirable. To mitigate this issue, the points between slices are randomly interpolated, (b) the point cloud after the inter-slice interpolation, (c) the 3D brain model after assigning an intensity to each point in the point cloud and surface reconstruction, (d) applying the same procedure to generate a color-coded 3D visualization heatmap of importance for a lesion within the brain.

5.3. Results

The patients (average age: 62 ± 15 years; 40% male and 60% female) had an average tumour size of 2 ± 1.03 cm and an average GPA of 2.2. The demographic and clinical attributes of the patients in this study are presented in Supplementary Table 4.1.

Figure 5.3 shows the training loss over 300 epochs for the models in this study before and after applying the BiT training scheme. The vanilla 3D residual network pretrained on the UFC101 and BraTS datasets (without BiT scheme) is the baseline model of this study. In pretraining/training of the models without the BiT recipe, the batch normalization layers were not replaced with group normalization and the weight standardization and BiT-HyperRule were not applied. Table 5.1 presents the performance of different models investigated in this study for radiotherapy outcome prediction. A careful investigation of Figure 5.3 and Table 5.1 demonstrates that incorporating the BiT scheme in development of the deep models for outcome prediction generally improves their performance in terms of convergence, loss, F1-score, and AUC on the independent test set. The F1-score and AUC may be considered the most important metrics presented in Table 5.1 because of the imbalance exists in the dataset.

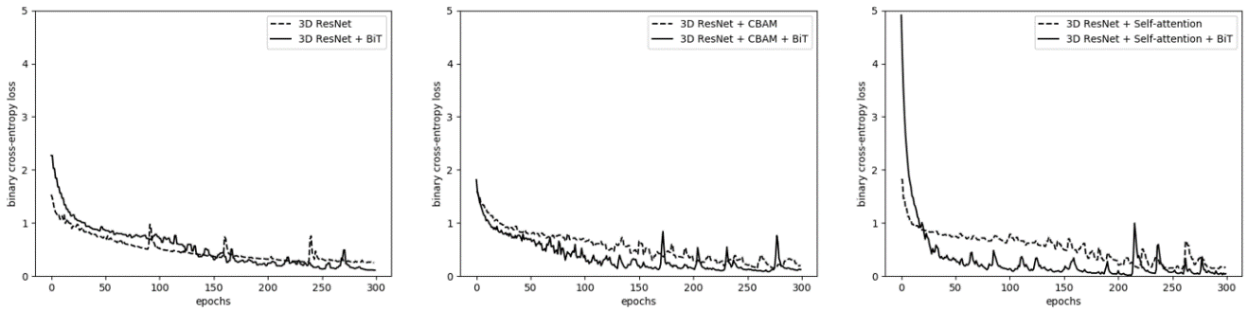

Figure 5.3- Training loss of the three models investigated in this study before and after applying BiT training scheme. Following the BiT training recipe generally led to faster convergence, smaller loss, and better performance overall.

Specifically, following the BiT training scheme, the models improved their AUC on the test set from 0.83 to 0.84, 0.87 to 0.88, and 0.88 to 0.91 for the 3D residual network, 3D residual network + CBAM attention, and 3D residual network + self-attention, respectively. From a different perspective, incorporating attention mechanisms also improved the model performances in terms of accuracy, AUC, and F1-score. While the vanilla 3D residual network could achieve an

F1-score of 75% on the test set, the 3D residual network + CBAM attention improved the F1-score by 2.8%. The F1-score was improved by 3.8% compared to the baseline model by including the self-attention mechanism in the 3D residual network. In particular, the proposed transformer-convolutional network architecture with BiT training demonstrated the best performance in terms of accuracy, AUC, and F1-score, with 8% and 5% improvements in AUC and F1-score, respectively, compared to the baseline model, on the independent test set. This is a considerable improvement in model's performance given the complicated task at hand. Further, the proposed model resulted in the most balanced sensitivity and specificity values compared to the other models despite the imbalanced dataset applied in the study. The effect of incorporating selected clinical features from Section 4.2.3 as extra channel inputs to the framework was also investigated. The results show little to no benefit from adding the clinical features to the models. This can be due to over complexing the input of 3D networks by adding the clinical features which lead to high input dimension, while the added features do not substantially complement those extracted by the network from MRI for therapy outcome prediction. It is noteworthy that the self-attention guided model with BiT training presented in this chapter has outperformed the best model presented in Chapter 4 (InceptionResNet + LSTM with clinical feature fusion) in terms of AUS (0.91 versus 0.86) and resulted in a more balanced sensitivity and specificity (82.4%/82.6% versus 76.5%/87%). Figure 5.4 shows the ROC curves for different models investigated in this study. Table 5.2 shows the experimental results of the networks with different pretraining settings and hyper-parameters.

Table 5.1- Results of radiotherapy outcome prediction for different models. Acc: Accuracy; Sens: sensitivity; Spec: specificity.

Network	Validation Set					Independent Test Set				
	Acc.	**Sens.**	**Spec.**	**AUC**	**F1-Score**	**Acc.**	**Sens.**	**Spec.**	**AUC**	**F1-Score**
3D Residual Network	80%	66.7%	88.9%	0.84	72.7%	80%	71%	**87%**	0.83	75%
3D Residual Network + Clinical Features	80%	83.3%	88.9%	0.84	76.9%	80%	71%	**87%**	0.82	75%
3D Residual Network + BiT	80%	83.3%	77.8%	0.86	76.9%	80%	82.4%	78.%	0.84	77.8%
3D Residual Network + CBAM Attention	80%	83.3%	77.8%	0.88	76.9%	80%	82.4%	78.2%	0.87	77.8%
3D Residual Network + CBAM Attention + Clinical Features	80%	66.7%	88.9%	0.86	72.7%	80%	88.2%	73.9%	0.87	78.9%
3D Residual Network + CBAM Attention + BiT	80%	100%	66.7%	0.88	80%	80%	**88.2%**	73.9%	0.88	78.9%
3D Residual Network + Self-attention	**86.7%**	**83.3%**	88.9%	0.89	**83.3%**	**82.5%**	76.5%	**87%**	0.88	78.8%
3D Residual Network + Self-attention + Clinical Features	**86.7%**	66.7%	**100%**	0.88	80%	80%	**88.2%**	73.9%	0.87	78.9%
3D Residual Network + Self-attention + BiT	**86.7%**	**83.3%**	88.9%	**0.93**	**83.3%**	**82.5%**	82.4%	82.6%	**0.91**	**80%**

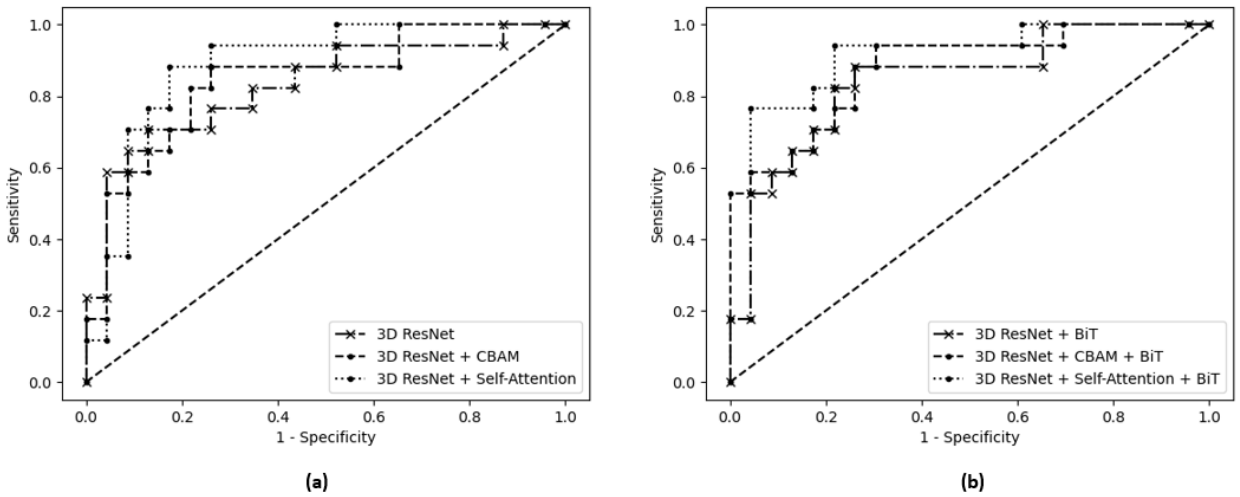

Figure 5.4- The ROC curves for (a) vanilla 3D residual network, 3D residual network + CBAM, and 3D residual network + self-attention, and (b) the same models trained with the BiT scheme.

An explanation of how an attention mechanism helps improving the performance of the deep models in therapy outcome prediction is as follows: characteristics of different regions within tumour and peritumoural areas on MRI contribute unequally to the likelihood of local response. Several studies demonstrate that tumour margin areas on MRI carry invaluable information regarding the responsiveness of brain metastasis to radiotherapy with possibly higher importance

for prediction modeling compared to the core areas [94], [158]. Attention mechanisms help the network to capture the subtle information latent within MRI by filtering out irrelevant data and focusing on regions which truly contribute to the network decisions. Moreover, self-attention facilitates capturing long-range dependencies inside the MRI volume, an important concept that simple 3D convolutional layers are not capable of because of their local nature and limited field of view. Comparing the performance of 3D residual network + CBAM attention and the vanilla 3D residual network, the former has outperformed the latter in terms of AUC and F1-Score, although the number of parameters is almost the same for these networks. This shows the benefit of incorporating attention mechanisms in this setting while it may not increase the network complexity considerably. Our further experiments with the 3D residual network with more parameters (~40M parameters) when extra layers were added to the network resulted in overfitting. This implies that the 3D Residual Network + Self-attention does not simply benefit from the increased number of network parameters but mainly from the structure of the attention layers incorporated. Figure 5.5 demonstrates the 3D visualization heatmaps for two representative lesions generated using the technique introduced in Section 5.2.5. The heatmaps show the contribution level of different regions within the volumetric ROI on the prediction of the proposed attention-guided model for each lesion in terms of local outcome. The 3D heatmaps can aid clinicians to examine the lesion volume thoroughly and inspect impactful regions for a predicted outcome which can eventually support their decision making in assessment, diagnosis, and treatment planning.

Table 5.2- Experimental results of the proposed framework on the validation set for 3D residual network, 3D residual network + CBAM attention, and 3D residual network + self-attention with different hyperparameters and pretraining settings.

Model	Pre-train (UCF101)	Pretrain (BraTS)	Learning Rate	Batch Size	Epochs	Accuracy	Sensitivity	Specificity
3D Residual Network	✓	✓	0.003	4	300	80%	66.7%	88.9%
3D Residual Network + BiT	✓	✓	0.003	4	300	80%	83.3%	77.8%
3D Residual Network + CBAM Attention	✓	✓	0.003	4	300	80%	83.3%	77.8%
3D Residual Network + CBAM Attention + BiT	✓	✓	0.003	4	300	80%	100%	66.7%
3D Residual Network + Self-attention	✓	✓	0.003	4	300	86.7%	83.3%	88.9%
3D Residual Network + Self-attention + BiT	✓	✓	**0.003**	**4**	**300**	**86.7%**	**83.3%**	**88.9%**
3D Residual Network	×	×	0.003	4	300	60%	66.7%	55.6%
3D Residual Network	×	✓	0.003	4	300	73.3%	66.7%	77.8%
3D Residual Network + BiT	×	✓	0.003	4	300	73.3%	83.3%	66.7%
3D Residual Network + Self-attention + BiT	×	✓	0.003	4	300	80%	83.3%	77.8%
3D Residual Network + Self-attention + BiT	✓	✓	0.00001	4	300	73.3%	83.3%	66.7%
3D Residual Network + Self-attention + BiT	✓	✓	0.01	4	300	80%	83.3%	77.8%
3D Residual Network + Self-attention + BiT	✓	✓	0.003	8	300	86.7%	100%	0.78%

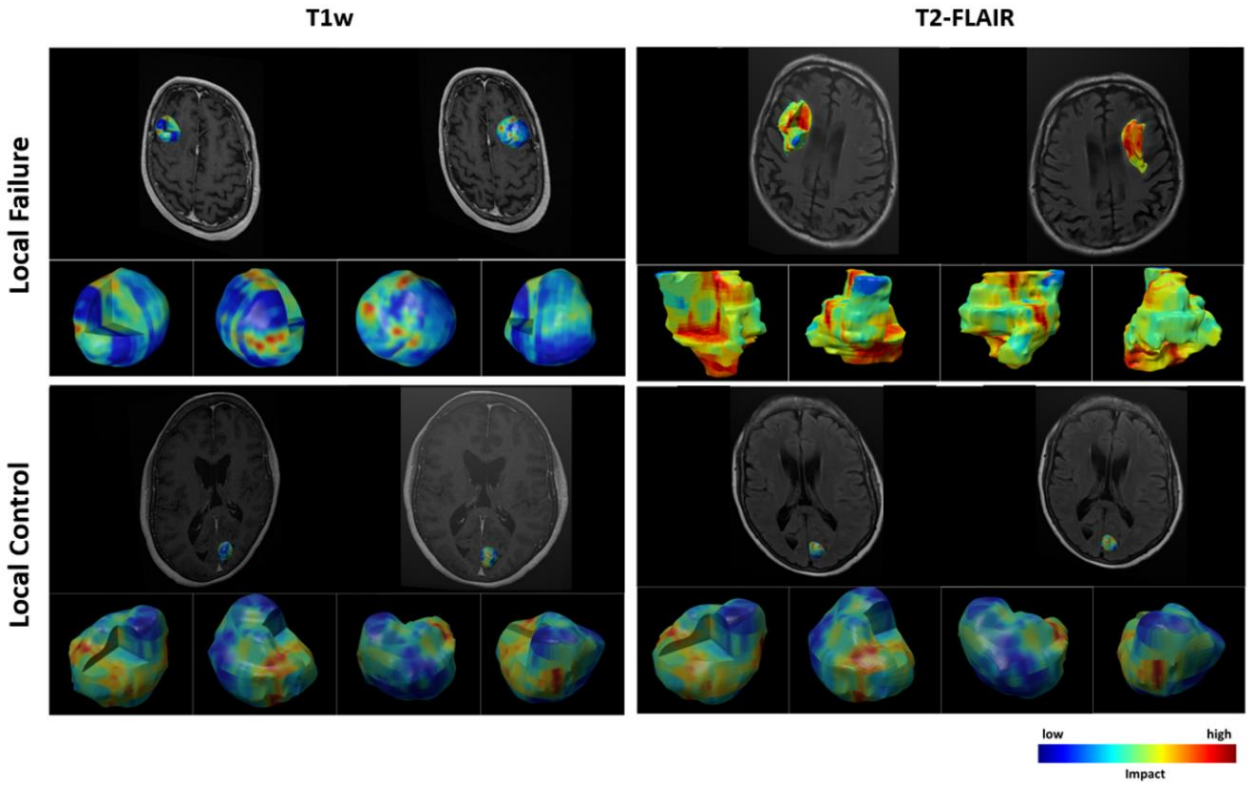

Figure 5.5- 3D visualization heatmaps corresponding to the two input channels (T1w and T2-FLAIR) of the 3D residual network with self-attention and BiT training for two representative lesions, one with an LF (top) and the other one with an LC (bottom) outcome. The user can inspect any desired area on the lesion/margin surface or inside the volumetric ROI and their correspondence with the MRI channels.

5.4. Discussion and Conclusion

An end-to-end 3D convolutional deep learning architecture with self-attention was introduced in this study to predict the local outcome in brain metastasis after radiotherapy. By employing 3D residual blocks in the proposed model, we investigated the possibility of early prediction of LF in brain metastasis treated with SRT using T1w and T2-FLAIR MRI volumes acquired at baseline. We further investigated the effect of incorporating attention mechanisms into the 3D residual network. The results show that the proposed model with self-attention mechanism outperforms the vanilla 3D residual network and the 3D residual network with CBAM attention in terms of accuracy, AUC, and F1-score. The proposed architecture combines residual learning with self-attention mechanism, allowing for full utilization of both global and local information while avoiding information loss. Specifically, the self-attention mechanism in the model takes into account long-range dependencies in the input MRI volumes while the residual connections allow the extracted information to persist throughout the network. We further improved the model's performance by following the BiT scheme for pretraining and hyperparameter tuning. A 3D visualization module was developed and coupled with the framework to show the important areas of lesion on MRI with higher impact on the model's decision. The visualization results confirm the findings of previous studies that the characteristics of tumour/lesion margin areas on T1w and T2-FLAIR images are important for predicting local outcome in brain metastasis treated with radiation therapy. In particular, these regions are among the high-impact regions to the predictions made by the proposed deep learning model with more attention gained from the model for therapy outcome prediction.

The findings of this study demonstrate the feasibility of early prediction of radiotherapy outcome for brain metastasis using only the features extracted from multi-modal MRI volumes. This study highlights the effect of adding attention mechanism to deep networks and the importance of pretraining in transferring knowledge to the fine-tuning step. When dealing with large models and large datasets (which is usually the case during pretraining) adhering to the BiT recipe allows for optimized training during the up-stream task and a computationally inexpensive fine-tuning protocol during the down-stream task to avoid a complex and costly hyper-parameter search. The obtained results are promising and encourage future studies on larger patient populations. The results of this study were obtained using an independent test set that was kept unseen during the model training and optimization. However, for a more rigorous evaluation of the efficacy and robustness of the models in the clinic, further investigations should be performed on larger patient cohorts and preferably with multi-institutional data.

Chapter 6

Conclusion and Future Work

6.1. Conclusion

Given the abundance of MRI and CT data acquired as part of standard of care for brain metastasis patients, machine learning in combination with medical imaging has a good potential to be adopted in the clinic to provide an efficient and non-invasive method of characterizing metastatic brain tumours, assessing tumour response and predicting the outcome of their treatment. Segmenting brain tumours automatically is a difficult challenge since tumours can arise in any location within the brain and can be of any size, number, shape, and contrast, but if managed accurately, could be a substantial aid to the clinicians who often have to go through the laborious task of manual segmentation. Especially in the case of radiotherapy outcome assessment, each patient goes through many follow-up scans and in order to assess the response, all baseline and follow-up scans had to be carefully segmented. As another challenge in the management of patients with brain metastasis, predicting the outcome of prescribed treatments before or early after the therapy is of extreme importance as a key to the personalized cancer therapeutics paradigm. The main contributions of this dissertation project to address the challenges described above are summarized below.

Chapter 2 Introduced an accurate and complex segmentation framework, capable of automatically segmenting brain tumours on the baseline and follow-up MRI with high accuracy. The promising results of the network show the possibility of accurate tumour segmentation on baseline and follow-up MRI. This is a significant contribution considering the huge burden that manual segmentation puts on radiotherapy clinical workflow.

Chapter 2 also introduced an automated system for the assessment of radiotherapy outcome in brain metastasis (local control/local failure and adverse radiation effect) using standard serial MRI based on RANO-BM criteria. By using the complex segmentation framework introduced in the same chapter and through assessing changes in tumour size on serial MRI, the automated system can assess tumour response with an accuracy of 91% in comparison to the response assessment by expert clinicians. Automatic monitoring and evaluation of radiotherapy outcome can streamline the radio-oncology workflow and facilitates precision oncology through fast, regular, and high-throughput response assessment.

Chapter 3 investigated the effect of tumour segmentation imperfections on the performance of MRI radiomic-based radiotherapy outcome prediction models. Despite many automatic segmentation networks introduced in the literature, their integration in quantitative imaging computational models has been hindered due to considerations regarding their imperfection. The results presented in Chapter 3 demonstrate that if the accuracy of a segmentation model is relatively high (a dice score of about 0.9 in our case) but not perfect, there would not be a significant performance difference between the ground-truth and automatic segmentation masks in radiomic modeling. Our results also confirmed that radiomic features are capable of providing important information for radiotherapy outcome predictions if they are used in conjunction with

machine learning. The findings of this chapter can facilitate the integration of automatic segmentation frameworks in high-throughput systems for imaging biomarker discovery.

Chapter 4 introduced a multi-modal deep sequence architecture for radiotherapy outcome prediction in brain metastasis *a priori* using treatment-planning MRI and standard clinical attributes. By fusing clinical features and deep-learning-based features extracted from MRI, an outcome prediction framework with an AUC of 0.86 was developed. A 2D visualization framework was also proposed to explain the network's decision. The encouraging findings of this study indicate the potential for early radiation therapy outcome prediction for brain metastases using deep learning of MRI and clinical features.

Chapter 5 proposed a novel 3D deep-learning architecture with residual connections and self-attention to predict the outcome of stereotactic radiation therapy in brain metastasis using volumetric MRI inputs. The proposed architecture outperformed the vanilla residual network and the residual network with CBAM attention in terms of accuracy, F1-score, and AUC. A new 3D visualization framework was also introduced to highlight the regions within the lesion carrying the most impactful information for radiotherapy outcome prediction. The visualization results show the importance of peri-lesional characteristics on treatment-planning MRI in predicting local outcome after radiotherapy. The findings of this chapter demonstrate the capability of volumetric MRI-derived self-attention-guided deep learning features radiotherapy outcome prediction for brain metastasis.

The results of the studies above indicate the significant potential of machine learning for the management of brain metastasis. Starting from the first step of MRI acquisition, machine

learning could be used as a supportive tool in the clinic to help clinicians with segmenting tumours, assessing therapy response, and predicting therapy outcome for possible early treatment adjustments. The contributions and findings of this dissertation are significant considering the heavy burden that manual tumour segmentation imposes on image-guided therapeutic systems in neuro-oncology, and the potential of the proposed methods for accelerating longitudinal tumour size analyses, streamlining image-guided therapy outcome evaluation workflows and optimizing treatments for individual patients to step forward towards a precision oncology paradigm.

6.2. Limitations and Future Work

The findings of this dissertation project, although significant, need to be further validated on larger patient cohorts with multi-center data. One potential future work is to implement the proposed automatic system for radiotherapy outcome assessment in a clinical setting to evaluate its performance while being fed prospective imaging data. This assessment would further investigate the efficacy of the framework in the clinic while exposing its potential limitations.

Since the studies presented in this dissertation have been performed using data acquired from brain metastasis patients treated with hypo-fractionated stereotactic radiation therapy, while tumours with a size of 5 mm and above have been present, the average tumour size was large (around 2.5 cm). This could potentially pose a limitation, as smaller tumours are usually harder to detect and delineate. To mitigate this limitation, we did some experiments on the performance of our proposed segmentation and assessment framework on small tumour sizes (refer to Chapter 2). The performance of the frameworks is lower than its average performance but still acceptable on small tumours (<1cm). More studies need to be carried out to investigate the proposed

segmentation framework, and if required adapt/optimize the network architectures, to improve its performance on smaller tumour sizes.

Transformer-based segmentation models are gaining popularity in literature [202]. It would be interesting to integrate these architectures into the longitudinal tumour segmentation frameworks to explore whether they can improve the performance of automatic therapy outcome assessment systems. We have already shown in Chapter 5 that incorporating self-attention is beneficial to convolutional deep network in medical image analysis for therapy response prediction. It would be interesting to see how pure transformers work in such applications.

The 2D and 3D visualization heatmaps introduced in the dissertation for the radiotherapy outcome prediction models can be investigated in future studies to assess for any significant differences they may manifest between the local control and local failure outcomes. Moreover, as a future work and with close collaboration with clinicians, the reliability and accuracy of these heatmaps can be rigorously studied.

There is a huge interest in differentiating between adverse radiation effect and true progression after SRT and SRS without diagnostic biopsy or surgical resection, yet with reliable sensitivity and specificity. One interesting line of research could be to investigate the possibility of predicting/detecting potential adverse radiation effect early after radiotherapy using image-guided deep-learning models to aid clinicians in detecting ARE non-invasively.

The methods, algorithms, and frameworks introduced in the dissertation are transferable to the studies of other malignancies of the central nervous system, including but not limited to glioma, GBM, and spinal tumours. For example, using the response assessment in neuro-oncology criteria

for gliomas [110] and the proposed segmentation framework introduced in chapter 2, the work is expandable from brain metastasis to gliomas. Moreover, the segmentation, therapy response assessment, and outcome prediction techniques could be adapted for different cancer sites and imaging modalities such as breast, prostate, and liver scanned with CT and ultrasound. The promising results presented in this dissertation endorse future endeavors to apply the method and techniques to other malignancies and imaging modalities.

References

[1] J. B. Posner and N. L. Chernik, "Intracranial metastases from systemic cancer.," *Adv. Neurol.*, vol. 19, pp. 579–92, 1978, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/570349.

[2] B. D. Fox, V. J. Cheung, A. J. Patel, D. Suki, and G. Rao, "Epidemiology of Metastatic Brain Tumors," *Neurosurg. Clin. N. Am.*, vol. 22, no. 1, pp. 1–6, Jan. 2011, doi: 10.1016/j.nec.2010.08.007.

[3] P. Sacks and M. Rahman, "Epidemiology of Brain Metastases," *Neurosurg. Clin. N. Am.*, vol. 31, no. 4, pp. 481–488, Oct. 2020, doi: 10.1016/j.nec.2020.06.001.

[4] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA. Cancer J. Clin.*, vol. 69, no. 1, pp. 7–34, Jan. 2019, doi: 10.3322/caac.21551.

[5] P. W. Sperduto *et al.*, "Estimating survival in patients with gastrointestinal cancers and brain metastases: An update of the graded prognostic assessment for gastrointestinal cancers (GI-GPA)," *Clin. Transl. Radiat. Oncol.*, vol. 18, pp. 39–45, Sep. 2019, doi: 10.1016/j.ctro.2019.06.007.

[6] D. N. Cagney *et al.*, "Incidence and prognosis of patients with brain metastases at diagnosis of systemic malignancy: a population-based study," *Neuro. Oncol.*, vol. 19, no. 11, pp. 1511–1521, Oct. 2017, doi: 10.1093/neuonc/nox077.

[7] P. W. Sperduto *et al.*, "Estimating Survival in Melanoma Patients With Brain Metastases: An Update of the Graded Prognostic Assessment for Melanoma Using Molecular Markers (Melanoma-molGPA)," *Int. J. Radiat. Oncol.*, vol. 99, no. 4, pp. 812–816, Nov. 2017, doi: 10.1016/j.ijrobp.2017.06.2454.

[8] J. Wong, A. Hird, A. Kirou-Mauro, J. Napolskikh, and E. Chow, "Quality of life in brain metastases radiation trials: a literature review," *Curr. Oncol.*, vol. 15, no. 5, pp. 25–45, Oct. 2008.

[9] M. C. Chamberlain, "Brain metastases: a medical neuro-oncology perspective," *Expert Rev. Neurother.*, vol. 10, no. 4, pp. 563–73, Apr. 2010, doi: 10.1586/ern.10.30.

[10] P. W. Sperduto *et al.*, "Diagnosis-specific prognostic factors, indexes, and treatment outcomes for patients with newly diagnosed brain metastases: a multi-institutional analysis of 4,259 patients.," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 77, no. 3, pp. 655–61, Jul. 2010, doi: 10.1016/j.ijrobp.2009.08.025.

[11] G. Rahmathulla, S. A. Toms, and R. J. Weil, "The Molecular Biology of Brain Metastasis," *J. Oncol.*, vol. 2012, pp. 1–16, 2012, doi: 10.1155/2012/723541.

[12] C. Nieder, O. Spanne, M. P. Mehta, A. L. Grosu, and H. Geinitz, "Presentation, patterns of care, and survival in patients with brain metastases," *Cancer*, vol. 117, no. 11, pp. 2505–2512, Jun. 2011, doi: 10.1002/cncr.25707.

[13] P. W. Sperduto *et al.*, “Beyond an Updated Graded Prognostic Assessment (Breast GPA): A Prognostic Index and Trends in Treatment and Survival in Breast Cancer Brain Metastases From 1985 to Today,” *Int. J. Radiat. Oncol.*, vol. 107, no. 2, pp. 334–343, Jun. 2020, doi: 10.1016/j.ijrobp.2020.01.051.

[14] P. W. Sperduto *et al.*, “Estimating Survival in Patients With Lung Cancer and Brain Metastases,” vol. 3, no. 6, p. 827, Jun. 2017, doi: 10.1001/jamaoncol.2016.3834.

[15] T. K. Owonikoko *et al.*, “Current approaches to the treatment of metastatic brain tumours,” *Nat. Rev. Clin. Oncol.*, vol. 11, no. 4, pp. 203–222, 2014, doi: 10.1038/nrclinonc.2014.25.Current.

[16] C. M. Carapella, N. Gorgoglione, and P. A. Oppido, “The role of surgical resection in patients with brain metastases,” *Curr. Opin. Oncol.*, vol. 30, no. 6, pp. 390–395, 2018, doi: 10.1097/CCO.0000000000000484.

[17] S. Campos *et al.*, “Brain metastasis from an unknown primary, or primary brain tumour? A diagnostic dilemma.,” *Curr. Oncol.*, vol. 16, no. 1, pp. 62–6, Jan. 2009, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19229374.

[18] E. W. Sankey *et al.*, “Operative and peri-operative considerations in the management of brain metastasis,” *Cancer Med.*, vol. 8, no. 16, pp. 6809–6831, Nov. 2019, doi: 10.1002/cam4.2577.

[19] R. A. Patchell *et al.*, “Postoperative Radiotherapy in the Treatment of Single Metastases to the Brain,” *JAMA*, vol. 280, no. 17, Nov. 1998, doi: 10.1001/jama.280.17.1485.

[20] A. Mahajan *et al.*, “Post-operative stereotactic radiosurgery versus observation for completely resected brain metastases: a single-centre, randomised, controlled, phase 3 trial,” *Lancet Oncol.*, vol. 18, no. 8, pp. 1040–1048, Aug. 2017, doi: 10.1016/S1470-2045(17)30414-X.

[21] P. D. Brown *et al.*, “Postoperative stereotactic radiosurgery compared with whole brain radiotherapy for resected metastatic brain disease (NCCTG N107C/CEC·3): a multicentre, randomised, controlled, phase 3 trial,” *Lancet Oncol.*, vol. 18, no. 8, pp. 1049–1060, Aug. 2017, doi: 10.1016/S1470-2045(17)30441-2.

[22] P. D. Brown, M. S. Ahluwalia, O. H. Khan, A. L. Asher, J. S. Wefel, and V. Gondi, “Whole-brain radiotherapy for brain metastases: Evolution or revolution?,” *J. Clin. Oncol.*, vol. 36, no. 5, pp. 483–491, 2018, doi: 10.1200/JCO.2017.75.9589.

[23] P. D. Brown *et al.*, “Effect of Radiosurgery Alone vs Radiosurgery With Whole Brain Radiation Therapy on Cognitive Function in Patients With 1 to 3 Brain Metastases,” *JAMA*, vol. 316, no. 4, p. 401, Jul. 2016, doi: 10.1001/jama.2016.9839.

[24] E. L. Chang *et al.*, “Neurocognition in patients with brain metastases treated with radiosurgery or radiosurgery plus whole-brain irradiation: a randomised controlled trial,” *Lancet Oncol.*, vol. 10, no. 11, pp. 1037–1044, Nov. 2009, doi: 10.1016/S1470-2045(09)70263-3.

[25] P. Mulvenna *et al.*, “Dexamethasone and supportive care with or without whole brain

radiotherapy in treating patients with non-small cell lung cancer with brain metastases unsuitable for resection or stereotactic radiotherapy (QUARTZ): results from a phase 3, non-inferiority," *Lancet*, vol. 388, no. 10055, pp. 2004–2014, Oct. 2016, doi: 10.1016/S0140-6736(16)30825-X.

[26] M. O'Beirn *et al.*, "The Expanding Role of Radiosurgery for Brain Metastases," *Medicines*, vol. 5, no. 3, p. 90, Aug. 2018, doi: 10.3390/medicines5030090.

[27] P. Venkatesan, "SRS is non-inferior to WBRT for brain metastases," *Lancet Oncol.*, vol. 19, no. 8, p. e386, Aug. 2018, doi: 10.1016/S1470-2045(18)30490-X.

[28] N. U. Lin *et al.*, "Response assessment criteria for brain metastases: proposal from the RANO group," *Lancet Oncol.*, vol. 16, no. 6, pp. e270–e278, Jun. 2015, doi: 10.1016/S1470-2045(15)70057-4.

[29] F. R. Vogenberg, C. Isaacson Barash, and M. Pursel, "Personalized medicine: part 1: evolution and development into theranostics.," *Pharmacol. Ther.*, vol. 35, no. 10, pp. 560–576, Oct. 2010.

[30] E. Y. Saito *et al.*, "Whole brain radiation therapy in management of brain metastasis: results and prognostic factors.," *Radiat. Oncol.*, vol. 1, p. 20, Jun. 2006, doi: 10.1186/1748-717X-1-20.

[31] H. Soliman, S. Das, D. A. Larson, and A. Sahgal, "Stereotactic radiosurgery (SRS) in the modern management of patients with brain metastases," *Oncotarget*, vol. 7, no. 11, pp. 12318–12330, Mar. 2016, doi: 10.18632/oncotarget.7131.

[32] P. W. Sperduto *et al.*, "Summary Report on the Graded Prognostic Assessment: An Accurate and Facile Diagnosis-Specific Tool to Estimate Survival for Patients With Brain Metastases," *J. Clin. Oncol.*, vol. 30, no. 4, pp. 419–425, Feb. 2012, doi: 10.1200/JCO.2011.38.0527.

[33] J. W. Yates, B. Chalmer, and F. P. McKegney, "Evaluation of patients with advanced cancer using the karnofsky performance status," *Cancer*, vol. 45, no. 8, pp. 2220–2224, Apr. 1980, doi: 10.1002/1097-0142(19800415)45:8<2220::AID-CNCR2820450835>3.0.CO;2-Q.

[34] P. W. Sperduto *et al.*, "Estimating Survival in Patients With Lung Cancer and Brain Metastases," *JAMA Oncol.*, vol. 3, no. 6, p. 827, Jun. 2017, doi: 10.1001/jamaoncol.2016.3834.

[35] "DDT Glossary." https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/ddt-glossary (accessed Mar. 22, 2020).

[36] A. Boire *et al.*, "Liquid biopsy in central nervous system metastases: a RANO review and proposals for clinical applications.," *Neuro. Oncol.*, vol. 21, no. 5, pp. 571–584, 2019, doi: 10.1093/neuonc/noz012.

[37] F. Xiao *et al.*, "Cerebrospinal fluid biomarkers for brain tumor detection: clinical roles and current progress.," *Am. J. Transl. Res.*, vol. 12, no. 4, pp. 1379–1396, 2020, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/32355549.

[38] F. Calabrese *et al.*, "Are There New Biomarkers in Tissue and Liquid Biopsies for the Early

Detection of Non-Small Cell Lung Cancer?," *J. Clin. Med.*, vol. 8, no. 3, Mar. 2019, doi: 10.3390/jcm8030414.

[39] K. Cyll *et al.*, "Tumour heterogeneity poses a significant challenge to cancer biomarker research," *Br. J. Cancer*, vol. 117, no. 3, pp. 367–375, Jul. 2017, doi: 10.1038/bjc.2017.171.

[40] W. B. Pope, "Brain metastases: neuroimaging.," *Handb. Clin. Neurol.*, vol. 149, pp. 89–112, 2018, doi: 10.1016/B978-0-12-811161-1.00007-4.

[41] K. Brindle, "New approaches for imaging tumour responses to treatment," *Nat. Rev. Cancer*, vol. 8, no. 2, pp. 94–107, Feb. 2008, doi: 10.1038/nrc2289.

[42] A. Sadeghi-Naini *et al.*, "Imaging innovations for cancer therapy response monitoring," *Imaging Med.*, vol. 4, no. 3, pp. 311–327, Jun. 2012, doi: 10.2217/iim.12.23.

[43] H. Mehrabian, J. Detsky, H. Soliman, A. Sahgal, and G. J. Stanisz, "Advanced Magnetic Resonance Imaging Techniques in Management of Brain Metastases," *Front. Oncol.*, vol. 9, Jun. 2019, doi: 10.3389/fonc.2019.00440.

[44] J. P. B. O'Connor *et al.*, "Imaging biomarker roadmap for cancer studies," *Nat. Rev. Clin. Oncol.*, vol. 14, no. 3, pp. 169–186, Mar. 2017, doi: 10.1038/nrclinonc.2016.162.

[45] T. Chithambaram and K. Perumal, "Brain Tumor Detection and Segmentation in MRI Images Using Neural Network," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 7, no. 3, pp. 155–164, 2017, doi: 10.23956/ijarcsse/V7I3/0164.

[46] L. Chen, Y. Wu, A. M. DSouza, A. Z. Abidin, A. Wismuller, and C. Xu, "MRI Tumor Segmentation with Densely Connected 3D CNN," 2018, doi: 10.1117/12.2293394.

[47] M. Havaei *et al.*, "Brain tumor segmentation with Deep Neural Networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017, doi: 10.1016/j.media.2016.05.004.

[48] B. H. Menze *et al.*, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015, doi: 10.1109/TMI.2014.2377694.

[49] J. Xue *et al.*, "Deep learning-based detection and segmentation-assisted management of brain metastases," *Neuro. Oncol.*, vol. 22, no. 4, pp. 505–514, 2020, doi: 10.1093/neuonc/noz234.

[50] J. Cho *et al.*, "Deep Learning-Based Computer-Aided Detection System for Automated Treatment Response Assessment of Brain Metastases on 3D MRI," *Front. Oncol.*, vol. 11, p. 739639, 2021, doi: 10.3389/fonc.2021.739639.

[51] E. Karami *et al.*, "Quantitative MRI Biomarkers of Stereotactic Radiotherapy Outcome in Brain Metastasis.," *Sci. Rep.*, vol. 9, no. 1, p. 19830, Dec. 2019, doi: 10.1038/s41598-019-56185-5.

[52] M. Cabezas *et al.*, "Survival prediction using ensemble tumor segmentation and transfer learning," Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.04274.

[53] A. Diamant, A. Chatterjee, M. Vallières, G. Shenouda, and J. Seuntjens, "Deep learning in head & neck cancer outcome prediction," *Sci. Rep.*, vol. 9, no. 1, p. 2764, Dec. 2019,

doi: 10.1038/s41598-019-39206-1.

[54] A. Jalalifar, H. Soliman, A. Sahgal, and A. Sadeghi-Naini, “A Cascaded Deep-Learning Framework for Segmentation of Metastatic Brain Tumors Before and After Stereotactic Radiation Therapy.,” *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 1063–1066, 2020, doi: 10.1109/EMBC44109.2020.9175489.

[55] D. Shen, G. Wu, and H. I. Suk, “Deep Learning in Medical Image Analysis,” *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017, doi: 10.1146/annurev-bioeng-071516-044442.

[56] K. J. Shanthi and M. S. Kumar, “Skull stripping and automatic segmentation of brain MRI using seed growth and threshold techniques,” in *2007 International Conference on Intelligent and Advanced Systems*, Nov. 2007, pp. 422–426, doi: 10.1109/ICIAS.2007.4658421.

[57] E. Karami *et al.*, “An Automatic Framework for Segmentation of Brain Tumours at Follow-up Scans after Radiation Therapy *,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2019, pp. 463–466, doi: 10.1109/EMBC.2019.8856858.

[58] A. Bleau and L. J. Leon, “Watershed-Based Segmentation and Region Merging,” *Comput. Vis. Image Underst.*, vol. 77, no. 3, pp. 317–370, Mar. 2000, doi: 10.1006/cviu.1999.0822.

[59] R. Malladi, J. A. Sethian, and B. C. Vemuri, “Shape modeling with front propagation: a level set approach,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 2, pp. 158–175, 1995, doi: 10.1109/34.368173.

[60] S. Supot, C. Thanapong, P. Chuchart, and S. Manas, “Segmentation of Magnetic Resonance Images Using Discrete Curve Evolution and Fuzzy Clustering,” in *2007 IEEE International Conference on Integration Technology*, Mar. 2007, pp. 697–700, doi: 10.1109/ICITECHNOLOGY.2007.4290409.

[61] A. Jalalifar, H. Soliman, M. Ruschin, A. Sahgal, and A. Sadeghi-Naini, “A Brain Tumor Segmentation Framework Based on Outlier Detection Using One-Class Support Vector Machine,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2020, pp. 1067–1070, doi: 10.1109/EMBC44109.2020.9176263.

[62] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” pp. 1–8, 2015, doi: 10.1007/978-3-319-24574-4_28.

[63] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” Jun. 2016, [Online]. Available: http://arxiv.org/abs/1606.06650.

[64] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” in *Computer Vision -- ECCV 2014*, 2014, pp. 818–833.

[65] O. Oktay *et al.*, “Attention U-Net: Learning Where to Look for the Pancreas,” Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.03999.

[66] N. S. Punn and S. Agarwal, “Inception U-Net Architecture for Semantic Segmentation to

Identify Nuclei in Microscopy Cell Images," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 16, no. 1, pp. 1–15, Apr. 2020, doi: 10.1145/3376922.

[67] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," Jul. 2018, [Online]. Available: http://arxiv.org/abs/1807.10165.

[68] M. Nishino *et al.*, "Personalized tumor response assessment in the era of molecular medicine: cancer-specific and therapy-specific response criteria to complement pitfalls of RECIST," *Am. J. Roentgenol.*, vol. 198, no. 4, pp. 737–45, Apr. 2012, doi: 10.2214/AJR.11.7483.

[69] L. G. Kessler *et al.*, "The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions," *Stat. Methods Med. Res.*, vol. 24, no. 1, pp. 9–26, Feb. 2015, doi: 10.1177/0962280214537333.

[70] A. Ahmed, P. Gibbs, M. Pickles, and L. Turnbull, "Texture analysis in assessment and prediction of chemotherapy response in breast cancer," *J. Magn. Reson. Imaging*, vol. 38, no. 1, pp. 89–101, Jul. 2013, doi: 10.1002/jmri.23971.

[71] C. J. Rose *et al.*, "Quantifying spatial heterogeneity in dynamic contrast-enhanced MRI parameter maps," *Magn. Reson. Med.*, vol. 62, no. 2, pp. 488–499, Aug. 2009, doi: 10.1002/mrm.22003.

[72] M. C. Preul, Z. Caramanos, R. Leblanc, J. G. Villemure, and D. L. Arnold, "Using pattern analysis ofin vivo proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors," *NMR Biomed.*, vol. 11, no. 4–5, pp. 192–200, Jun. 1998, doi: 10.1002/(SICI)1099-1492(199806/08)11:4/5<192::AID-NBM535>3.0.CO;2-3.

[73] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," *Radiology*, vol. 278, no. 2, pp. 563–577, Feb. 2016, doi: 10.1148/radiol.2015151169.

[74] S. Rizzo *et al.*, "Radiomics: the facts and the challenges of image analysis," *Eur. Radiol. Exp.*, vol. 2, no. 1, p. 36, Dec. 2018, doi: 10.1186/s41747-018-0068-z.

[75] H. J. W. L. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat. Commun.*, vol. 5, no. 1, p. 4006, Sep. 2014, doi: 10.1038/ncomms5006.

[76] S. Bae *et al.*, "Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction," *Radiology*, vol. 289, no. 3, pp. 797–806, Dec. 2018, doi: 10.1148/radiol.2018180200.

[77] E. Karami *et al.*, "Quantitative MRI Biomarkers of Stereotactic Radiotherapy Outcome in Brain Metastasis," *Sci. Rep.*, vol. 9, no. 1, p. 19830, Dec. 2019, doi: 10.1038/s41598-019-56185-5.

[78] A. Mouraviev *et al.*, "Use of Radiomics for the Prediction of Local Control of Brain Metastases After Stereotactic Radiosurgery.," *Neuro. Oncol.*, Jan. 2020, doi: 10.1093/neuonc/noaa007.

[79] Y. Lin, Z. Nie, and H. Ma, “Structural Damage Detection with Automatic Feature-Extraction through Deep Learning,” *Comput. Civ. Infrastruct. Eng.*, vol. 32, no. 12, pp. 1025–1046, Dec. 2017, doi: 10.1111/mice.12313.

[80] W. Zhao and S. Du, “Spectral–Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016, doi: 10.1109/TGRS.2016.2543748.

[81] C. You *et al.*, “Class-Aware Generative Adversarial Transformers for Medical Image Segmentation,” Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.10737.

[82] L. Yang *et al.*, “NuSeT: A deep learning tool for reliably separating and analyzing crowded cells,” *PLOS Comput. Biol.*, vol. 16, no. 9, p. e1008193, Sep. 2020, doi: 10.1371/journal.pcbi.1008193.

[83] C. You, R. Zhao, L. Staib, and J. S. Duncan, “Momentum Contrastive Voxel-wise Representation Learning for Semi-supervised Volumetric Medical Image Segmentation,” May 2021, [Online]. Available: http://arxiv.org/abs/2105.07059.

[84] C. You, Y. Zhou, R. Zhao, L. Staib, and J. S. Duncan, “SimCVD: Simple Contrastive Voxel-Wise Representation Distillation for Semi-Supervised Medical Image Segmentation,” Aug. 2021, [Online]. Available: http://arxiv.org/abs/2108.06227.

[85] A. Lagree *et al.*, “A review and comparison of breast tumor cell nuclei segmentation performances using deep convolutional neural networks,” *Sci. Rep.*, vol. 11, no. 1, p. 8025, 2021, doi: 10.1038/s41598-021-87496-1.

[86] P. Afshar, A. Mohammadi, K. N. Plataniotis, A. Oikonomou, and H. Benali, “From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities,” *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 132–160, Jul. 2019, doi: 10.1109/MSP.2019.2900993.

[87] E. Wetzer, J. Gay, H. Harlin, J. Lindblad, and N. Sladoje, “When Texture Matters: Texture-Focused Cnns Outperform General Data Augmentation and Pretraining in Oral Cancer Detection,” in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2020, pp. 517–521, doi: 10.1109/ISBI45749.2020.9098424.

[88] Q. Lyu, C. You, H. Shan, and G. Wang, “Super-resolution MRI through Deep Learning,” Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.06776.

[89] S. J. Cho, L. Sunwoo, S. H. Baik, Y. J. Bae, B. S. Choi, and J. H. Kim, “Brain metastasis detection using machine learning: a systematic review and meta-analysis,” *Neuro. Oncol.*, vol. 23, no. 2, pp. 214–225, Feb. 2021, doi: 10.1093/neuonc/noaa232.

[90] R. M. Auchter *et al.*, “A multiinstitutional outcome and prognostic factor analysis of radiosurgery for resectable single brain metastasis,” *Int. J. Radiat. Oncol.*, vol. 35, no. 1, pp. 27–35, Apr. 1996, doi: 10.1016/S0360-3016(96)85008-5.

[91] P. K. Sneed *et al.*, “Adverse radiation effect after stereotactic radiosurgery for brain metastases: incidence, time course, and risk factors,” *J. Neurosurg.*, vol. 123, no. 2, pp.

373–386, Aug. 2015, doi: 10.3171/2014.10.JNS141610.

[92] M. T. Truong *et al.*, “Results of surgical resection for progression of brain metastases previously treated by gamma knife radiosurgery,” *Neurosurgery*, vol. 59, no. 1, pp. 86–97; discussion 86-97, Jul. 2006, doi: 10.1227/01.NEU.0000219858.80351.38.

[93] H. Kano, D. Kondziolka, J. Lobato-Polo, O. Zorro, J. C. Flickinger, and L. D. Lunsford, “T1/T2 Matching to Differentiate Tumor Growth From Radiation Effects After Stereotactic Radiosurgery,” *Neurosurgery*, vol. 66, no. 3, pp. 486–492, Mar. 2010, doi: 10.1227/01.NEU.0000360391.35749.A5.

[94] E. Karami *et al.*, “Quantitative MRI Biomarkers of Stereotactic Radiotherapy Outcome in Brain Metastasis,” *Sci. Rep.*, vol. 9, no. 1, p. 19830, Dec. 2019, doi: 10.1038/s41598-019-56185-5.

[95] A. Tiwari, S. Srivastava, and M. Pant, “Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019,” *Pattern Recognit. Lett.*, vol. 131, pp. 244–260, Mar. 2020, doi: 10.1016/j.patrec.2019.11.020.

[96] Z. Liu *et al.*, “Deep Learning Based Brain Tumor Segmentation: A Survey,” Jul. 2020, [Online]. Available: http://arxiv.org/abs/2007.09479.

[97] N. Gordillo, E. Montseny, and P. Sobrevilla, “State of the art survey on MRI brain tumor segmentation,” *Magn. Reson. Imaging*, vol. 31, no. 8, pp. 1426–1438, 2013, doi: 10.1016/j.mri.2013.05.002.

[98] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, “A deep learning model integrating FCNNs and CRFs for brain tumor segmentation,” *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2018, doi: 10.1016/j.media.2017.10.002.

[99] A. Işın, C. Direkoğlu, and M. Şah, “Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods,” *Procedia Comput. Sci.*, vol. 102, pp. 317–324, 2016, doi: 10.1016/j.procs.2016.09.407.

[100] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, “Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks,” *Med. Image Underst. Anal.*, pp. 1–12, 2017, doi: 10.1007/978-3-319-60964-5_44.

[101] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 2015, [Online]. Available: http://arxiv.org/abs/1505.04597.

[102] R. Ranjbarzadeh, A. Bagherian Kasgari, S. Jafarzadeh Ghoushchi, S. Anari, M. Naseri, and M. Bendechache, “Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images,” *Sci. Rep.*, vol. 11, no. 1, p. 10930, Dec. 2021, doi: 10.1038/s41598-021-90428-8.

[103] M. Islam, V. VS, V. J. M. Jose, N. Wijethilake, U. Utkarsh, and H. Ren, “Brain Tumor Segmentation and Survival Prediction using 3D Attention UNet,” Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.00985.

[104] T. Magadza and S. Viriri, “Deep Learning for Brain Tumor Segmentation: A Survey of State-of-the-Art,” *J. Imaging*, vol. 7, no. 2, p. 19, Jan. 2021, doi: 10.3390/jimaging7020019.

[105] E. Gates, J. G. Pauloski, D. Schellingerhout, and D. Fuentes, “Glioma Segmentation and a Simple Accurate Model for Overall Survival Prediction,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer International Publishing, 2019, pp. 476–484.

[106] L. Pei, L. Vidyaratne, M. M. Rahman, and K. M. Iftekharuddin, “Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images,” *Sci. Rep.*, vol. 10, no. 1, p. 19726, Dec. 2020, doi: 10.1038/s41598-020-74419-9.

[107] Y. Zhu *et al.*, “Semi-Automatic Segmentation Software for Quantitative Clinical Brain Glioblastoma Evaluation,” *Acad. Radiol.*, vol. 19, no. 8, pp. 977–985, Aug. 2012, doi: 10.1016/j.acra.2012.03.026.

[108] H. Mehrabian, K. L. Desmond, H. Soliman, A. Sahgal, and G. J. Stanisz, “Differentiation between Radiation Necrosis and Tumor Progression Using Chemical Exchange Saturation Transfer.,” *Clin. Cancer Res.*, vol. 23, no. 14, pp. 3667–3675, Jul. 2017, doi: 10.1158/1078-0432.CCR-16-2265.

[109] A. Sinha and J. Dolz, “Multi-scale guided attention for medical image segmentation,” Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.02849.

[110] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nat. Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021, doi: 10.1038/s41592-020-01008-z.

[111] D. Oft *et al.*, “Volumetric Regression in Brain Metastases After Stereotactic Radiotherapy: Time Course, Predictors, and Significance,” *Front. Oncol.*, vol. 10, p. 590980, Jan. 2020, doi: 10.3389/fonc.2020.590980.

[112] O. Sagi and L. Rokach, “Ensemble learning: A survey,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, Jul. 2018, doi: 10.1002/widm.1249.

[113] B. Vellayappan *et al.*, “Diagnosis and Management of Radiation Necrosis in Patients With Brain Metastases.,” *Front. Oncol.*, vol. 8, p. 395, 2018, doi: 10.3389/fonc.2018.00395.

[114] N. U. Lin *et al.*, “Challenges relating to solid tumour brain metastases in clinical trials, part 1: patient population, response, and progression. A report from the RANO group,” *Lancet Oncol.*, vol. 14, no. 10, pp. e396–e406, Sep. 2013, doi: 10.1016/S1470-2045(13)70311-5.

[115] K. Douri *et al.*, “RANO-BM response criteria verification study in a SRS-treated cohort,” *Can. J. Neurol. Sci.*, vol. 49, no. s1, pp. S45–S45, Jun. 2022, doi: 10.1017/cjn.2022.229.

[116] G. Fischedick, U. Haverkamp, and A. Fischedick, “Are Three-Dimensional Volumetric Measurements of Brain Metastasis the Future for Disease Control? A Comparative Study,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 99, no. 2, pp. E659–E660, Oct. 2017, doi: 10.1016/j.ijrobp.2017.06.2193.

[117] M. J. Follwell *et al.*, “Volume specific response criteria for brain metastases following salvage stereotactic radiosurgery and associated predictors of response,” *Acta Oncol. (Madr).*, vol. 51, no. 5, pp. 629–35, May 2012, doi: 10.3109/0284186X.2012.681066.

[118] A. L. Stockham *et al.*, “Conventional MRI does not reliably distinguish radiation necrosis from tumor recurrence after stereotactic radiosurgery,” *J. Neurooncol.*, vol. 109, no. 1, pp. 149–158, Aug. 2012, doi: 10.1007/s11060-012-0881-9.

[119] I. M. Dequesada, R. G. Quisling, A. Yachnis, and W. A. Friedman, “Can Standard Magnetic Resonance Imaging Reliably Distinguish Recurrent Tumor From Radiation Necrosis After Radiosurgery For Brain Metastases? A Radiographic-pathological Study,” *Neurosurgery*, vol. 63, no. 5, pp. 898–904, Nov. 2008, doi: 10.1227/01.NEU.0000333263.31870.31.

[120] J. S. Detsky *et al.*, “Differentiating radiation necrosis from tumor progression in brain metastases treated with stereotactic radiotherapy: utility of intravoxel incoherent motion perfusion MRI and correlation with histopathology,” *J. Neurooncol.*, vol. 134, no. 2, pp. 433–441, Sep. 2017, doi: 10.1007/s11060-017-2545-2.

[121] Y. Yunqi *et al.*, “Quantitative MR Perfusion for the Differentiation of Recurrence and Radionecrosis in Hypoperfusion and Hyperperfusion Brain Metastases After Gamma Knife Radiosurgery,” *Front. Neurol.*, vol. 13, p. 823731, 2022, doi: 10.3389/fneur.2022.823731.

[122] R. Wiggenraad *et al.*, “Pseudo-progression after stereotactic radiotherapy of brain metastases: lesion analysis using MRI cine-loops,” *J. Neurooncol.*, vol. 119, no. 2, pp. 437–443, Sep. 2014, doi: 10.1007/s11060-014-1519-x.

[123] N. Lamba, P. Y. Wen, and A. A. Aizer, “Epidemiology of brain metastases and leptomeningeal disease,” *Neuro. Oncol.*, vol. 23, no. 9, pp. 1447–1456, Sep. 2021, doi: 10.1093/neuonc/noab101.

[124] Q. T. Ostrom, C. H. Wright, and J. S. Barnholtz-Sloan, “Brain metastases: epidemiology,” 2018, pp. 27–42.

[125] N. Lamba *et al.*, “Racial disparities in supportive medication use among older patients with brain metastases: a population-based analysis,” *Neuro. Oncol.*, vol. 22, no. 9, pp. 1339–1347, Sep. 2020, doi: 10.1093/neuonc/noaa054.

[126] L. L. Muldoon *et al.*, “Chemotherapy Delivery Issues in Central Nervous System Malignancy: A Reality Check,” *J. Clin. Oncol.*, vol. 25, no. 16, pp. 2295–2305, Jun. 2007, doi: 10.1200/JCO.2006.09.9861.

[127] V. A. Venur, V. Karivedu, and M. S. Ahluwalia, “Systemic therapy for brain metastases,” 2018, pp. 137–153.

[128] S. Amin, M. J. Baine, J. L. Meza, and C. Lin, “Association of Immunotherapy With Survival Among Patients With Brain Metastases Whose Cancer Was Managed With Definitive Surgery of the Primary Tumor,” *JAMA Netw. Open*, vol. 3, no. 9, p. e2015444, Sep. 2020, doi: 10.1001/jamanetworkopen.2020.15444.

[129] A. S. Berghoff, V. A. Venur, M. Preusser, and M. S. Ahluwalia, “Immune Checkpoint Inhibitors in Brain Metastases: From Biology to Treatment,” *Am. Soc. Clin. Oncol. Educ. B.*, no. 36, pp. e116–e122, May 2016, doi: 10.1200/EDBK_100005.

[130] G. Xiao *et al.*, “Immune checkpoint inhibitors for brain metastases in non-small-cell lung cancer: from rationale to clinical application,” *Immunotherapy*, vol. 13, no. 12, pp. 1031–

1051, Aug. 2021, doi: 10.2217/imt-2020-0262.

[131] H. Okada *et al.*, "Immunotherapy response assessment in neuro-oncology: a report of the RANO working group," *Lancet Oncol.*, vol. 16, no. 15, pp. e534–e542, Nov. 2015, doi: 10.1016/S1470-2045(15)00088-1.

[132] C. Nieder and M. P. Mehta, "Prognostic indices for brain metastases – usefulness and challenges," *Radiat. Oncol.*, vol. 4, no. 1, p. 10, Dec. 2009, doi: 10.1186/1748-717X-4-10.

[133] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016, doi: 10.1148/radiol.2015151169.

[134] S. Rizzo *et al.*, "Radiomics: the facts and the challenges of image analysis," *Eur. Radiol. Exp.*, vol. 2, no. 1, 2018, doi: 10.1186/s41747-018-0068-z.

[135] Z. Bodalal, S. Trebeschi, T. D. L. Nguyen-Kim, W. Schats, and R. Beets-Tan, "Radiogenomics: bridging imaging and genomics," *Abdom. Radiol.*, vol. 44, no. 6, pp. 1960–1984, Jun. 2019, doi: 10.1007/s00261-019-02028-w.

[136] M. Diehn *et al.*, "Identification of noninvasive imaging surrogates for brain tumor gene-expression modules," *Proc. Natl. Acad. Sci.*, vol. 105, no. 13, pp. 5213–5218, Apr. 2008, doi: 10.1073/pnas.0801279105.

[137] P. O. Zinn *et al.*, "Radiogenomic Mapping of Edema/Cellular Invasion MRI-Phenotypes in Glioblastoma Multiforme," *PLoS One*, vol. 6, no. 10, p. e25451, Oct. 2011, doi: 10.1371/journal.pone.0025451.

[138] A. Mouraviev *et al.*, "Use of radiomics for the prediction of local control of brain metastases after stereotactic radiosurgery," *Neuro. Oncol.*, vol. 22, no. 6, pp. 797–805, Jun. 2020, doi: 10.1093/neuonc/noaa007.

[139] J. J. M. van Griethuysen *et al.*, "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.

[140] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.

[141] C. DING and H. PENG, "MINIMUM REDUNDANCY FEATURE SELECTION FROM MICROARRAY GENE EXPRESSION DATA," *J. Bioinform. Comput. Biol.*, vol. 03, no. 02, pp. 185–205, Apr. 2005, doi: 10.1142/S0219720005001004.

[142] C.-Y. Liao *et al.*, "Enhancement of Radiosurgical Treatment Outcome Prediction Using MRI Radiomics in Patients with Non-Small Cell Lung Cancer Brain Metastases," *Cancers (Basel).*, vol. 13, no. 16, p. 4030, Aug. 2021, doi: 10.3390/cancers13164030.

[143] J. Jin, H. Zhu, Y. Teng, Y. Ai, C. Xie, and X. Jin, "The Accuracy and Radiomics Feature Effects of Multiple U-net-Based Automatic Segmentation Models for Transvaginal Ultrasound Images of Cervical Cancer," *J. Digit. Imaging*, Mar. 2022, doi: 10.1007/s10278-022-00620-z.

[144] Y. Teng *et al.*, "The Effects of Automatic Segmentations on Preoperative Lymph Node Status Prediction Models With Ultrasound Radiomics for Patients With Early Stage Cervical Cancer," *Technol. Cancer Res. Treat.*, vol. 21, p. 153303382210993, Jan. 2022, doi: 10.1177/15330338221099396.

[145] C. C. Schag, R. L. Heinrich, and P. A. Ganz, "Karnofsky performance status revisited: reliability, validity, and guidelines.," *J. Clin. Oncol.*, vol. 2, no. 3, pp. 187–193, Mar. 1984, doi: 10.1200/JCO.1984.2.3.187.

[146] A. Kapadia *et al.*, "Temporal evolution of perfusion parameters in brain metastases treated with stereotactic radiosurgery: comparison of intravoxel incoherent motion and dynamic contrast enhanced MRI.," *J. Neurooncol.*, vol. 135, no. 1, pp. 119–127, Oct. 2017, doi: 10.1007/s11060-017-2556-z.

[147] K. L. Desmond *et al.*, "Chemical exchange saturation transfer for predicting response to stereotactic radiosurgery in human brain metastasis," *Magn. Reson. Med.*, vol. 78, no. 3, pp. 1110–1120, Sep. 2017, doi: 10.1002/mrm.26470.

[148] E. Karami, M. Ruschin, H. Soliman, A. Sahgal, G. J. Stanisz, and A. Sadeghi-Naini, "An MR Radiomics Framework for Predicting the Outcome of Stereotactic Radiation Therapy in Brain Metastasis," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2019, pp. 1022–1025, doi: 10.1109/EMBC.2019.8856558.

[149] D. E. Rumelhart and J. L. McClelland, "Learning Internal Representations by Error Propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, D. Rumelhart and J. McClelland, Eds. MIT Press, 1987, pp. 318–362.

[150] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[151] D. R. Wilson and T. R. Martinez, "The general inefficiency of batch training for gradient descent learning," *Neural Networks*, vol. 16, no. 10, pp. 1429–1451, Dec. 2003, doi: 10.1016/S0893-6080(03)00138-2.

[152] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.

[153] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*, 2009, pp. 1–8, doi: 10.1145/1553374.1553380.

[154] F. Chollet and others, "Keras." GitHub, 2015, [Online]. Available: https://github.com/fchollet/keras.

[155] Martín Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems." 2015, [Online]. Available: https://www.tensorflow.org/.

[156] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[157] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Comput. Sci. Eng.*, vol. 9, no. 3,

pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.

[158] D. A. Hardesty and P. Nakaji, “The Current and Future Treatment of Brain Metastases,” *Front. Surg.*, vol. 3, May 2016, doi: 10.3389/fsurg.2016.00030.

[159] T. P. Coroller *et al.*, “CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma,” *Radiother. Oncol.*, vol. 114, no. 3, pp. 345–350, Mar. 2015, doi: 10.1016/j.radonc.2015.02.015.

[160] H. C. Kniep *et al.*, “Radiomics of Brain MRI: Utility in Prediction of Metastatic Tumor Type,” *Radiology*, vol. 290, no. 2, pp. 479–487, Feb. 2019, doi: 10.1148/radiol.2018180946.

[161] A. Sadeghi-Naini *et al.*, “Low-frequency quantitative ultrasound imaging of cell death in vivo,” *Med. Phys.*, vol. 40, no. 8, p. 082901, Jul. 2013, doi: 10.1118/1.4812683.

[162] A. S. Panayides *et al.*, “AI in Medical Imaging Informatics: Current Challenges and Future Directions,” *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 7, pp. 1837–1857, Jul. 2020, doi: 10.1109/JBHI.2020.2991043.

[163] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[164] G. Nicora, F. Vitali, A. Dagliati, N. Geifman, and R. Bellazzi, “Integrated Multi-Omics Analyses in Oncology: A Review of Machine Learning Methods and Tools,” *Front. Oncol.*, vol. 10, Jun. 2020, doi: 10.3389/fonc.2020.01030.

[165] A. Mouraviev *et al.*, “Use of Radiomics for the Prediction of Local Control of Brain Metastases After Stereotactic Radiosurgery.,” *Neuro. Oncol.*, Jan. 2020, doi: 10.1093/neuonc/noaa007.

[166] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, “A survey: Deep learning for hyperspectral image classification with few labeled samples,” *Neurocomputing*, vol. 448, pp. 179–204, Aug. 2021, doi: 10.1016/j.neucom.2021.03.035.

[167] D. R. Sarvamangala and R. V. Kulkarni, “Convolutional neural networks in medical image understanding: a survey,” *Evol. Intell.*, vol. 15, no. 1, pp. 1–22, Mar. 2022, doi: 10.1007/s12065-020-00540-3.

[168] M. Wang, Q. Zhang, S. Lam, J. Cai, and R. Yang, “A Review on Application of Deep Learning Algorithms in External Beam Radiotherapy Automated Treatment Planning,” *Front. Oncol.*, vol. 10, Oct. 2020, doi: 10.3389/fonc.2020.580919.

[169] Y. J. CHA *et al.*, “Prediction of Response to Stereotactic Radiosurgery for Brain Metastases Using Convolutional Neural Networks,” *Anticancer Res.*, vol. 38, no. 9, pp. 5437–5445, Sep. 2018, doi: 10.21873/anticanres.12875.

[170] S. A. Jalalifar, H. Soliman, A. Sahgal, and A. Sadeghi-Naini, “Predicting the outcome of radiotherapy in brain metastasis by integrating the clinical and MRI-based deep learning features,” *Med. Phys.*, Jul. 2022, doi: 10.1002/mp.15814.

[171] M.-H. Guo *et al.*, “Attention mechanisms in computer vision: A survey,” *Comput. Vis.*

Media, Mar. 2022, doi: 10.1007/s41095-022-0271-y.

[172] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[173] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers," 2020, pp. 213–229.

[174] J. Fu *et al.*, "Dual Attention Network for Scene Segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 3141–3149, doi: 10.1109/CVPR.2019.00326.

[175] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. R. Martin, and S.-M. Hu, "PCT: Point cloud transformer," Dec. 2020, doi: 10.1007/s41095-021-0229-5.

[176] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[177] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification," Jan. 2018, [Online]. Available: http://arxiv.org/abs/1801.09927.

[178] A. Rao, J. Park, S. Woo, J.-Y. Lee, and O. Aalami, "Studying the Effects of Self-Attention for Medical Image Analysis," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2021, pp. 3409–3418, doi: 10.1109/ICCVW54120.2021.00381.

[179] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Oct. 2019, pp. 1971–1980, doi: 10.1109/ICCVW.2019.00246.

[180] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer, Cham, 2018, pp. 3–19.

[181] N. S. Shaik and T. K. Cherukuri, "Multi-level attention network: application to brain tumor classification," *Signal, Image Video Process.*, vol. 16, no. 3, pp. 817–824, Apr. 2022, doi: 10.1007/s11760-021-02022-0.

[182] S. Bakas *et al.*, "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge," Nov. 2018, [Online]. Available: http://arxiv.org/abs/1811.02629.

[183] J. Cheng, "brain tumor dataset," 2017, doi: 10.6084/m9.figshare.1512427.v5.

[184] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," 2021, [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy.

[185] A. Vaswani *et al.*, “Attention is All you Need,” in *Advances in Neural Information Processing Systems*, 2017, vol. 30, [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[186] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, “Self-Attention Generative Adversarial Networks,” in *Proceedings of the 36th International Conference on Machine Learning*, 2019, vol. 97, pp. 7354–7363, [Online]. Available: https://proceedings.mlr.press/v97/zhang19d.html.

[187] A. Kolesnikov *et al.*, “Big Transfer (BiT): General Visual Representation Learning,” in *Computer Vision -- ECCV 2020*, 2020, pp. 491–507.

[188] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[189] K. He, X. Zhang, S. Ren, and J. Sun, “Identity Mappings in Deep Residual Networks,” in *Computer Vision -- ECCV 2016*, Springer International Publishing, 2016, pp. 630–645.

[190] X. Yang, “An Overview of the Attention Mechanisms in Computer Vision,” *J. Phys. Conf. Ser.*, vol. 1693, no. 1, p. 012173, Dec. 2020, doi: 10.1088/1742-6596/1693/1/012173.

[191] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local Neural Networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 7794–7803, doi: 10.1109/CVPR.2018.00813.

[192] K. Soomro, A. R. Zamir, and M. Shah, “UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild,” *CRCV-TR-12-01*, Nov. 2012.

[193] S. Bakas *et al.*, “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features,” *Sci. Data*, vol. 4, no. 1, p. 170117, Dec. 2017, doi: 10.1038/sdata.2017.117.

[194] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Feb. 2015, [Online]. Available: http://arxiv.org/abs/1502.03167.

[195] Y. Wu and K. He, “Group Normalization,” *Int. J. Comput. Vis.*, vol. 128, no. 3, pp. 742–755, Mar. 2020, doi: 10.1007/s11263-019-01198-w.

[196] S. Qiao, H. Wang, C. Liu, W. Shen, and A. Yuille, “Micro-Batch Training with Batch-Channel Normalization and Weight Standardization,” Mar. 2019, [Online]. Available: http://arxiv.org/abs/1903.10520.

[197] T.-Y. Lin *et al.*, “Microsoft COCO: Common Objects in Context,” in *Computer Vision – ECCV 2014*, 2014, pp. 740–755.

[198] L. M. Zintgraf, T. S. Cohen, T. Adel, and M. Welling, “Visualizing Deep Neural Network Decisions: Prediction Difference Analysis,” Feb. 2017, [Online]. Available: http://arxiv.org/abs/1702.04595.

[199] Geetha M and Rakendu R, “An improved method for segmentation of point cloud using Minimum Spanning Tree,” in *2014 International Conference on Communication and Signal Processing*, Apr. 2014, pp. 833–837, doi: 10.1109/ICCSP.2014.6949960.

[200] M. Kazhdan, M. Bolitho, and H. Hoppe, “Poisson surface reconstruction,” in *Proceedings of the fourth Eurographics symposium on Geometry processing*, 2006, vol. 7.

[201] A. Maiti and D. Chakravarty, “Performance analysis of different surface reconstruction algorithms for 3D reconstruction of outdoor objects from their digital images.,” *Springerplus*, vol. 5, no. 1, p. 932, 2016, doi: 10.1186/s40064-016-2425-9.

[202] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, “Segmenter: Transformer for Semantic Segmentation,” May 2021, [Online]. Available: http://arxiv.org/abs/2105.05633.

[203] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.03385.

[204] C. Szegedy *et al.*, “Going Deeper with Convolutions,” Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.4842.

[205] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning,” Feb. 2016, [Online]. Available: http://arxiv.org/abs/1602.07261.

[206] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to Sequence Learning with Neural Networks,” Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.3215.

[207] D. Bahdanau, K. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.0473.

[208] T. B. Brown *et al.*, “Language Models are Few-Shot Learners,” May 2020, [Online]. Available: http://arxiv.org/abs/2005.14165.

[209] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Oct. 2018, [Online]. Available: http://arxiv.org/abs/1810.04805.

Appendices

Appendix A: Details of Deep-Learning Modules and Architectures

Multi-scale Self-guided Attention Network (MSGA): In an effort to model long-range dependencies and overcome the redundancy of encoder-decoder networks, the multi-scale self-guided attention network has been introduced [109]. Especially if images are processed in 2D, target structures on the medical images generally exhibit intra- and inter-class variability in size, shape, and texture. Traditional CNNs for segmentation provide local feature representations because of their local receptive fields. Local feature representations may result in possible discrepancies between features corresponding to the pixels with the same label since long-range contextual information is not effectively recorded. This could cause intra-class inconsistency, which might eventually affect how well the model performs the pixel classification. To address this issue, an attention mechanism has been incorporated into the MSGA network. First, a multi-scale method is used to capture the global context. The directed attention modules, which are made up of a stack of position and channel self-attention modules, are then supplied with newly learnt characteristics at various scales. The stack of attention modules will assist in gradually filtering out noise and accentuating pertinent information, while the position and channel self-attention modules will aid to adaptively integrate local characteristics with their global dependencies. Features at multiple scales are extracted from a ResNet-101 network and go through a corresponding attention module with position and channel self-attention. Using an encoder-

decoder architecture, the attention is guided to remove any redundant information while highlighting pertinent information. The network is finally trained using a combination of loss of reconstruction ($L_{Rec_{total}}$), loss of guided attention ($L_{G_{total}}$), and loss of segmentation results at different scales ($L_{Seg_{total}}$). Each of total losses are calculated using the sum of losses at each scale. The total loss is calculated as:

$$L_{total} = \alpha L_{Seg_{total}} + \beta L_{G_{total}} + \gamma L_{Rec_{total}} \tag{a.1}$$

where $L_{Seg_{total}} = \sum_{s=0}^{S} L_{Seg_{F'}}^{s} + L_{Seg_{A}}^{s}$, $L_{Rec_{total}} = \sum_{s=0}^{S} L_{Rec}^{s}$, and $L_{G_{total}} = \sum_{s=0}^{S} L_{G}^{s}$.

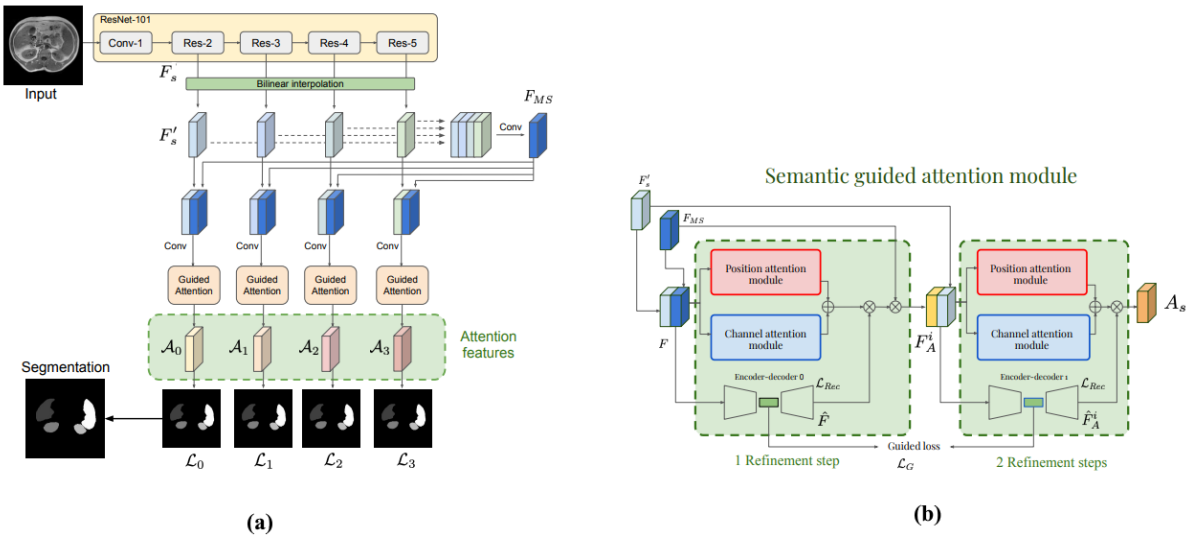

Figure a.1- (a) Overall architecture of the multi-scale self-guided attention network. (b) The schematic of the guided attention module consists of position attention module, channel attention module, and encoder-decoder structure.

Figure a.1(a) shows the overall architecture of MSGA network. Features from different layers of ResNet-101 are extracted resulting in a multi-scale group of features. These features are then interpolated to the same size, concatenated and convolved. The results are then fed to guided attention modules depicted in Figure a.1(b). In this module, position and channel attention modules

with the guidance coming from the parallel encoder-decoder architecture model long range dependencies and discard redundant information.

nnUNet: nnUNet proposes a robust common configuration (fixed parameters) for training UNets for the task of segmentation. The self-adapting feature extraction process offered by the nnUNet adapts to the input data, increasing its robustness against the data's fluctuation and noise. It employs a nested UNet architecture, which is trained using a mix of 2D and 3D input data, to enhance the network's performance on 3D data. Additionally, by joining together many feature maps tailored to different modalities, it expands the UNet architecture to support various modalities. In some instances, it outperforms other state-of-the-art segmentation algorithms, while it has demonstrated a competitive performance on several imaging datasets. Instead of using a new network architecture, nnU-Net achieves its high performance by systematizing the difficult manual model configuration process, which was previously handled either by laborious manual tweaking or purely empirical approaches with practical constraints.

InceptionResNetV2: Residual connections and inception blocks have been central in many advances in computer vision in recent years. The InceptionResNetV2 borrowed the idea from both and proposed combining inception architecture with residual connections. Residual connections were first introduced by He *et. al.* to ease the training of very deep networks [203]. A basic residual block is demonstrated in Figure a.2(a). In the deep residual learning regime, instead of stacking non-linear layers to find the desired underlaying mapping $\mathcal{H}(x)$, nonlinear layers are stacked to fit a residual mapping $\mathrm{F}(x) \coloneqq \mathcal{H}(x) - x$. Theoretically, it is easier to optimize the residual mapping rather than original mapping. In the extreme case, if the optimal solution is identity mapping, it would be easier to push $\mathrm{F}(x)$ to zero rather than learning identity mapping by

stacking nonlinear layers [203]. The idea of inception module is to make networks wider rather than deeper by having filters of multiple size operating on the same level instead of stacking them [204]. Extra 1×1 convolutions were added in the developed framework to limit the number of output channels by dimensionality reduction (Figure a.2(b)). In InceptionResNetV2, residual connections allow for the network to have more layers while inception blocks make the networks wider [205]. Figure a.3 depicts the compressed view of the InceptionResnetV2 network adapted in the developed framework. Notice the residual connection replaced concatenation filter in the last layer of inception blocks.

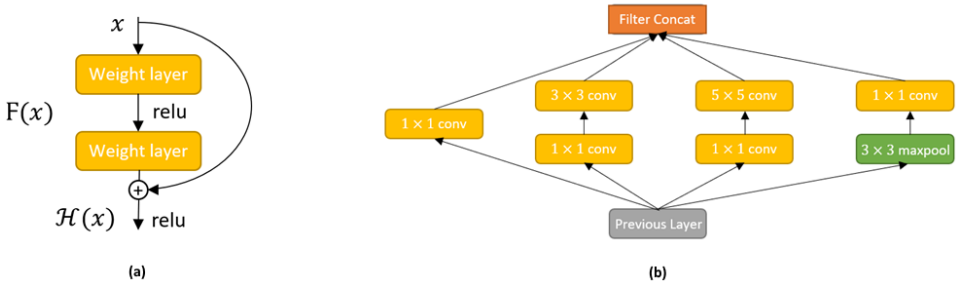

Figure a.2- (a) A residual building block, and (b) Inception module with dimension reduction.

Recurrent Networks: Recurrent neural networks allow information to persist over time through feedback connections [150]. LSTM is a special kind of recurrent neural network powerful at learning long-term dependencies. By having internal mechanism called gate, LSTM can regulate the flow of information. These gates can learn which data in a sequence is important to keep or discard. By doing that, it can pass relevant information down the long chain of sequences to make predictions [150]. Figure a.4(a) demonstrates inside of an LSTM cell. The forget gate decides whether the information should be kept, or it is redundant and should be discarded. In the input gate, the cell state is updated based on the multiplication of tanh output and sigmoid output. The

output gate decides what the next hidden state should be. These gates together control the flow of information from one cell state to the another and help with maintaining long-range dependencies.

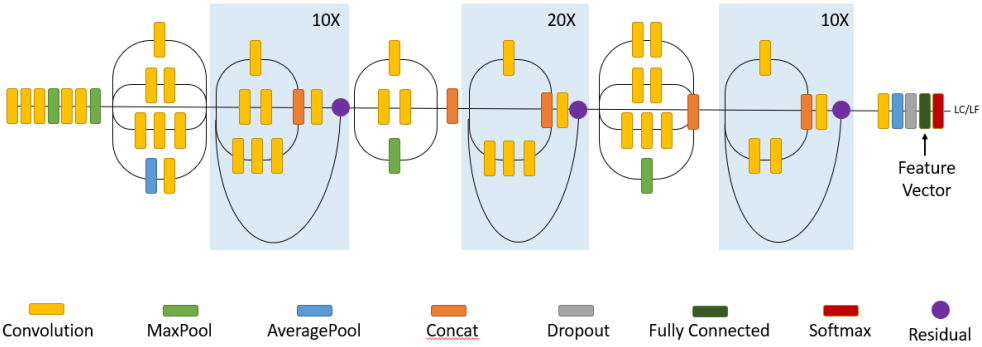

Figure a.3- A compressed view of InceptionResNetV2. X shows number of times an Inception block with residual connection is repeated.

The sequence to sequence (Seq2Seq) models [206] are a special kind of recurrent neural networks and are usually utilized to solve natural language processing problems such as machine translation, image captioning, question answering, etc. Most common Seq2Seq architectures consist of an encoder and a decoder. Both the encoder and decoder are LSTM models. Encoder reads the input sequence and summarizes information into a context vector. The context vector is then fed to the decoder and the decoder tries to make accurate predictions based on the context vector. Figure a.4(b) depicts the system overview of a Seq2Seq model. Instead of relying solely on the hidden state of the last LSTM cell as the context vector, the linear sum of the hidden states from all LSTM cells in the encoder could be calculated as a context vector that brings about a seq2seq model with attention [207].

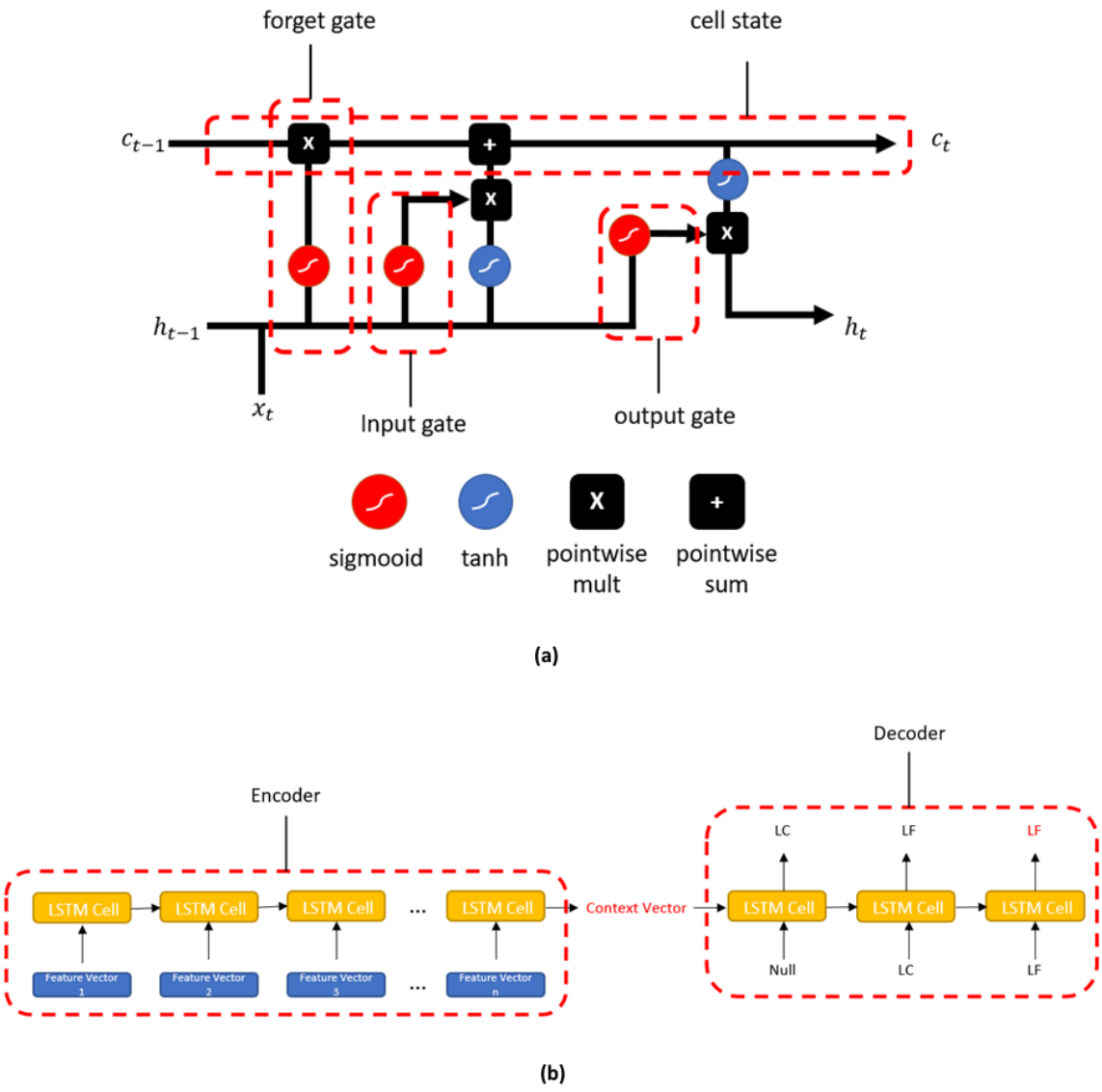

Figure a.4- (a) Inside of an LSTM cell with *input gate*, *output gate*, *forget gate*, and *cell state*. $h_{t-1}, c_{t-1}, x_t. h_t, c_t$ are hidden state of previous layer, cell state of previous layer, input, hidden state of current layer and cell state of current layer, respectively, and (b) System overview of a Seq2Seq model with encoder and decoder.

Transformers: A Transformer, applied in Chapter 4, is a novel architecture which obviates the need for recurrence and relies entirely on an attention mechanism to find global dependencies between input and output. Initially introduced by Vaswani et al. [185] in the paper "Attention is All You Need", Transformer achieved state-of-the-art performance compared to recurrent

networks while being more parallelized and requiring significantly less time to train. Similar to Seq2Seq, the Transformer employs and encoder-decoder style, however, eliminating recurrence and replacing it with attention mechanisms allows for substantially greater parallelization than RNNs and CNNs. The Transformer was introduced as a network for sequence-to-sequence translation but since it's proposal, a huge variety of networks has been introduced based solely on encoder or decoder part of Transformers. GPT-3 [208] is an example of a network which is built on the decoder part of the transformer that produces human-like text with very high quality. The encoder part of Transformer on the other hand has been utilized in architectures such as BERT [209] for question-answering and text classification tasks.

As discussed earlier, Transformer has an encoder-decoder architecture and uses self-attention mechanism to infer dependencies. Each encoder or decoder consists of modules that contain feed-forward and attention layers. Figure a.5(a) demonstrates an architecture of Transformers. With (key, value) vector pairings, the attention layer employs a trainable associative memory. From a series of *N* inputs, the query and key matrices are generated and packed into the following matrices:

$$X \in \mathbb{R}^{N\times D};\ Q = XW_Q \in \mathbb{R}^{N\times D};\ \ K = XW_K \in \mathbb{R}^{N\times D} \tag{a.2}$$

where X is the sequence of *N* inputs with dimension *D*, *Q* and *K* are the query and key matrices, and W_Q and W_K are linear transformations with trainable parameters. The output of the attention is a weighted sum of the N-value matrix $V \in \mathbb{R}^{N\times D}$ and is calculated by the following formula:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{D}}\right)V \tag{a.3}$$

Instead of offering a single attention head, Vaswani et al. [185] offered multi-head attention, which entails applying h self-attention functions to the input.

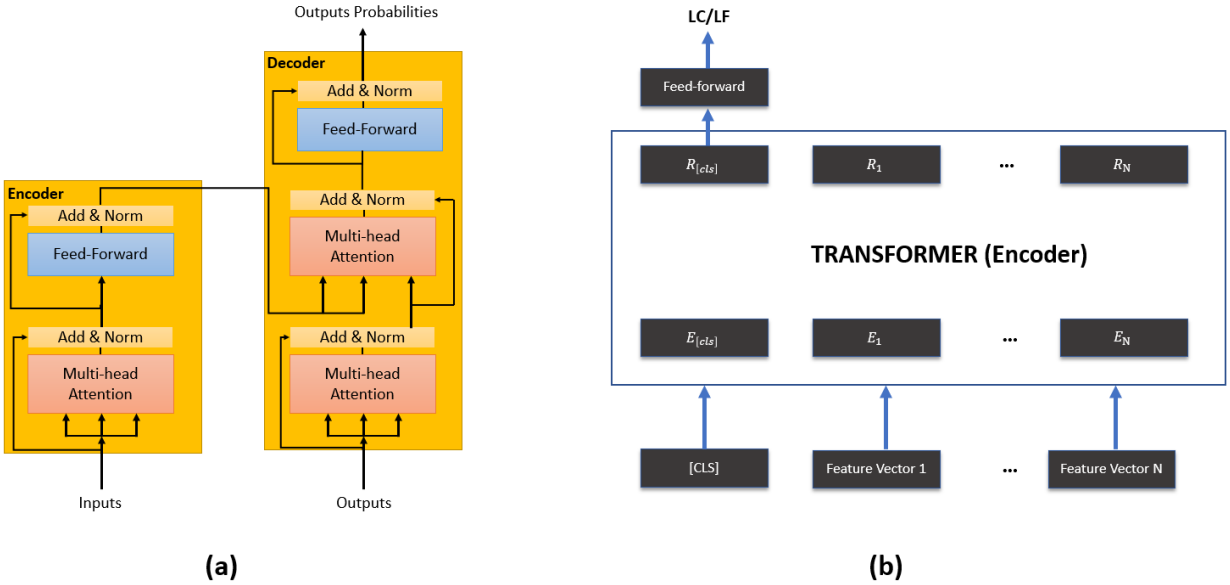

Figure a.5- (a) The architecture of a Transformer for sequence-to-sequence tasks. In practice, a stack of multiple encoders and decoders are used. Each encoder consists of multi-head attention later, residual connections, and feed-forward layer to prepare the input of next encoder/decoder block. (b) Classification using Encoder architecture of Transformer. A [cls] token is added to the input series and in the output of encoder, only the [cls] part is used for classification.

3D Residual Network: The 3D residual network applied in Chapter 5 is an extension of residual networks or ResNets with 3D components instead of 2D ones [49]. Historically, the trend of making networks deeper to increase their modeling capabilities was hindered by vanishing gradient. After numerous applications of the chain rule, the gradients from which the loss function is derived simply drop to zero when the network is too deep. As a consequence, the weights at higher layers never update their values, hence no learning takes place. By introducing skip connections, gradients can flow backward from deeper layers to initial filters directly via these connections. Skip connections enable the network to easily model the identity function, where the output of a function becomes its input. More specifically, instead of learning the output function

H(x) = f(x), the output is changed into *H(x) = f(x) + x*. Simply, by setting *f(x)=0*, *H(x)* becomes the identity function. The most important part of the architecture is the stack of residual blocks and skip connections to preserve the gradient. Further to preserving gradient, another reason that skip connections prove useful is the fact that the learned features correlate to lower semantic information retrieved from the input in prior levels. That information become too abstract if the skip connections are not utilized in this architecture.

Convolutional Block Attention Module (CBAM): When it comes to feed-forward convolutional neural networks, CBAM is a simple yet effective attention module [180]. Given an intermediate feature map, the module progressively infers attention maps along two different dimensions, *i.e.*, channel and spatial, and then multiplies the attention maps by the input feature tensors to perform adaptive feature refinement on the intermediate feature tensors. The fact that CBAM is a lightweight and universal module means that it can be smoothly integrated into any CNN architecture with minimal overhead and that it is trainable from start to finish alongside the base CNNs.

As mentioned above, CBAM consists of two sequential separate attention mechanisms, channel attention, and spatial attention. Because each channel of a feature tensor may be considered as a feature detector, channel attention is focused on 'what' is significant in the context of an input image when using feature tensors. Channel attention begins by aggregating spatial information from the feature tensor using both average-pooling and max-pooling processes, resulting in the generation of two separate spatial context descriptors for each feature map: F_{avg}^{c} and F_{max}^{c}, which denote average-pooled features and max-pooled features, respectively. Afterwards, both descriptors are forwarded to a shared network, which generates the channel

attention map $M_c \in R^{1\times1\times1\times C}$, where C is the number of channels. The shared network is made up of a multi-layer perceptron (MLP) with one hidden layer. Following the application of the shared network to each descriptor, the resulting feature tensors are combined by applying element-wise summing to form a single feature tensor. To summarize, the channel attention map is computed using Equation (a.4), where σ denotes the sigmoid function.

$$M_c(F) = \sigma(MLP\big(AvgPool(F)\big) + MLP\big(MaxPool(F)\big)) \quad \text{(a.4)}$$

The next step is to compute spatial attention. In order to build a spatial attention map, the spatial attention module uses the inter-spatial relationship between features. At the other end of the spectrum from channel attention, spatial attention focuses on 'where' informative features are located in the image, and it is considered a complement to the channel attention. Using average-pooling and max-pooling operations along the channel axis, the spatial attention map can be calculated. A convolution layer is applied to the concatenated feature descriptor in order to construct a spatial attention map based on it, *i.e.*, $M_s(F) \in R^{X\times Y\times Z}$. The channel information in the feature tensor is aggregated via the use of two pooling processes, resulting in the generation of two 3D maps: $F^s_{avg} \in R^{X\times Y\times Z\times1}$ and $F^s_{max} \in R^{X\times Y\times Z\times1}$, that denote the average-pooled features over the channel, and the max-pooled features, respectively. Formally, the spatial attention map is computed using Equation (a.5)

$$M_s(F) = \sigma(f^{7\times7\times7}\big(\big[F^s_{avg}; F^s_{max}\big]\big)) \quad \text{(a.5)}$$

The channel and spatial modules are applied to the intermediate feature maps sequentially and output the refined features. In our proposed architecture, the CBAM block was added right before

the 3D average pooling layer to filter out irrelevant information and focus on important details for classification.

The Self-attention Module: The self-attention module applied in Chapter 5 and is defined as a tensor mapping that transforms the input tensor to a query, a key, and a value tensor. The key and value are learned features extracted by convolution blocks, and the query determines which values to focus on for the learning process. The role of 3D convolution blocks ($1 \times 1 \times 1$ convolutions) before the key, query, and value is to perform linear transformations on the input feature tensors. The key, query, and value vectors are denoted by $k(x)$, $q(x)$, and $v(x)$ and are calculated as $k(x) = W_k x$, $q(x) = W_q x$, and $v(x) = W_v x$, where W_k, W_q, and W_v are all $1 \times 1 \times 1$ convolution filters and x is the feature tensor coming from the previous layer. After reshaping to permit matrix multiplications, $k(x), q(x)$ and $v(x) \in R^{N \times C}$, where C is the number of channels and $N = X \times Y \times Z$ is the number of elements in the input feature tensor. The self-attention map α can be calculated using Equation (a.6).

$$\alpha_{i,j} = \frac{\exp(q(x_i)k(x_j)^T)}{\sum_{i=1}^{n} \exp(q(x_i)k(x_j)^T)}. \quad \text{(a.6)}$$

In this equation, $a_{i,j}$ is the correlation between the feature element i and other feature elements, and j is the index of corresponding output position. The output of the attention branch is $o = (o_1, o_2, \dots, o_N)^T \in R^{N \times C}$, where o_j is calculated using Equation (a.7).

$$o_j = \sum_{i=1}^{N} a_{i,j} v(x_i). \quad \text{(a.7)}$$

Finally, a $1 \times 1 \times 1$ convolution (W_o) is applied to the reshaped output ($\in \mathbb{R}^{X \times Y \times Z \times C}$) to keep the number of channels consistent between the input and output feature tensors of the attention layer ($Output = W_o o$). The 3D self-attention module is added to the architecture after each residual block to ensure deriving long-range dependencies along with the convolution layers that mostly capture local features and dependencies.