# Deep Learning-based Tumour Heterogeneity Analysis with Multiparametric Magnetic Resonance Imaging

THE UNIVERSITY OF SYDNEY

Yue Xia

[redacted]

Supervisor: Professor Jinman Kim

Associate Supervisor: Dr Euijoon Ahn

A thesis submitted in fulfilment of the requirements

for the degree of Master of Philosophy

March 2025

# Statement of Originality

I certify that:

To the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes;

I understand that failure to comply with the Student Plagiarism: Academic Board Policy: Academic Dishonesty and Plagiarism can lead to the University commencing proceedings against me for potential student misconduct under the Thesis and Examination of Higher Degrees by Research Procedures 2024.

This Work is substantially my own, and to the extent that any part of this Work is not my own, I have indicated that it is not my own by Acknowledging the Source of that part or those parts of the Work.

Name: Yue Xia

Signature: Date: 31 March 2025

# Authorship Attribution Statement

Chapter 3 of this thesis is based on a publication Y. Xia, Y. Yuan, E. Ahn and J. Kim, "Multi-Phase and Hierarchical Unsupervised Learning Framework for Glioblastoma Sub-Region Segmentation in MRI Sequences," 2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Perth, Australia, 2024. I designed the method, conducted theexperiments, analysed the data and drafted the manuscript.

Chapter 4 of this thesis is based on a publication Y. Xia, Y. Yuan, E. Ahn and J. Kim, " Heterogeneity-Aware Deep Learning for Tumour Classification in Multi-Parametric MRI: A Dual-Stream Framework with Unsupervised Sub-Region Alignment,", in preparation. I designed the method, conducted the experiments, analysed the data and drafted the manuscript.

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Yue Xia Date: 30 March 2025

As supervisor for the candidature upon which thesis is based, I can confirm that the authorship attribution statements above are correct.

Professor Jinman Kim Date: 31 March 2025

# Abstract

Tumour heterogeneity, characterised by spatial and temporal variations in cellular morphology and molecular profiles within the tumour tissues, impacts oncology diagnosis, prognosis, and treatment outcomes. Medical image analysis, particularly through multiparametric magnetic resonance imaging (mpMRI), provides a non-invasive approach to capture these complex heterogeneity patterns, enabling a more objective and comprehensive assessment of tumour biology. mpMRI supports more personalised treatment planning and reduces the need for invasive diagnostic procedures, ultimately benefiting patient care. However, the intrinsic variability in heterogenous tumours poses challenges in accurately delineating tumour regions and predicting therapeutic responses, often leading to inconsistent clinical interpretations. Because of this, labelling and analysing heterogeneity sub-regions in mpMRI are time-consuming tasks and requires experienced expertise.

Recent advances in computational image analysis techniques such as radiomics and deep learning models has demonstrated capability to automate the mpMRI analysis and thereby reducing clinician's workload. Radiomics approaches extract handcrafted features from tumour sub-regions to summarise tumour heterogeneity. However, handcrafted features strongly rely on prior knowledge of imaging patterns correlated with tumour biology and patient outcomes, and also requires manual annotation of the sub-regions. Although deep learning models excel at extracting detailed pixel-level heterogeneity features without explicit human intervention, most approaches do not incorporate tumour sub-region information, thereby limiting their ability to capture tissue heterogeneity and the complex spatial relationships between adjacent sub-region structures.

This thesis introduces a novel deep-learning framework that addresses the challenges of tumour heterogeneity in mpMRI modalities to enhance tumour heterogeneity analysis in medical image and its use in downstream tasks including image segmentation and classification. The framework comprises of two main components. First component is an unsupervised semantic segmentation method developed to delineate tumour sub-regions automatically. This method effectively captures the intrinsic structure of heterogeneous tumours by leveraging a multi-phase training strategy that combines coarse segmentation with refined, self-supervised learning enhanced by sparse spatial continuity and context-based hierarchical loss functions. Second, we propose a heterogeneity-aware deep learning method for tumour classification that integrates machine-generated sub-region labels with dual-stream feature extraction for both local heterogeneity and global image information. A learnable alignment module is employed to standardise sub-region labels across different imaging modalities, enabling the extraction of both local heterogeneity features and global contextual information. The proposed framework was demonstrated on two mpMRI datasets: (i) brain tumour with BraTS2021 for sub-region segmentation and, (ii) liver tumour with LLD-MMRI2023 for cancer classificaiton, in comparison with the state-of-the-art methods.

# Acknowledgements

I would like to express my deepest gratitude to Professor Jinman Kim. His visionary leadership and constant encouragement have been pivotal to my academic growth. His insightful feedback and innovative ideas have helped shape the direction of my research.

I am deeply thankful to Dr Euijoon Ahn for his critical insights and detailed guidance. His willingness to share his expertise and provide constructive feedback has greatly enhanced the quality of my work. The clarity and depth of his advice have been instrumental in refining my research.

I wish to express my heartfelt appreciation to Dr. Xia Tian. His technical support and practical advice were invaluable in overcoming the hurdles encountered during this project. Dr Tian's commitment to excellence and his readiness to assist whenever needed have greatly contributed to the successful completion of this thesis.

I also immensely thankful to all friends and mentors from our research group. Thank you all for being there through the tough journey in research and I won't be finishing this without supports from each one of you. I hope that you continue to experience success and fulfillment in your own endeavours!

Lastly, I would like to express my thanks to my family and Xiaoyu Liang, whose love and encouragement have been my foundation. I would not be possible to start such a wonderful journey without you. For this, I am grateful.

# Contents

# List of Figures

# List of Tables

# Thesis Publications

The following publications have been produced or in preparation over the course of the candidature:

Accepted paper:

Y. Xia, Y. Yuan, E. Ahn and J. Kim, "Multi-Phase and Hierarchical Unsupervised Learning Framework for Glioblastoma Sub-Region Segmentation in MRI Sequences," *2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Perth, Australia, 2024.

Paper under preparation:

Y. Xia, Y. Yuan, E. Ahn and J. Kim, " Heterogeneity-Aware Deep Learning for Tumour Classification in Multi-Parametric MRI: A Dual-Stream Framework with Unsupervised Sub-Region Alignment,". For submission to *IEEE Journal of Biomedical and Health Informatics*.

# Chapter 1. Introduction

## 1.1. Motivation

Tumour heterogeneity represents a significant clinical and research challenge due to its substantial influence on cancer progression, therapeutic response, and patient outcomes [1]. It encompasses variations observed among different tumours (inter-tumour heterogeneity) and variations within individual tumours (intra-tumour heterogeneity). Such heterogeneity is characterised by spatial and temporal differences in cellular morphology, molecular profiles, and phenotypic behaviours, resulting in diverse tumour progression patterns and clinical manifestations [2]. These variations significantly influence patient responses to treatment, underscoring the necessity for accurate, comprehensive, and reproducible assessments of tumour heterogeneity to guide clinical decisions and personalised care strategies.

Medical imaging techniques, such as computed tomography (CT) and magnetic resonance imaging (MRI), have been widely utilised for the early detection and diagnosis of cancer. They are a non-invasive method to depict tumour heterogeneity information. Among the imaging modalities, multiparametric MRI (mpMRI) has been widely studied for its ability to assess tumour tissues and capture tumour heterogeneity information [3]. It utilises magnetic fields and radiofrequency pulses to generate images of patient organs. mpMRI utilises a range of parameters, including various radiofrequency pulses and gradient settings, to create distinct intensity distributions that effectively differentiate tissue types and reveal detailed information on tissue-specific heterogeneity. Due to these advantages, mpMRI is extensively used for tumour-related clinical decisions, including screening, diagnosis, and prognosis evaluation, across various tumour sites, such as the liver and the brain structures [4, 5].

In mpMRI, tumour heterogeneity manifests as varied intensities, shapes, and textures across tumour regions, reflecting underlying molecular, cellular, and histopathological variations. For example, these observable characteristics allow the differentiation and delineation of clinically relevant tumour sub-regions, such as necrotic cores, enhancing tumour regions, and peritumoral oedema in glioblastoma (GBM) brain cancer tumours [6, 7]. Clinicians rely on this information to inform diagnosis, establish prognosis, and determine therapeutic strategies, including genomics-guided medication selection [8]. These clinical evaluations, however, are done manually which is time consuming, subjective to the expert interpretation, and requiring highly specialised skills. This subjective process introduces variability, impedes standardisation, and may reduce diagnostic consistency and reliability across clinical settings.

Radiomics represents a computational method developed to quantitatively characterise and analyse tumour sub-region heterogeneity within medical images by extracting high-dimensional quantitative imaging features. Radiomics pipeline in sub-region analysis typically derive manual or semi-automatic imaging biomarkers (features) from tumour sub-regions previously defined by clinicians. Subsequently, machine learning models use these extracted features to gain insights into tumour characteristics, predict clinical outcomes, or determine molecular profiles [9]. Recent studies, such as those by Li et al. [10] and Zhang et al. [11], have demonstrated the potential of radiomics features extracted from tumour sub-regions in determining clinically relevant genetic information, including the O6-methylguanine DNA methyltransferase (MGMT) methylation status from mpMRI. However, despite these promising results, reliance on manual or semi-automatic tumour sub-region annotations poses significant limitations regarding scalability, reproducibility, and generalisation across diverse clinical scenarios.

Recent advances in deep learning offer a compelling solution by providing end-to-end frameworks that can learn hierarchical, data-driven features directly from medical imaging data. Deep learning methods, such as convolutional neural networks (CNNs), have already demonstrated state-of-the-art performance in clinical decision-support tasks for medical imaging [12, 13]. By automatically extracting and fusing information across various feature scales throughout the convolution process, these models have the potential to capture subtle tissue heterogeneity and complex spatial interdependencies that are often missed by traditional radiomics pipelines. However, based on literature search, there are no methods that explored the analysis of sub-regions to aid in detecting localised tumour heterogeneity and further enhancing diagnostic accuracy.

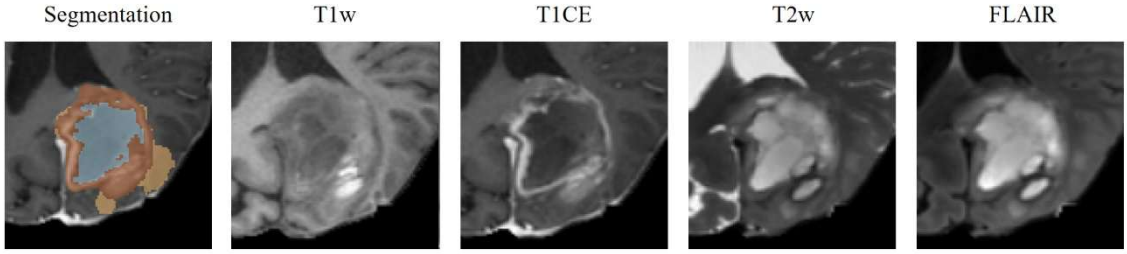


**Figure 1.1.** The four common MRI modalities used in the brain; the images are acquired from patients diagnosed with GBM tumours. The blue-shaded segmentation mask indicates the necrosis area, brown represents the enhanced tumour, and orange represents the oedema region.

One challenge in tumour subregion-based image analysis is that the subregion labels are either manually annotated or semi-automatic e.g., manually labelling the machine learning outputs such as with cluster analysis segmentation method. Due to the complexity in manual labelling, tumour subregions are typically segmented in coarse regions e.g., the Brain Tumor Segmentation Challenge 2021 (BraTS2021) [53] datasets that contain subregion labels for GBM comprise of three number of labels illustrated in Figure 1.1, which although captures the main subregions, do not offer finer details that may further differentiate the tumour regions.

Another challenge is that despite the promising capabilities of radiomics and deep learning methods, they are not designed to capture both the local sub-region-based and global heterogeneity information simultaneously. Current radiomics pipelines primarily rely on handcrafted, predefined features derived from clinician-defined tumour regions, frequently neglecting intricate local interactions between neighbouring tumour sub-regions and lacking integration of global contextual information [14]. Conversely, although deep learning approaches excel at extracting detailed pixel-level heterogeneity features without explicit human intervention, most of these methods do not explicitly leverage clinically relevant sub-region analysis that they are trained in an end-to-end manner to directly mapping input images to output labels [12, 13]. These approaches are usually trained in an end-to-end manner i.e. from image direct to label. This limits their potential to fully capture tissue heterogeneity for each tumour sub-region and complex spatial relationships between adjacent anatomical structures. Therefore, these limitations underscore the need for end-to-end frameworks that capture detailed local sub-region heterogeneity and comprehensive global tumour context for improved diagnostic and therapeutic outcomes.

## **1.2.** Aims and Objectives

In this thesis, a new framework is introduced to address the identified challenges in tumour heterogeneity analysis through (i) leveraging a context-based unsupervised semantic segmentation for GBM and, (ii) adapting the sub-region analysis approach from radiomics pipeline in a deep learning method for tumour sub-region analysis. The proposed framework leverages machine-generated sub-region labels to obtain heterogeneity information for determining tumour inter- or intra-heterogeneity and also expands upon sub-region analysis approaches in radiomics by adapting deep learning

techniques for tumour classification. The research in this thesis addresses the following two aims:

1. Develop an unsupervised semantic segmentation method designed explicitly for GBM. A novel method is proposed that introduces multi-phase training and hierarchical loss based on the intrinsic structure of GBM. The proposed method mitigates the need for human labelling of GBM sub-regions and thus potentially improving efficiency in downstream tasks involving heterogeneity analysis.

2. Adapting a deep learning method that integrates unsupervised segmentation of sub-region labels for tumour classification and MGMT status prediction. A novel dual-stream feature extraction method is proposed to effectively captures local sub-region heterogeneity through local feature extraction stream while integrating global contextual information via the global stream. A learnable alignment for sub-regions is also acquired from the unsupervised segmentation. This enabled the method to leverage the heterogeneity information captured from computer-generated sub-region labels, and demonstrating effective enhancement in tumour disease classification performance.

## 1.3. Thesis Structure

The remainder of this thesis is structured as follows:

Chapter 2 provides the essential background knowledge necessary for readers to understand the thesis. It provides an overview of medical images, particularly mpMRI images, radiomics methods in sub-region analysis, and current state-of-the-art deep learning.

Chapters 3 and 4 provide a detailed contribution to this thesis. Chapter 3 presents the unsupervised sub-region segmentation method designed for GBM. The spatial-heterogeneity insights uncovered in Chapter 3 provide the conceptual foundation for the heterogeneity-aware classification strategy developed in Chapter 4. Chapter 4 describes the heterogeneity-based deep learning framework that leverages sub-regions for tumour classification.

Chapter 6 concludes this thesis's contribution and discusses the future works in tumour heterogeneity analysis.

# Chapter 2. Background

This chapter begins by introducing the fundamentals of medical imaging, focusing on their representation in computational approaches and processing techniques. It includes a detailed overview of mpMRI modalities, emphasising their role in characterising tissue heterogeneity composition and subtle physiological features. The chapter then focuses on two principal computational methods: radiomics and deep learning. For radiomics, the discussion encompasses handcrafted feature extraction and the general radiomics pipeline, with particular attention given to pipeline that utilise sub-region analysis to capture the intrinsic variability within tumours. For deep learning, both unsupervised segmentation frameworks and supervised classification frameworks are examined, showcasing their capabilities in automatically learning hierarchical features and improving the detection and characterisation of tumour regions. These sections provide a comprehensive foundation for the audience to understand the two chapters that follow, which introduce the main contributions of this thesis.

## 2.1. Medical Image

Medical images are digital representations of anatomical structures or physiological processes within the human body. They are composed of picture elements, commonly referred to as pixels in 2D images and voxels in 3D images. A pixel (short for “picture element”) represents the smallest unit in a 2D digital image and holds a single intensity value. In volumetric imaging, such as MRI, images are acquired in stacks of 2D slices, where each pixel extends into the third dimension to form a voxel (volume element), thereby capturing spatial information in three dimensions [15].

Spatial resolution refers to the physical size of each pixel or voxel, typically measured in millimetres. It defines the smallest detail that can be distinguished in an image. For instance, a voxel with a spatial resolution of 1 mm × 1 mm × 1 mm represents a cubic millimetre of tissue. Higher spatial resolution enables better discrimination of fine structural details within the tissue [16].

In a standard RGB image, the intensity of each channel was determined by the amount of corresponding colour. In medical imaging, the context of the intensity depends on the imaging technique used. Each imaging technique, based on its unique physical principles and parameters used, will produce distinct information about the organ. For instance, while X-ray and CT intensities reflect variations in tissue density and composition, ultrasound intensities depend on acoustic impedance and MRI intensities are influenced by proton density and relaxation times [17].

Contrast resolution, on the other hand, denotes the ability to distinguish differences in intensity values between adjacent tissue types or regions. Differentiating between normal and pathological tissues is crucial in medical imaging. In mpMRI, contrast resolution is manipulated using different imaging sequences to highlight specific tissue characteristics [18]. The following subsection provides an overview of the modality used in mpMRI for chapters 3 and 4.

### 2.1.1. Multiparametric MRI (mpMRI) modalities

mpMRI combines multiple imaging modalities, each exploiting different contrast mechanisms, to comprehensively characterise tumours. These sequences provide complementary information about tissue properties and are invaluable in delineating heterogeneous tumour sub-regions (e.g., distinguishing enhancing tumour core,

necrotic areas, and surrounding oedema) [6]. Below, we summarised the used mpMRI modalities and their imaging characteristics.

Common modalities used for the brain tumour are visualised in Figure 1.1. The description of each modality is as follows [19]:

1. T1-weighted (T1w): T1-weighted images are acquired using short repetition time (TR) and echo time (TE) to emphasise differences in longitudinal relaxation. They render fat- and protein-rich tissues as bright and water-rich regions as dark, providing high anatomical detail and baseline tumour characterisation.
2. Contrast-Enhanced T1-weighted (T1CE): T1CE is performed after administering a gadolinium-based contrast agent that shortens T1 relaxation times. It delineates tumour boundaries by brightening regions with abnormal vascularity or blood–brain barrier disruption, highlighting viable tumour components.
3. T2-weighted (T2w): T2-weighted imaging uses long TR and TE to accentuate differences in transverse relaxation, making it highly sensitive to water content. It depicts water-rich tissues (such as oedema and cysts) as bright, while fatty and denser tissues appear darker, thereby revealing fluid-related abnormalities.
4. T2- Fluid Attenuated Inversion Recovery (FLAIR): FLAIR is a specialised T2 sequence that uses an inversion recovery pulse to null the signal from free fluids, such as cerebrospinal fluid. This technique highlights oedema, gliosis, or tumour infiltration as bright signals against a dark background, thereby enhancing lesion visibility.

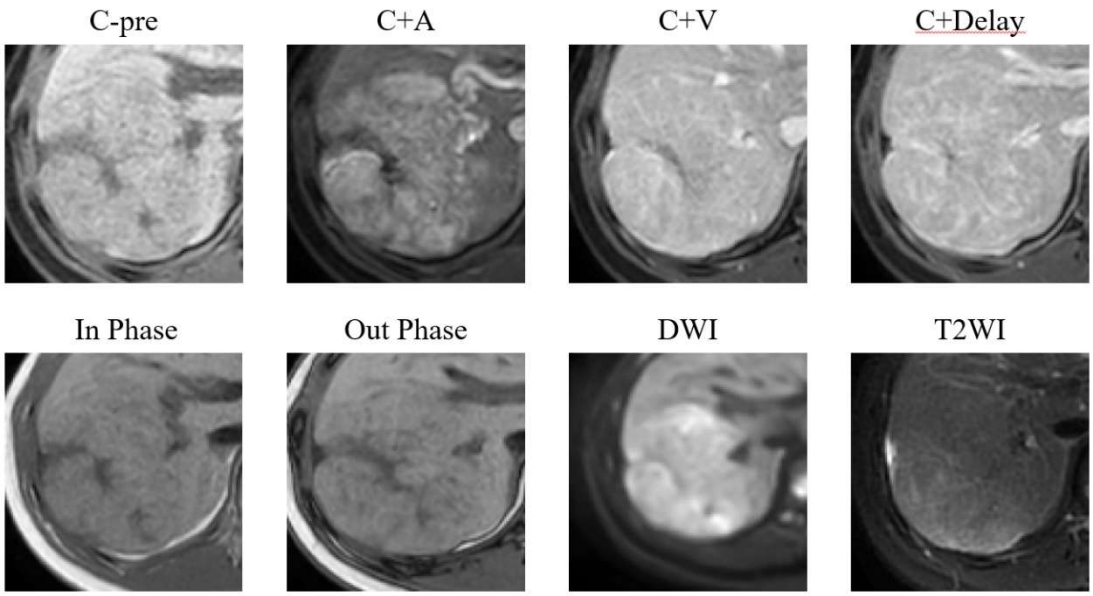


**Figure 2.1.** The eight common MRI modalities used in liver; the images are acquired from patients diagnosed with Hepatocellular carcinoma.

.

Common modalities used for the liver structure are visualised in Figure 2.2. The description of each modality is as follows [20]:

1. Pre-Contrast (C-pre): Pre-contrast imaging is a baseline T1-weighted scan performed before the administration of any contrast agent. It establishes the inherent signal characteristics of the tissue, serving as a reference to differentiate native properties from contrast-induced changes.
2. Arterial Phase Post-Contrast (C+A): Arterial phase imaging is conducted immediately after contrast injection to capture the early passage of the contrast agent through the arterial system. It accentuates highly vascularised tumour regions by highlighting rapid contrast uptake, indicative of aggressive angiogenesis.
3. Venous Phase Post-Contrast (C+V): Venous phase imaging is acquired typically 60–90 seconds post-injection, once the contrast has equilibrated in the venous system. It displays the full extent of tumour enhancement by reflecting intermediate perfusion and extracellular contrast distribution.

4. Delayed Phase Post-Contrast (C+Delay): Delayed phase imaging is obtained several minutes after contrast injection, allowing time for the contrast agent to equilibrate between vascular and interstitial spaces. It highlights areas with persistent contrast retention, differentiating leaky tumour regions from those undergoing washout and aiding in distinguishing active tumour from treatment effects.
5. In-Phase and Out-of-Phase Imaging: In-phase and out-of-phase imaging are paired T1 gradient-echo sequences acquired at different echo times to exploit the chemical shift between fat and water. They reveal mixed tissue composition by showing signal cancellation in voxels containing both fat and water, thereby detecting microscopic fat deposits.
6. Diffusion-Weighted Imaging (DWI): DWI applies strong diffusion-sensitizing gradients to capture the motion of water molecules within tissues. It differentiates tissue characteristics by showing restricted diffusion (indicative of high cellularity) as a bright signal and free diffusion (in necrotic or cystic areas) as a darker signal, with corresponding ADC maps.
7. T2-weighted (T2w), as mentioned above.

### 2.1.2. Medical Image Segmentation

Medical image segmentation involves partitioning a medical image into regions, which usually requires a pixel-level label to represent a region of interest [21]. This is crucial for tumour-related diseases, as it can highlight tumour regions and distinguish them from surrounding normal tissues. For heterogeneous tumours like GBM, segmentation also requires the separation of tumour sub-regions by differentiable tissue appearance in the image [6]. However, the manual labelling process is time-consuming and influenced by individual biases and competence levels [22].

By leveraging computerised medical analysis algorithms, the segmentation process can be facilitated by computers. Current computerised medical image segmentation can be divided into three approaches based on employed techniques: (i) supervised segmentation that uses manual labels to train the algorithms, (ii) unsupervised segmentation that employs algorithms to distinguish intensity present in the image and generate ROI without human intervention and labels and (iii) semi-automated segmentation that employs pixel level labels acquired from unsupervised segmentation then requires human to identify the class of each region. Semi-automated segmentation is typically used in sub-region-related segmentations and analyses, which will be introduced in the radiomics section.

In supervised segmentations, deep learning has demonstrated superior performance in various diseases compared to other methods [23]. Ronneberger et al. [24] proposed U-net for biomedical image segmentation and utilises a down-sampling path to capture context and a symmetric up-sampling path that maintains spatial relationships between input (image) and output (segmentation labels). Furthermore, a recently proposed framework, transformer, adapts the same dual-path manner and employs attention mechanisms to achieve a better global reception field, which improves segmentation results [25]. However, the majority of existing studies employ supervised learning approaches that depend on large amounts of pixel-level annotated datasets. This may hinder the generalisation capabilities and overall performance of supervised learning approaches.

For unsupervised segmentation, clustering algorithms could be used to generate unannotated sub-region labels. One conventional machine learning approach for clustering algorithms is k-means [26]. It leverages the intensity distribution present in the images and partitions them into regions by a pre-defined number of centres. However, k-means uses intensity values only and it would consider unconnected regions as the same type. A deep learning-based clustering method incorporates a

spatial constraint to facilitate the separation of unconnected regions. It enhances the continuity of segmented regions by leveraging the spatial relationships present in the image. Some radiomics methods for sub-region analysis aim to eliminate manual labels; they would use a unsupervised segmentation approach to generate sub-regions in tumours.

### 2.1.3. Medical image classification

Medical image classification involves labelling categories to a medical image based on the given image. It is essential for clinical diagnosis, treatment planning and prognosis evaluation. In this thesis, we focus on tumour heterogeneity-related classification. One is inter-tumour heterogeneity, which characterises the difference between different tumours and aims to distinguish their type. The other one is intra-tumour heterogeneity, which characterises the differences among patients with the same tumour and provides insights into genetic biomarkers that can then be used to aid personalised treatment. Genetic biomarkers have become essential tools for stratifying patients into meaningful subgroups and predicting treatment efficacy and outcomes. For example, MGM) status in GBM influences tumour cell response to alkylating chemotherapy agents, necessitating alternative treatment strategies to enhance patient prognosis and survival rates [27, 28]. Consequently, incorporating genetic biomarkers into clinical practice is crucial for optimising therapeutic outcomes, individualising patient care, and improving clinical prognosis.

In the clinic, the acquisition of genetic information, typically achieved through tissue biopsy procedures, is vital for individualised treatment planning and is directly associated with improved clinical outcomes [29]. However, biopsy procedures present several significant challenges. They are invasive and costly, frequently associated with patient discomfort and potential complications [30, 31]. Additionally, they are subject

to inaccuracies arising from the spatial distribution of genetic heterogeneity within the tumour, often requiring repeated procedures to characterise tumour genetics [32] accurately. This limitation highlights the critical need for alternative or complementary non-invasive assessment tools.

The conventional computational approach to medical image classification typically involves i) image feature extraction, ii) feature selection and iii) classification [33]. In the feature extraction phase, quantitative imaging characteristics that are informative for the target labels are identified using either handcrafted features or deep learning approaches that automatically learn useful representations from the training data. Next, feature selection involves choosing a subset of these extracted features to reduce the dimensionality of the data, thereby improving the efficiency and accuracy of the machine learning algorithm. Finally, during the classification stage, the selected features are fed into a machine learning model that assigns a label or category to the image.

In contrast, deep learning-based approaches offer an end-to-end solution for medical image classification by automatically learning image features that are directly relevant to the associated labels, eliminating the need for explicit feature engineering [34]. As a result, these deep learning techniques have achieved state-of-the-art performance in various tasks.

## 2.2. Radiomics

Radiomics is a field of medical imaging analysis that extracts various quantitative features from radiographic images to characterise tumour phenotype in a non-invasive manner. Conventional radiomic features are predefined features extracted from medical images, such as voxel volume, surface area, Gray Level Co-occurrence Matrix (GLCM),

and Gray Level Size Zone Matrix (GLSZM), to gain high-level semantic features in shape, volume, and texture [35]. Radiomic features can capture intratumor heterogeneity in a quantitative way and are easily used for further classification and other statistical-based analyses. A significant challenge in glioblastoma analysis is accounting for tumour heterogeneity. By converting medical images into mineable data, radiomics enables the correlation of imaging features with underlying histopathology, genomics, and clinical outcomes. In oncology, radiomic analyses have shown promise for tumour detection, prognosis, and prediction of treatment response, often serving as imaging biomarkers of disease aggressiveness [36, 37]. Heterogeneous regions in a tumour, such as necrosis, oedema, and enhancing regions, can significantly influence prediction outcomes. Several studies have demonstrated the correlation between these regions and MGMT methylation status, suggesting the need for models that capture texture information from these heterogeneous regions [38].

### 2.2.1. Radiomics for sub-region analysis

Radiomics for sub-region analysis involves partitioning a tumour volume into spatially distinct habitats or subregions, each assumed to exhibit more homogeneous imaging characteristics. This approach is motivated by the inherent heterogeneity of tumours, which often comprise regions with varying cellularity, necrosis, vascularity, and other microenvironmental factors. By employing manually or semi-automatically generated sub-region masks, radiomics enables the extraction and summarization of quantitative features from each sub-region, thereby providing a more nuanced characterization of tumour heterogeneity. For example, Li et al. [10] proposed a radiomic framework that extracts imaging features using GBM sub-region labels defined by clinicians, while Zhang et al. [11] leveraged an unsupervised segmentation method to automatically generate unannotated sub-region labels. Although both

approaches have demonstrated the potential of sub-region analysis in medical imaging, they are still constrained by the reliance on manual labelling processes.

## 2.3. Deep learning

Deep learning has emerged as a sophisticated approach for medical image analysis, enabling the automatic learning of complex feature representations directly from imaging data. Convolutional neural networks (CNNs) have demonstrated remarkable success in various tasks, including tumour segmentation, classification, and outcome prediction, often achieving performance on par with or exceeding that of human experts in specific domains [39]. Unlike radiomics, which relies on handcrafted features engineered by humans, deep learning models iteratively learn their features from training examples, capturing subtle patterns and interactions in the data that might be difficult to encode manually. Within the context of tumour imaging, deep learning methods have been applied to analyse multimodal inputs (such as combining MRI, CT, and PET channels) and to incorporate 3D spatial context, thereby further enhancing the characterisation of complex lesions. This section discusses two branches of deep learning relevant to this thesis: unsupervised segmentation frameworks, which aim to delineate tumours or subregions without annotated labels, and supervised classification frameworks, which learn to predict known labels (e.g., tumour type or patient outcome) from imaging data. Both approaches have been explored to tackle the challenges of tumour heterogeneity in medical imaging, as outlined below.

### 2.3.1. Unsupervised Segmentation Frameworks

Unsupervised segmentation refers to methods that partition images into meaningful regions without the need for manually annotated training labels. The goal

is to discover inherent structure in the image data – for example, separating a tumour from surrounding tissue or dividing a tumour into internally consistent subregions – purely based on patterns in the images themselves [40]. Classical unsupervised segmentation techniques in medical imaging include intensity thresholding, region growing, and clustering algorithms such as k-means or fuzzy c-means, which group pixels or voxels based on similar intensity or texture characteristics [41]. These traditional approaches have been used to identify tumour boundaries or tissue compartments without supervision. Still, they often rely on simple intensity differences and may struggle with noisy data or complex anatomy. In recent years, deep learning has been leveraged to enhance unsupervised segmentation by learning more sophisticated feature representations. Deep unsupervised segmentation frameworks typically involve neural networks that learn to encode the image into a latent representation from which a segmentation can be derived [42]. For instance, a deep network might be trained to cluster pixel embeddings or reconstruct the input image, with the network's latent space organised such that distinct regions (e.g., foreground vs. background or different tissue types) separate naturally. The key advantage is that the network can capture high-level features (edges, textures, and even subtle tumour-specific patterns) that beyond intensity distribution, enabling more robust segmentation of complex, heterogeneous tumours without the need for ground-truth masks [43].

DeepCluster [44] is a framework designed initially for unsupervised feature learning. DeepCluster iteratively clusters the features extracted by a convolutional neural network and then uses the resulting cluster assignments as pseudo-labels to further train the network. When applied to segmentation, DeepCluster can be used to partition an image into coherent regions by clustering intermediate feature maps rather than raw pixel values. In this setting, the network alternates between grouping features into clusters and updating its parameters to better discriminate between these clusters, thereby gradually refining the segmentation. This iterative process enables the network

to capture complex structures and subtle heterogeneity in tumour images. Moreover, because the clustering is performed in a high-dimensional feature space, DeepCluster can leverage learned representations that encapsulate both intensity and textural information, which is particularly advantageous for delineating the intricate subregions within heterogeneous tumours.

Kim et al. [45] propose an unsupervised image segmentation framework that treats segmentation as a differentiable clustering problem on pixel features. The method alternates between assigning pixels to cluster labels (based on the similarity of their CNN-derived feature embeddings) and updating the network parameters in a self-supervised loop, similar to iterative deep clustering. This training strategy optimises a joint objective that encourages pixels with similar features to share the same label and enforces spatial continuity (neighbouring pixels are likely assigned the same label) while also promoting the use of a large number of distinct segment labels to avoid trivial solutions. As a result, the model learns to produce fine-grained yet coherent segmentation maps without any ground-truth annotations, effectively discovering object and region boundaries through feature clustering alone. The approach is fully differentiable end-to-end, yielding segmentations that are both semantically meaningful and spatially smooth and showcasing the potential of purely self-supervised learning for image segmentation.

### 2.3.2. Supervised Classification Frameworks

In supervised deep learning frameworks, models are trained on labelled datasets to learn a direct mapping from input images (or image-derived features) to a corresponding target output, such as a class label, disease status, or clinical outcome. In the context of tumor imaging, supervised classification networks are widely used to predict labels that are diagnostically or prognostically relevant. For example,

distinguishing between tumours and non-tumours, classifying tumour subtypes (based on histology or genetic profile), or predicting response to therapy – all based on the imaging appearance [46]. CNN-based classifiers take image data (2D slices, 3D volumes, or multimodality combinations) as input and process them through multiple convolutional layers to extract high-level features, which are then fed to fully connected layers or other classifiers to output the predicted class. Through training on many labeled examples, the network's filters become tuned to image patterns that correlate with the labels. Supervised deep learning frameworks have been successfully applied across many tumor types and imaging modalities. For tumor heterogeneity analysis, a typical application is predicting molecular or histopathological subtypes directly from imaging. In many cases, deep learning classifiers have achieved higher accuracy than traditional approaches. For instance, in lung cancer, a CNN-based model predicting 2-year survival outperformed a conventional radiomics model that utilized handcrafted texture and shape features, suggesting that the deep model captured prognostic heterogeneity patterns that radiomics did not [47, 48].

DenseNet [49], or the Dense Convolutional Network, is distinguished by its dense connectivity pattern, where every layer is directly connected to all subsequent layers within a dense block. Each layer receives as input the concatenated feature maps of all preceding layers, fostering extensive feature reuse throughout the network. Such connectivity not only promotes efficient parameter usage by reducing redundancy but also significantly improves gradient flow during training, thereby mitigating the vanishing gradient problem commonly encountered in deep networks. DenseNet is organised into dense blocks—groups of convolutional layers that incrementally add new features—interleaved with transition layers that perform downsampling and feature compression. The transition layers employ 1×1 convolutions and pooling operations to control the dimensionality of feature maps and maintain computational efficiency. This design results in a network that learns rich, multi-scale representations;

early layers capture fine, local texture details, while later layers integrate these features into more abstract, global representations. The dense connectivity also functions as an implicit form of deep supervision, as each layer has direct access to the final output's gradients, further enhancing training stability and convergence speed.

Transformers are a class of deep neural architecture that rely entirely on self-attention mechanisms to capture dependencies within an input sequence [50]. Unlike traditional convolutional or recurrent networks, Transformers compute attention scores between all pairs of input tokens simultaneously, enabling them to model long-range relationships without sequential processing. Their architecture is centred around a multi-head attention mechanism, where the input is projected into several subspaces, and each head independently learns to focus on different aspects of the data. The outputs from these attention heads are then concatenated and passed through a position-wise feed-forward network. Residual connections and layer normalisation wrap around these sublayers to facilitate training by preserving gradient flow and stabilising activations. Positional encodings are added to the input embeddings to retain order information, as the self-attention mechanism itself is permutation invariant. In recent adaptations for vision tasks, variants such as UniFormer [51] have emerged, integrating the strengths of Transformers with convolutional operations.

UniFormer leverages local attention in its early layers—mimicking the behavior of convolutional networks—to capture fine-grained spatial details and reduce feature redundancy. In deeper layers, architecture transitions to global self-attention, thereby allowing each image patch to interact with every other patch and capture holistic contextual relationships. This staged approach combines the efficiency and inductive biases of CNNs with the global modelling capacity of Transformers, yielding a balanced architecture that efficiently extracts both local textures and global semantic context. Such a hybrid design is particularly effective for complex vision tasks where both detailed spatial analysis and comprehensive contextual understanding are critical.

# Chapter 3. Unsupervised Glioblastoma Sub-Region Segmentation in MRI Modalities

In this chapter, to supplement unsupervised learning for GBM sub-region segmentation, a novel multi-phase and hierarchical unsupervised learning method is proposed, specifically for the segmentation of GBM sub-regions using mpMRI. The proposed method innovates by leveraging intrinsic image features and tumour spatial relationships encoded in mpMRI without reliance on annotated data. The principal novelty of this method is the exploitation that different tumour sub-regions exhibit unique textural and contrast characteristics among the different sub-regions, which can be effectively leveraged by the proposed hierarchical loss functions. This contribution is demonstrated in enhancing the heterogeneity tumour analysis efficiency based on automatic sub-regions segmentation.

## 3.1. Contributions

The main contributions of the proposed method are as follows:

- **Multi-Phase Training:** The proposed method uses a phased training approach where initial training phases focus on coarse segmentation based on semantic features, followed by successive phases that refine the segmentation based on finer, more specific characteristics. This strategy enables progressive learning and adaptation to complex patterns in the data, thereby

enhancing the model's ability to distinguish between tumour sub-regions.

- **Context-based Hierarchical Loss:** This loss function adapts to the natural hierarchy of tumour anatomy, allowing the model to refine segmentation further as it learns the intrinsic structures of the tumour regions. This loss enforces the predicted mask to be spatially consistent with the natural structure of GBM among the predicted tumour sub-regions.

## 3.2. Proposed Method

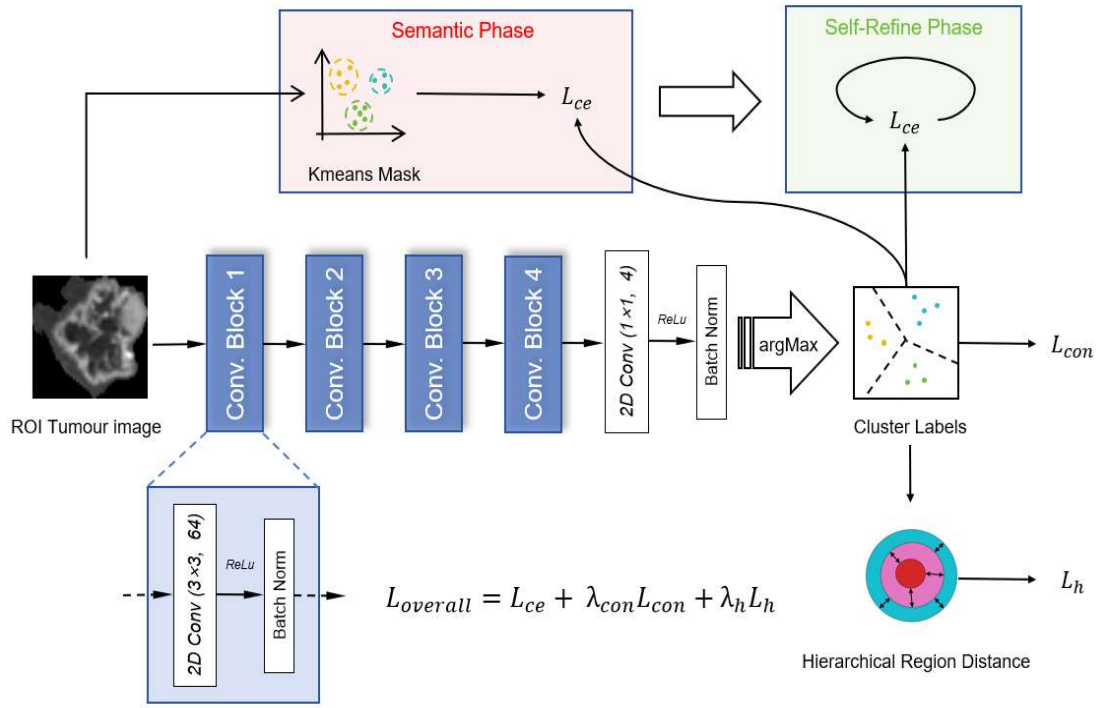


**Figure 3.1.** Proposed Multi-phase training architecture for our unsupervised segmentation method, which includes a semantic phase (light red box) and a self-refine phase (light green box). The hierarchical region distance was retrieved from the clustered labels.

### 3.2.1. Multi-phase training Architecture

The schematic of the proposed unsupervised GBM segmentation model is shown in Figure 3.1. It utilises the intensity distribution difference in sub-regions of mpMRI images to generate clustering pseudo-labels for sub-region segmentation. Given an image $x_n$, a convolutional segmentation network $F$ train to generates a response map $S_n = F(x_n)$. The response map is normalised to $\hat{S}_n$ , which has a mean of 0 and a standardised variance of 1. The final segmentation $C_n$ is obtained by selecting the channel number with the maximum response value in $\hat{S}_n$, where response map can be seen as logit indicating the likelihood that each pixel belongs to the cluster represented by each channel. The network $F$ was trained in both the semantic phase and self-refine phase using different pseudo-labels. In the semantic phase, the network $F$ was trained to learn semantic information by minimising the cross-entropy loss between $C_n$ and the pseudo-label $\hat{Y}$ generated by k-means [26]. The self-refine phase was trained to learn refined segmentation maps via feature similarity by minimising the cross-entropy loss between $\hat{S}_n$ and $C_n$ . Sparse spatial loss and context-based hierarchical loss are also used to train the model to learn spatial relationships of spatial continuity and regional arrangements.

## 3.2.2. Loss function

The proposed loss function $L$ consist of: A cross-entropy loss to learn representation of semantic information present in pseudo-label during semantic phase and feature similarity in self-refine phase; A sparse spatial continuity loss to constrain spatial continuity; A Context-based hierarchical loss to constrain predicted $C_n$ follows the natural hierarchical structure of GBM. The $L$ denoted as follows:

$$L = L_{ce}(t) + \lambda_{con} L_{con}(\hat{S}_n) + \lambda_h L_h(C_n)$$

$$L_{ce}(t) = \begin{cases} L_{ce}(\hat{Y}_n, \hat{S}_n) & \text{where } t \leq \mu \\ L_{ce}(\hat{S}_n, C_n) & \text{otherwise} \end{cases}$$

where $t$ represents current epoch number, $\mu$ represents the desired epoch number for semantic phase, $\lambda_{con}$ and $\lambda_h$ represents the weight for sparse spatial continuity loss and hierarchical loss, respectively.

### 3.2.3. Cross-entropy loss for unsupervised clustering

In the proposed architecture, the network $F$ is trained by reducing the cross-entropy loss between $\hat{S}_n$ and pseudo-labels instead of ground truth. During the semantic phase, k-means algorithm was adopted to create a soft mask based on the MRI images, which serves as a pseudo-label for guiding the initial clustering process to learn semantic information in the image effectively. As training progresses, the framework employs a transition from using the k-means mask to the argmax label $C_n$ learned by the network $F$, which would further refine the segmentation map by prompt clustering. The cross-entropy loss for semantic phase is defined as follows:

$$L_{ce} = -\sum_{i=1}^{N}\sum_{c=1}^{C}\hat{Y}_{i,c} ln C_{i,c}$$

where $c$ is the cluster index, $\hat{Y}$ denotes the pseudo-label, which is the k-means mask for the semantic phase. The cross-entropy loss for self-refine phase is defined as follows:

$$L_{ce} = -\sum_{i=1}^{N}\sum_{c=1}^{C} C_{i,c} ln \hat{S}_{i,c} .$$

### 3.2.4. Sparse spatial continuity loss

The sparse spatial continuity loss used in the method is based on L1-norm regularisation, an approach that has been widely used in clustering-based segmentation to enhance clustering and generalisation [52]. The encouraged sparsity enhances the understanding of the spatial relationships between pixels and regions in the image,

promotes spatial connectivity among regions, and thus facilitates smoother transitions between adjacent pixels within the same tumour subregion. The L1-norm is used to measure the vertical and horizontal sparsity. The loss is defined as follows:

$$L_{con} = \sum_{w=1}^{W} \sum_{h=1}^{H} \left\| \hat{S}_{w,h+1} - \hat{S}_{w,h} \right\| + \left\| \hat{S}_{w+1,h} - \hat{S}_{w,h} \right\|$$

where $w$ and $h$ denote the width and height of the response map, respectively. $\hat{S}_{w,h}$ represents the pixel value at coordinate $(w, h)$ in the normalised segmentation map.

### 3.2.5. Context-based hierarchical loss

In the proposed method, the natural hierarchy of GBM heterogeneity is characterised by a hierarchical loss, which ensures that the segmented region of interest (ROI) is partitioned with respect to a logical spatial arrangement of the tumour sub-regions. Channels (sub-regions) are sorted based on their spatial characteristics and proximity to the overall centroid of the segmentation. For each pair of adjacent sorted subregions, the distance between their boundaries is computed using the Euclidean distance transform. The definition is as follows:

$$L_h = \sum_{c=1}^{C-1} \frac{\sum_{i,j} \left( \| (i,j) - (k,l) \|_{1-M_c} \times M_{c+1,i,j} \right)}{\sum_{i,j} M_{c+1,i,j}}$$

where $M_c$ represents the extracted mask from $C_n$ and ( $\| (i,j) - (k,l) \|_{1-M_c}$ calculates the Euclidean distance from pixel $(i, j)$ to the nearest background pixel.

## 3.3. Experiment

### 3.3.1. Dataset description

The proposed method was evaluated on the 2021 BraTS dataset [53]. The BraTS21 dataset includes 1,251 annotated MRI data of patients with GBM. It is pre-processed as follows. MRI subjects have been skull-stripped and correctly registered, with a shape of 240 × 240 × 155. Each MRI scan includes four registered sequences: T1, T1Gd, T2, and T2-FLAIR. For each patient, the necrotic core and enhancing tumour are labelled using T1Gd, corresponding to hypo- and hyper-intense areas, respectively. The oedema regions are labelled using T2-FLAIR. The ground truths of tumour regions are refined automated segmentations validated by two senior radiologists.

### 3.3.2. Evaluation metrics

The performance of the proposed unsupervised method was evaluated by using two standard metrics: Dice similarity coefficient (Dice) and sensitivity (SEN), which are defined by the following equations:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}$$

$$SEN = \frac{TP}{TP + FN}$$

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

Due to the nature of the clustering method, the corresponding sub-region of a specific cluster cannot be retrieved. The ground truth labels are aligned with the predicted clusters by calculating the Dice scores for each cluster-region pair, determining the cluster with the highest Dice score for each region, and then mapping

these clusters to the respective areas to create aligned labels. If a particular type of sub-region is not present in the ground truth, the corresponding metrics are not calculated.

### 3.3.3. Training details

To incorporate the BraTS21 dataset into the 2D network, the slice with the largest tumour region was selected. The entire tumour region was masked by the ground truth and cut out using the smallest enclosing rectangle, then resized to 128 × 128.

The k-means clustering was computed on the processed tumour image with a predefined number of three clusters based on GBM sub-region definition, as described in Section 2.2. The network was first trained using the k-means mask for 150 epochs and then fine-tuned using the argmax clustered mask for another 450 epochs. The Adam optimiser used with a uniform learning rate of 0.001 and set to $\lambda$ 1 for both $\lambda_c$ and $\lambda_h$ in all phases. The proposed framework was implemented using PyTorch and trained on an NVIDIA RTX 4090 GPU.

For evaluation purposes, U-Net [24] was selected to demonstrate the upper bound of performance in this task as a fully supervised approach. The U-Net was trained for 500 epochs using a learning rate of 1e-5 and a batch size of 32.

### 3.3.4. Baselines

The Common unsupervised segmentation methods that use clustering methods as the baselines, including k-means [26], DeepCluster [44], and Differentiable Feature Clustering (DFC) [45] are selected. k-means is used to segment the image into predefined 4 clusters, which follow the context of 3 sub-regions plus background. DeepCluster is trained using a feature-level k-means mask with the Adam optimiser

and a learning rate of 0.0001 for 2000 epochs. DFC employs a similar approach in the self-refine phase, focusing on feature similarity and training with the Adam optimiser at a learning rate of 0.001 for 600 epochs.

## 3.4. Results

**Table 3.1.** Evaluation of performance on BraTS21 dataset with DICE and SEN. The best results are bolded, and the second best is underlined.

| Methods | Dice | | | | Sensitivity | | | |
|---|---|---|---|---|---|---|---|---|
| | Necrotic core | Enhanced tumour | Oedema | Overall | Necrotic core | Enhanced tumour | Oedema | Overall |
| U-Net | 0.6096 | 0.8534 | 0.9121 | 0.7974 | 0.6594 | 0.8559 | 0.8968 | 0.8087 |
| k-means [26] | <u>0.3833</u> | <u>0.7137</u> | 0.6829 | <u>0.6041</u> | 0.7730 | 0.6871 | 0.6236 | 0.6801 |
| DeepCluster [44] | 0.3057 | 0.6170 | <u>0.7090</u> | 0.5439 | 0.7806 | 0.6842 | <u>0.6665</u> | 0.7000 |
| DFC [45] | 0.3035 | 0.6317 | **0.7700** | 0.5805 | **0.8486** | **0.8600** | **0.7793** | **0.8207** |
| Proposed method | **0.4144** | **0.7580** | 0.7087 | **0.6359** | <u>0.8106</u> | <u>0.7831</u> | 0.6424 | <u>0.7266</u> |

Table 3.1 presents the performance evaluation on the BraTS21 dataset using Dice and sensitivity metrics across three sub-regions including necrotic core, enhanced tumour, and oedema and overall. Under complete supervision, the model achieved the highest scores, with an overall Dice of 0.7974 and an overall sensitivity of 0.8087. Among the unsupervised methods, while DFC achieves the highest sensitivity (0.8207 overall), the proposed method demonstrates more balanced performance, with an overall Dice score of 0.6359 and an overall sensitivity of 0.7266. Notably, the overall approach improves the segmentation of the necrotic core (Dice of 0.4144) compared to k-means (0.3833) and DeepCluster (0.3057), and it maintains competitive performance across the enhanced tumour and oedema regions.

**Table 3.2** Ablation study result on BraTS21 dataset with DICE and SEN. The best results are bolded, and the second best is underlined.

| Method | Dice | | | | Sensitivity | | | |
|---|---|---|---|---|---|---|---|---|
| | Necrotic core | Enhanced tumour | Oedema | Overall | Necrotic core | Enhanced tumour | Oedema | Overall |
| semantic phase | 0.3790 | 0.6823 | <u>0.7049</u> | 0.6023 | 0.7933 | 0.7173 | **0.6515** | <u>0.7088</u> |
| self-refine phase | 0.3239 | 0.6127 | 0.6311 | 0.5334 | 0.7658 | 0.7510 | 0.5425 | 0.6678 |
| both phase | 0.3831 | 0.7284 | 0.6758 | 0.6033 | 0.7773 | 0.7546 | 0.5989 | 0.6922 |
| both phase +$L_{con}$ | <u>0.4009</u> | <u>0.7294</u> | 0.6994 | <u>0.6188</u> | <u>0.8089</u> | **0.7843** | 0.6293 | 0.7222 |
| both phase +$L_{con}$ +$L_h$ | **0.4144** | **0.7580** | **0.7087** | **0.6359** | **0.8106** | <u>0.7831</u> | <u>0.6424</u> | **0.7266** |

Table 3.2 extends this evaluation through an ablation study, dissecting the contributions of different training phases and loss components. The semantic phase alone yields an overall Dice score of 0.6023, whereas the self-refine phase alone yields a score of 0.5334. However, combining both phases improves the overall Dice to 0.6033. Further integration of the sparse spatial continuity loss ($L_{con}$) increases the overall Dice score to 0.6188, and the subsequent addition of the context-based hierarchical loss ($L_h$) yields the highest overall Dice score of 0.6359. These results confirm that each component—the multi-phase training strategy along with the continuity and hierarchical loss contributes significantly to the segmentation performance, validating the effectiveness of the proposed approach in handling the heterogeneous nature of tumour sub-regions.

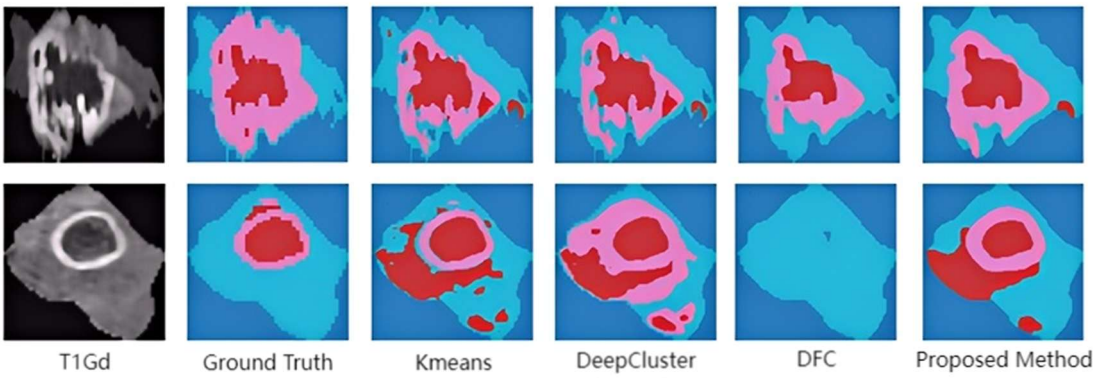


**Figure 3.2.** Segmentation results on BraTS21 dataset. From left to right: T1Gd sequence of an MRI (1st column), ground truth segmentation (2nd column), segmentation results of comparison methods (3rd-5th columns), and the results of the proposed method (6th column). Dark Blue is the background, light blue represents oedema, pink indicates the enhanced tumour, and red represents the necrotic core. The two wors represent two distinct heterogenous tumour cases.

Figure 3.2 illustrates that, while k-means, DeepCluster, and DFC can broadly capture some tumour regions, they often misclassify the necrotic core or produce inconsistent boundaries compared to the ground truth. By contrast, the proposed method more accurately delineates each sub-region. This indicates that our approach provides more precise and spatially coherent segmentations than other unsupervised clustering methods.

## 3.5. Discussion

The result in Table 1 demonstrates that the proposed method outperformed the standard clustering methods. The sensitivity of the proposed method is lower than that of DFC because it relies on k-means clustering to initialise the sub-regions, which can hinder its ability to accurately capture these regions initially. However, the higher performance in Dice metric demonstrated a more precise segmentation for the proposed method. Compared to the other methods, the proposed method achieved a more balanced performance across each sub-region. This suggests the effectiveness of the

multi-phase training approach and context-based hierarchical loss. For the three sub-regions presented in GBM, the overall performance of necrotic core segmentation is comparatively lower than other regions. As shown in Figure 3.2, despite the discontinuity of the clusters, the unsupervised methods struggled to distinguish between the necrotic core and the oedema, assigning them to different clusters. This may be due to the similarity in intensity distribution between the two sub-regions across all the sequences. Despite this, the proposed method significantly outperformed the other three unsupervised methods in segmenting the necrotic core region, achieving a 3% higher result compared to the second-best method. This implies that the sparse continuity loss and hierarchical loss effectively provide additional spatial relationship aspects, allowing for more nuanced and connected pixel grouping and helping the model segment this region.

An ablation study was conducted to demonstrate the efficacy of each module in proposed framework, the results of which are presented in Table 2. When only using k-means or argmax to generate pseudo-labels for training the model, it achieves Dice scores of 0.60 and 0.53, respectively. Training using the multi-phase approach achieved a Dice score of 0.61, and this further increased to 0.62 when the sparse spatial continuity loss was added. Combining all components yielded the highest Dice score of 0.64. This demonstrates that the individual components of the proposed framework are essential in improving the results. The proposed approach is easily transferable and can be integrated as an enhancement module in other unsupervised segmentation frameworks. The approach is adaptive to any pseudo-labels provided and can effectively retrieve the semantic information from the given pseudo-labels. The context-based hierarchical loss constrains the spatial relation between segmented sub-regions from the mask level. This adaptability allows hierarchical loss to be used not only in unsupervised learning but also in supervised learning. I will investigate its application using other imaging modalities or in the context of different diseases in the future.

The visualised segmentation results are illustrated in Figure 3.1. The necrotic core was hard to be distinguished from the oedema region in all segmentation method. This difficulty likely stems from the similar intensity distributions between these two regions in certain water-insensitive imaging modalities. Additionally, the complexity of the enhanced tumour region in ground truth further complicates its differentiation from the oedema region. The proposed method improves upon existing techniques by incorporating a hierarchical loss that guides the segmentation masks to conform to the inherent GBM structures. The DFC method is hard to fine-tune with the continuity loss; in cases like the second row, the DFC method considers the whole tumour as the same entity. While the proposed method uses a semantic phase to initiate segmentation to capture sub-regions at a coarse level, effectively capturing sub-regions and overcoming the continuity loss issue.

Despite the proposed approach having higher performance among unsupervised methods, several limitations were identified. When the pseudo-labels provided in the semantic phase cannot characterise sub-region information, it is likely the model will not learn the semantic information even with continued training. This suggests that the proposed method is highly dependent on the initial semantic pseudo-labels. It is possible to apply other unsupervised learning methods, such as contrastive learning or high-level feature similarity constraints, to reduce the dependency on the initial labels. Another limitation is that the experiments were conducted on 2D images rather than 3D; adapting the proposed method to support 3D input will be part of the future work. The proposed method relied on the cropped tumour region, as the focus was on segmenting the sub-regions. This limitation could be mitigated by adopting whole-tumour segmentation algorithms, e.g., CGC [54], as a pre-process to automatically segment the whole tumour before providing it to the proposed method.

## 3.6. Conclusion

In this study, a new unsupervised method was proposed with a novel multi-phase training approach and a context-based hierarchical loss for GBM sub-region segmentation using mpMRI. The proposed method enabled the network to obtain semantic information from pseudo-labels effectively and ensures that the segmentation results adhere to the native hierarchical structure present in GBM. Evaluation on the BraTS21 dataset demonstrated an improved performance of our model compared to k-means and other common unsupervised clustering methods.

# Chapter 4. Heterogeneity awareness based deep learning framework for tumour classification

To characterise and exploit intra-tumour heterogeneity, a novel heterogeneity-aware deep learning classification (HA-DLC) method is proposed for tumour disease classification task. Figure 2.2 is an overview of the proposed method, comprising of dual-stream feature extraction: the first stream is local heterogeneity feature extracted from the Modality-wise Segmentation Module, which leverages sub-region labels to obtain heterogeneity information, and the second stream is global features extracted from the Cascaded Feature Extraction Network using whole image. The proposed method innovates in the use of existing radiomics approach for sub-region-based feature extraction. To achieve generalisability to work on different tumour classification task, a pseudo label generation module is proposed that generate consistent pseudo sub-region labels through heterogeneous sub-region for any given tumour cases. The proposed method was evaluated using two datasets consist with different mpMRI sequence sets and different tasks to demonstrate the robustness and adaptability of the proposed method.

## 4.1. Contributions

The main contributions of this chapter are as follows:

- **Learnable Alignment for Sub-Region Heterogeneity:** An innovative alignment module is introduced that integrates complementary features derived from the dual streams and incorporates sub-region heterogeneity labels from unlabelled data. This module utilises a modality-wise alignment network to generate 'relation-maps' for sub-region labels produced by unsupervised segmentation. The resulting relation-maps cluster index in unaligned sub-regions which mitigates inconsistencies from unfixed cluster indexes. This enables the module to leverage intrinsic sub-region information without relying on large labelled (manual) data. In contrast to the method described in Chapter 3, this module is applicable to multiple tumour types. Moreover, it preserves relation-maps throughout the classification training process, ensuring that the generated subregions are directly linked to the captured heterogeneity.
- **Heterogeneity and Global Dual-Stream Feature Extraction:** Two complementary feature extraction 'streams' are proposed to capture both local heterogeneity and global features. The Modality-wise Segmentation Module as local heterogeneity stream is trained to extract local heterogeneity features by hard-sharing the feature in both segmentation and classification. In parallel, a global module extracts robust contextual features from the entire image.

# 4.2. Proposed Method

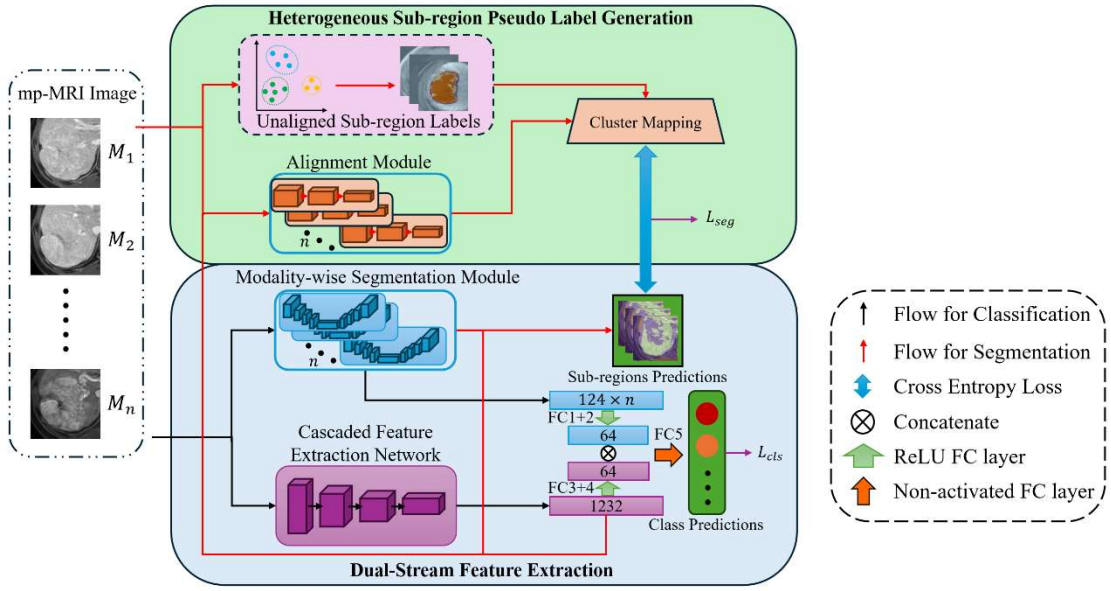


**Figure 4.1.** Overview of our Heterogeneity-Aware Deep Learning for tumour classification. It comprises a heterogeneous sub-region pseudo label generation module and dual-stream feature extraction module. Segmentation loss is derived from the aligned cluster sub-region labels and sub-region label predictions.

## 4.2.1. Overview

Figure 4.1 illustrates the proposed heterogeneity-aware deep learning framework, which comprises two main modules: (1) heterogeneous sub-region pseudo label generation and (2) dual-stream feature extraction. Unaligned sub-region labels are initially derived in the first module by applying an unsupervised semantic segmentation technique to each mpMRI volume. Because these labels are inherently inconsistent across different volumes (i.e., the same cluster index may refer to different tissue sub-types in other images), an Alignment Module is introduced to enforce label consistency. Specifically, the Alignment Module consists of $n$ Align Networks, where $n$ is the number of modalities in mpMRI. Each Align Network takes three inputs: (i) the corresponding modality in the mpMRI image, (ii) the sub-region probability map obtained by applying a softmax function to the response map produced by the

corresponding segmentation network, and (iii) a global deep feature extracted by a Cascaded Feature Extraction Network (CFEN). Using these inputs, the Align Network generates a relational table that maps corresponding sub-regions across different volumes to the same label, thereby converting unaligned labels into aligned sub-region pseudo labels suitable for supervised training.

The second dual-stream feature extraction module comprises a segmentation network and the CFEN. The Modality-wise Segmentation Module comprises multiple segmentation networks that segment sub-regions and extract distinct, heterogeneous features for each modality. Segmentation backbone outputs are trained by aligned sub-region pseudo labels using segmentation loss $L_{seg}$. Concurrently, 114 heterogeneous features per modality from the segmentation network's encoder and 1,232 global features are each fed through two fully connected layers. FC1 and FC3 each have 512 neurons, and FC2 and FC4 each have 64 neurons; all FC1-4 employ Rectified Linear Unit (ReLU) activations. The non-activated FC layer (FC5) utilises concatenated features from FC2 and FC4 to generate final classification predictions. The classification loss $L_{cls}$ is defined as the cross-entropy between the predicted labels and the ground-truth labels. The overall framework is optimised by minimising the combined loss, including $L_{cls}$ and $L_{seg}$.

## 4.2.2. Modality-wise Segmentation Module

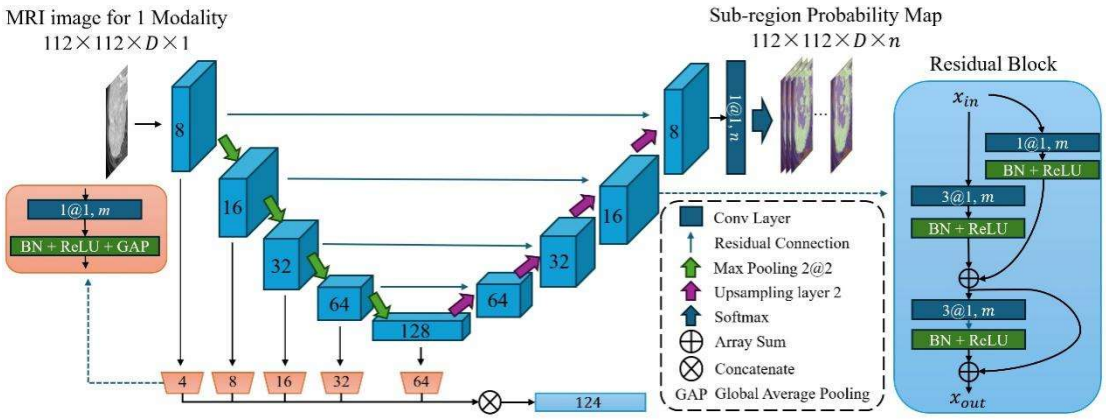


**Figure 4.2.** The segmentation network architecture for Modality-wise Segmentation Module. It comprises encoder and decoder streams constructed from residual blocks, as shown on the right. The local heterogeneity features were extracted from the output of each encoder residual block.

The modality-wise segmentation module consists of $n$ segmentation networks, where $n$ is the number of modalities in the mp-MRI. The structure ensures that each network can consistently extract distinct modality-specific heterogeneous features and segment sub-regions for each modality. The segmentation backbone of the segmentation network in the Modality-wise Segmentation Module is shown in Figure 4.2. It is built upon a customised 3D residual U-Net architecture [24]. Each segmentation network comprises two primary branches: a downsampling encoder and an upsampling decoder with skip connections. The encoder employs five cascaded residual blocks to extract deep features for classification and segmentation tasks. The decoder mirrors the first four residual blocks in the encoder and receives residuals from skip connections to complete sub-region segmentation. A max pooling layer with a kernel size of 2 × 2 × 2 was used to downsample between encoder blocks, and an upsampling layer with a kernel size of 2 × 2 × 2 was used to connect between decoder blocks.

A residual block contains two convolutional layers with kernel size 3×3×3, stride size 1×1×1 and kernel number of $m$. The second convolutional layer receives output from the first convolutional layer and a non-linear residual connection through a convolutional layer with a kernel size of 1×1×1, a stride size of 1×1×1, and a kernel number of $m$. The output of the residual block is the second convolutional layer, with a shortcut residual connection to the first convolutional layer. Batch normalisation and ReLU activation follow all convolutional layers. The final sub-region segmentation is derived from the last residual block, followed by a convolutional layer with a kernel size of 3×3×3, a kernel number of c, representing the predefined number of sub-regions, and a softmax activation function. The heterogeneous local features are extracted from the output of each residual block in the encoder using a convolutional layer with a kernel size of 1×1×1 and half the number of kernels of the corresponding residual block. This is followed by batch normalisation, ReLU activation, and a Global Average Pooling layer to obtain 124 deep features for each modality.

### 4.2.3. Alignment Module

The Align Module is designed to standardise sub-region probability maps across different imaging volumes by learning relational mapping. It ensures that similar tissue sub-regions across patients receive consistent labels, mitigating inconsistencies in unsupervised sub-region segmentation. This module consists of n alignment networks, where n is the number of modalities in the input mp-MRI. Each alignment network accepts three inputs: the corresponding modality in the mp-MRI image, the sub-region probability map for the modality and global deep features from the cascaded feature extraction network. Two modified DenseNet [49] were used to extract deep features from the mp-MRI image and the sub-region probability map. Each DenseNet in the alignment network consists of three dense blocks, and 112 deep features are extracted

from each block using the same approach as in the Modality-wise Segmentation Module. The deep features derived from DenseNet and global deep features are fed into individual fully connected (FC) layers with 64 nodes, then concatenated together and fed into an FC layer with c nodes to generate the relational table R to map unaligned labels into aligned sub-region pseudo labels.

### 4.2.4. Unsupervised Label Generation

To extract the sub-region level segmentation labels for supplement tumour heterogeneity information, pseudo sub-region labels were generated by using an unsupervised clustering approach. These pseudo labels serve as initial sub-regions, which are later refined through an alignment process to ensure consistency and accuracy. Two different unsupervised methods are included for sub-region label generation: (i) Intensity-based (k-means) [26] and (ii) Spatial and intensity-based (Differentiable Feature Clustering (DFC)) [45]. For k-means, the whole-tumour segmentation provided was used to generate the tumour ROI first, before performing clustering. For DFC, two strategies were employed: one is the same as the k-means; the other is to perform clustering at the whole image level and then retain the mask in the tumour ROI only, as DFC also considered spatial relations between the pixels present in the image. Finally, the aligned pseudo labels are used as supervisory signals to train the segmentation network, ensuring that the same sub-regions are identified consistently in every patient.

### 4.2.5. Cascaded feature extraction network

The cascaded feature extraction network is based on Uniformer [51]. Specifically, a Uniformer-small configuration was used in the framework to extract global features. The Uniformer-small consists of two local multi-head relation aggregator blocks and two global multi-head relation aggregator blocks. The same approach was used in the segmentation network to extract global features. The output from each block is fed into a convolutional layer with a kernel size equal to the number of channels in production, followed by batch normalisation, ReLU activation, and a Global Average Pooling layer to obtain 1232 deep features in total.

### 4.2.6. Loss

The heterogeneity-aware deep learning framework was trained end-to-end by combining segmentation loss and classification loss, as defined in the following equation.

$$L = L_{cls} + L_{seg}$$

where $L_{cls}$ is a cross-entropy loss for the classification task, and $L_{seg}$ is a cross-entropy loss for the segmentation task.

The classification loss $L_{cls}$ is defined as follows:

$$L_{cls} = -\sum_{i=1}^{M} Y_i ln\hat{Y}_i$$

where $M$ is the total number of classes, $Y$ denotes the one if $i$ is the ground truth label else zero and $\hat{Y}$ denotes the probability of class i.

The segmentation loss $L_{seg}$ is defined as follows:

$$L_{seg} = \frac{-\sum_{i=1}^{N}\sum_{c=1}^{C}\hat{C}_{i,c} ln \hat{S}_{i,c}}{N}$$

where $c$ is the cluster index, $N$ is the modality number, $\hat{C}$ denotes the aligned pseudo-label generated by unsupervised segmentation, and $\hat{S}$ denotes the probability map obtained from the response map outputted from the Modality-wise Segmentation Module.

## 4.3. Experiment

### 4.3.1. Dataset

The proposed method was evaluated on two public datasets: (1) Liver Lesion Diagnosis Challenge on Multi-phase MRI (LLD-MMRI2023) dataset [55] and (2) Radiogenomic Classification from RSNA-ASNR-MICCAI Brain Tumor Segmentation Challenge 2021 (BraTS2021 Radiogenomic Classification) dataset [53].

LLD-MMRI2023 comprises 498 patients, utilising eight MRI modalities, including T2W, DWI, in-phase, out-of-phase, pre-contrast, arterial, portal venous, and delayed phases. The task is to classify the type of lesion present in the image, including four benign types (Hepatic hemangioma, Hepatic abscess, Hepatic cysts, and Focal nodular hyperplasia) and three malignant types (Intrahepatic cholangiocarcinoma, Liver metastases, and Hepatocellular carcinoma). The provided tumour segmentation masks were generated through MedSAM [56].

The BraTS2021 Radiogenomic Classification dataset comprises 585 cases, featuring four MRI modalities: T1-weighted post-contrast (T1Gd), T1-weighted pre-contrast (T1w), T2-weighted (T2), and Fluid Attenuated Inversion Recovery (FLAIR). The task is to classify the methylation status of methylguanine-DNA methyltransferase (MGMT) promoter. To acquire the tumour segmentation masks, we used the patient ID

to match the images and masks in the RSNA-ASNR-MICCAI Brain Tumor Segmentation Challenge 2021 datasets. Cases that could not be found in the segmentation dataset were excluded and obtained 578 registered mp-MRI images with tumour segmentation masks.

### 4.3.2. Image preprocessing and unsupervised segmentation

For the LLD-MMRI2023 dataset, the tumour region of interest (ROI) was extracted through the provided lesion bounding box, which clinicians annotate. To standardise the shape of the input mp-MRI image, the ROI was extended to 128 × 128 × 16. Minimum-maximum normalisation was used to normalise the intensity.

For the RSNA-MICCAI Brain Tumor Radiogenomic Classification dataset, the matched mp-MRI images in the BraTS21 datasets are preprocessed by resampling, registration, skull stripping, and normalisation. The ROI region was extracted based on the bounding box of the whole tumour mask and then resized the ROI to 128×128×32.

For k-means-generated labels, the provided whole-tumour masks were used to crop the tumour region and run k-means for each modality. For DFC-generated labels, the entire image was used to obtain clusters and then crop the ROI, as the spatial relationship between pixels affects the clustering result.

### 4.3.3. Comparison methods

The leaderboard of the RSNA-MICCAI Brain Tumor 2021 (BraTS21) Radiogenomic Classification competition [57], vision graph neural network for BraTS21 Radiogenomic Classification [58] and Uniformer [51], which was used in the leaderboard solution for LLD-MMRI2023 were selected as comparison method. SOTA

models in other tasks including ResNet-50 [60], DenseNet-121 [49], EfficientNet-b7 [61] are also selected `for performance comparison. The MGMT leaderboard solution utilised a 3D ResNet-10 as its backbone, combining the 2D image with the largest brain area from each modality on the depth axis. To employ this framework in the LLD-MMRI2023 dataset, the image by the largest lesion region was selected and combined it with the same approach. Vision graph neural networks for MGMT prediction employed the PyramidViG-S structure and utilised FLAIR, T1W, and T2 fused 2D RGB images as input. We adopted the input to fuse all eight modalities for the LLD-MMRI2023 dataset. The Uniformer-small structure was used and the same input as the HA-DLC framework for the Uniformer.

### 4.3.4. Experimental settings

In this experiment, I used different parameters for unsupervised segmentation. For k-means, k values ranging from 3 to 9 cluster centres were used to determine the effect of complexity and the number of clusters. For DFC, 2, 0.5 and 0.001 were used as the continuity loss to generate detailed sub-regions.

I conducted an ablation study to gradually remove key modules from HA-DLC, demonstrating their efficacy on the BraTS21 Radiogenomic Classification dataset. The settings are shown as follows: (1) Classification backbone only – where the whole image was used in the classification without the segmentation and alignment components; (2) Classification loss only – where the segmentation loss was removed from the loss function; (3) Whole tumour segmentation – where the whole tumour mask, without sub-region labels, where used and, (4) Ground truth sub-region segmentation – where sub-regions used available from the BraTS 2021 Radiogenomic Classification dataset as the ground truth sub-region labels.

### 4.3.5. Evaluation metrics

For LLD-MMRI2023, the average of the F1 and Kappa scores was used as the primary evaluation metric in the competition. For BraTS21 Radiogenomic Classification, the Area Under the Curve was used as the primary evaluation metric in the competition. Additionally, accuracy is included as a supplement to performance evaluation. The definition of mentioned metrics is defined as follows:

$$F1\ Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1\ Score\ +\ Kappa\ score\ average = \frac{F1\ Score +\ Cohen's\ Kappa}{2}$$

where $TP$ is true positive, $FP$ is false positive, $TN$ is true negative and $FN$ is false negative in confusion matrix.

### 4.3.6. Implementation Details

The HA-DLC method and other comparison methods were implemented using Pytorch. An Adam optimiser was used with a batch size of 4 for LLD-MMRI2023 and a batch size of 2 for BraTS21 Radiogenomic Classification. Based on the difference in tumour size in depth, the image was resized to 128×128×16 for LLD-MMRI2023 and 128×128×32 for BraTS21 Radiogenomic Classification. For both datasets, data

augmentation applied was to the images, including random cropping to 112×112×depth-2, random flipping in three axes, and random rotation of up to 45 degrees. The learning rate was initially set to 1e-4, with a step learning rate scheduler implemented in PyTorch Image Models (timm) [59], using a decay rate of 0.1, and trained on a single 24GB NVIDIA RTX 4090 GPU. All models were trained and validated through 5-fold cross-validation.

## 4.4. Results

### 4.4.1. Performance of the Classification tasks

**Table 4.1.** Evaluation of performance on LLD-MMRI2023 and BraTS 2021 Radiogenomic Classification Datasets, including accuracy and primary metrics. The best result is bolded, and second best is underlined.

| Method | LLD-MMRI2023 | | MGMT-Classification | |
|---|---|---|---|---|
| | Accuracy | F1-Score + Kappa Average | Accuracy | AUC |
| Kaggle Leaderbroad | 0.677 ± 0.036 | 0.629 ± 0.048 | 0.624 ± 0.030 | 0.640 ± 0.036 |
| ViG-S | 0.719 ± 0.020 | 0.681 ± 0.021 | 0.603 ± 0.023 | 0.647 ± 0.023 |
| Resnet-50 | 0.761 ± 0.021 | 0.725 ± 0.020 | 0.622 ± 0.031 | 0.666 ± 0.012 |
| Efficientnet-b7 | 0.669 ± 0.054 | 0.610 ± 0.076 | 0.595 ± 0.063 | 0.657 ± 0.021 |
| Densenet-121 | 0.735 ± 0.057 | 0.700 ± 0.061 | 0.603 ± 0.023 | <u>0.667</u> ± 0.028 |
| Uniformer-S | 0.809 ± 0.020 | 0.778 ± 0.029 | 0.592 ± 0.019 | 0.606 ± 0.015 |
| HA-DLC (k-means: k = 5) | **0.843** ± 0.027 | **0.817** ± 0.043 | **0.659** ± 0.024 | 0.659 ± 0.026 |
| HA-DLC (DFC: μ = 0.001) | <u>0.825</u> ± 0.040 | <u>0.797</u> ± 0.047 | <u>0.633</u> ± 0.043 | **0.671** ± 0.043 |

Table 4.1 presents the classification results comparing HA-DLC to the state-of-the-art counterparts. The LLD-MMRI2023 dataset involves classifying distinct tumour types (inter-tumour heterogeneity), while the BraTS21 Radiogenomic Classification dataset requires classifying MGMT methylation status in glioblastoma (intra-tumour heterogeneity). The best pseudo-subregion label results were selected for HA-DLC using k-means (k = 5) in LLD-MMRI2023 and DFC (μ = 0.001) in BraTS2021 Radiogenomic Classification. The varying performance is presented in Section C below. The results indicate that HA-DLC outperformed all comparison methods, achieving an average F1-Score and Kappa of 0.817 on the LLD-MMRI2023 dataset and an AUC of 0.671 on the BraTS 2021 Radiogenomic Classification dataset. The proposed model attained accuracies of 0.843 and 0.633 on the LLD-MMRI2023 and BraTS 2021 datasets, respectively.

## 4.4.2. Ablation Studies

**Table 4.2.** Ablation Study Result of LLD-MMRI2023 and BraTS 2021 Radiogenomic Classification Datasets with primary evaluation metric. The best result is bolded, and second best is underlined.

| Method | Segmentation Backbone | Whole Tumour Segmentation | Sub-region Segmentation | LLD-MMRI2023 | MGMT-Classification |
|---|---|---|---|---|---|
| | | | | F1+Kappa Average | AUC |
| Classification Backbone only | | | | 0.778 ± 0.030 | 0.606 ± 0.015 |
| Without seg loss | √ | | | 0.768 ± 0.029 | 0.623 ± 0.054 |
| Whole Tumour Segmentation | √ | √ | | <u>0.802</u> ± 0.031 | <u>0.650</u> ± 0.021 |
| Sub-region Segmentation | √ | | √ | **0.817** ± 0.043 | **0.671** ± 0.037 |
| Ground Truth Segmentation | √ | | √ | NA | **0.681** ± 0.022 |

Table 4.2 summarises the ablation study results on the two datasets, focusing on the impact of segmentation goals on performance. The baseline classification backbone achieves an F1-score of 0.778 and a Kappa of 0.778 on LLD-MMRI2023, as well as an AUC of 0.606 on BraTS 2021. When the model is trained without the segmentation loss, its performance on LLD-MMRI2023 decreases slightly to 0.768, while the AUC on BraTS 2021 improves to 0.623. Incorporating whole-tumour segmentation improves the metrics to 0.802 and 0.650, respectively, and further employing sub-region segmentation enhances the performance to 0.817 on LLD-MMRI2023 and 0.671 on BraTS 2021. Additionally, ground-truth segmentation (available only for BraTS 2021) yields an AUC of 0.681.

### 4.4.3. Evaluation of pseudo sub-region labels generation strategies

**Table 4.3.** Result of the framework using k-means with a different number of centres on LLD-MMRI2023 and BraTS 2021 Radiogenomic Classification. The best result is bolded, and second best is underlined.

| Cluster Number | LLD-MMRI2023 | MGMT-Classification |
|---|---|---|
| | F1 + Kappa Average | AUC |
| 3 Centers | 0.799 ± 0.031 | 0.652 ± 0.046 |
| 4 Centers | 0.800 ± 0.036 | 0.652 ± 0.016 |
| 5 Centers | **0.817** ± 0.043 | <u>0.659</u> ± 0.026 |
| 6 Centers | 0.798 ± 0.038 | **0.660** ± 0.030 |
| 7 Centers | <u>0.807</u> ± 0.031 | 0.650 ± 0.050 |
| 8 Centers | 0.796 ± 0.034 | 0.650 ± 0.025 |
| 9 Centers | 0.782 ± 0.038 | **0.660** ± 0.023 |

The framework was trained using k-means clustering with varying numbers of centres. For the LLD-MMRI2023 dataset, the best performance, with an F1-score and Kappa of 0.817, is observed with five centres. Regarding the MGMT-classification task from the BraTS 2021 dataset, the AUC values remain relatively stable, ranging between 0.652 and 0.660 across cluster numbers from 3 to 9.

**Table 4.4.** Result of the framework using DFC with different coefficients of continuity loss on LLD-MMRI2023 and BraTS 2021 Radiogenomic Classification. The best result is bolded, and second best is underlined.

| DFC Continuity Loss Coefficient | LLD-MMRI2023 | | MGMT-Classification | |
|---|---|---|---|---|
| | F1+Kappa Average | Max Cluster Number | AUC | Max Cluster Number |
| Whole Image Level Clustering | | | | |
| mu = 2 | 0.780 ± 0.034 | 10 | 0.652 ± 0.030 | 9 |
| mu = 0.5 | <u>0.809</u> ± 0.045 | 17 | 0.664 ± 0.044 | 14 |
| mu = 0.001 | 0.797 ± 0.047 | 29 | **0.671** ± 0.043 | 24 |
| Tumour ROI Level Clustering | | | | |
| mu = 2 | 0.798 ± 0.037 | 7 | <u>0.664</u> ± 0.037 | 7 |
| mu = 0.5 | 0.793 ± 0.032 | 8 | 0.657 ± 0.054 | 8 |
| mu = 0.001 | **0.812** ± 0.028 | 15 | 0.656 ± 0.018 | 16 |

The DFC results are presented for two clustering approaches: whole image level and tumour ROI level. For whole-image-level clustering, as the continuity loss coefficient (μ) decreases from 2 to 0.001, the maximum number of clusters increases (from 10 to 29), with corresponding F1-score and Kappa score values ranging from 0.780 to 0.809 and AUC values from 0.652 to 0.671. Under tumour ROI level clustering, μ values ranging from 2 to 0.001 yield maximum cluster numbers from 7 to 15, with F1-Score and Kappa values between 0.793 and 0.812, and AUC values between 0.657 and 0.664. These experiments demonstrate the influence of clustering granularity and continuity constraints on the performance of the pseudo sub-region generation.

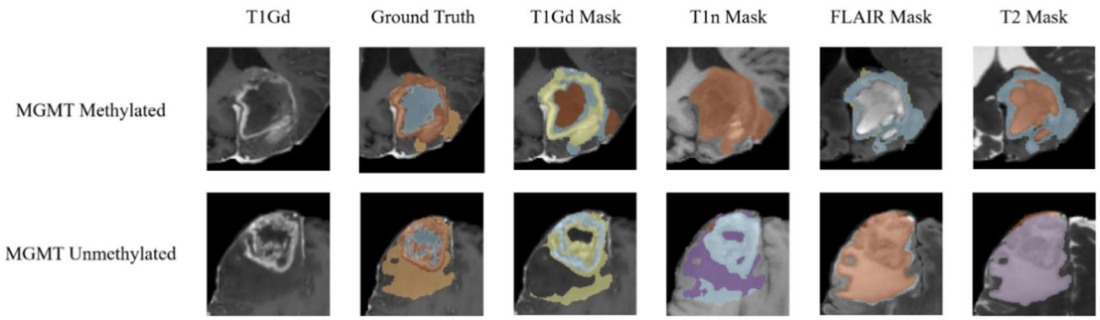


**Figure 4.3.** The original mpMRI image and pseudo sub-region labels of the BraTS 2021 Radiogenomic Classification dataset for two patients correctly predicted by the HA-DLC. The blue-shaded ground truth mask indicates the necrosis area, brown represents the enhanced tumour, and orange represents the oedema region. The different colours in the aligned clustered mask indicate different sub-region labels.

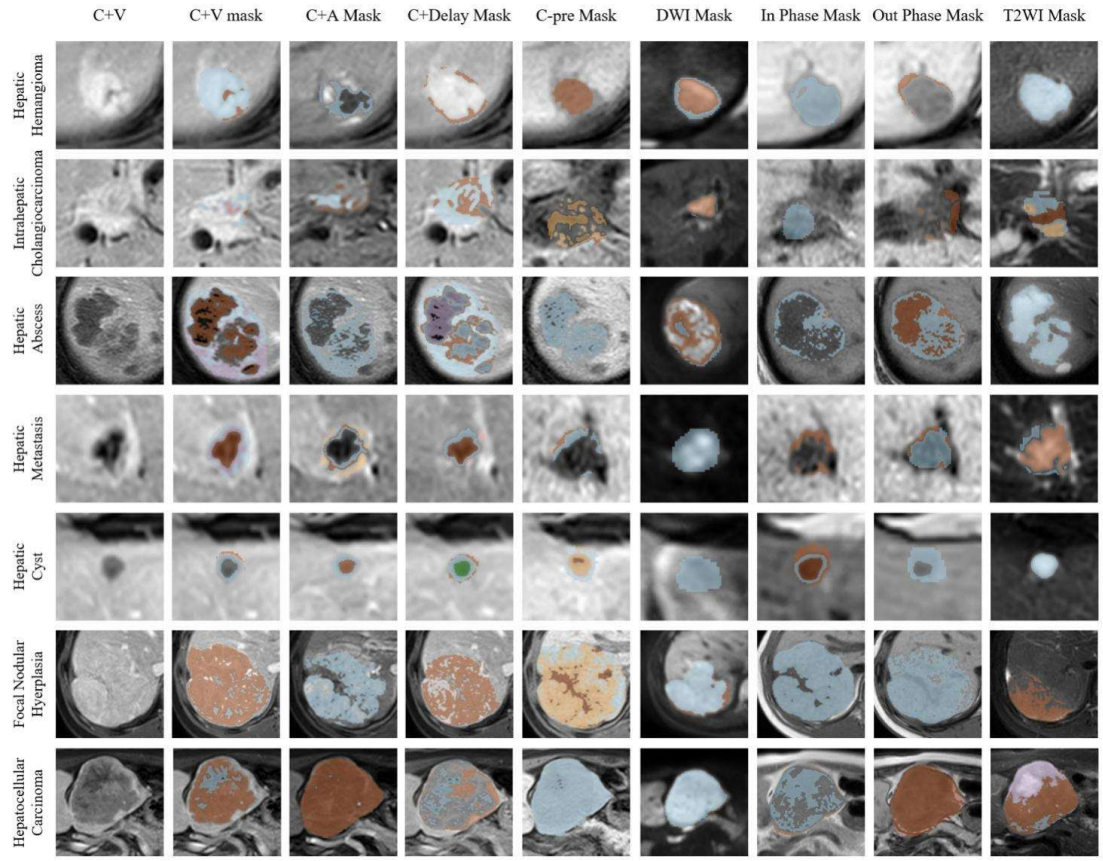


**Figure 4.4.** The original mp-MRI image and aligned pseudo sub-region labels of LLD-MMRI2023 for seven correctly predicted cases, organised by rows for each class, using the proposed HA-DLC framework. The modality with the pseudo sub-region labels is labelled in columns. The different colours in pseudo sub-region labels indicate different sub-region labels.

### 4.4.4. Evaluation of Aligned sub-region labels

Figure 4.3 illustrates the effectiveness of the HA-DLC framework in segmenting tumour sub-regions for two correctly classified patients from the BraTS 2021 Radiogenomic Classification dataset. The sub-regions in the ground truth masks are generally aligned with the pseudo sub-region labels generated by the clustering approach in T1Gd, which is used to segment sub-regions in glioblastoma. The variation in sub-region labels across different modalities indicate the model's ability to capture spatial and contrast-enhanced tumour characteristics. Figure 4.4 presents the segmentation results for seven correctly classified cases from LLD-MMRI 2023. The segmentation masks across modalities demonstrate consistent sub-region delineation, reflecting the framework's ability to adapt across diverse imaging sequences and anatomical structures.

## 4.5. Discussions

The HA-DLC demonstrated improved classification performance in inter- and intra-tumour heterogeneity classification tasks across two diverse MRI datasets. These results support the following key findings: (1) the proposed method outperformed its state-of-the-art counterpart in both heterogeneous classification datasets of MGMT and LLD-MMRI2023; (2) Sub-region segmentation enhanced classification performance compared to whole-tumour segmentation, demonstrating the importance of using sub-regions; (3) Pseudo sub-region labels generation strategies influenced sub-region granularity, in terms of partitioning the tumour into greater number of sub-regions compared to the ground truth, and thus leading to improved performance.

HA-DLC demonstrated high performance across two datasets that captured inter- and intra- tumour heterogeneity information for tumour type classification tasks (Table

4.1). This performance advantage suggests that using pseudo sub-region labels to extract local heterogeneity features for classification is meaningful.

In the ablation study result, the classification performance remained at the similar level when compared to the classification backbone only (Table 4.2). This suggests that the enhanced performance of HA-DLC does not intrinsically benefit from more parameters which are weights and bias in the networks. When the Modality-wise Segmentation Module was used to segment the entire tumour, its performance was higher compared to the results that was trained without the segmentation loss (Table 4.2). This improvement indicates that features extracted via the proposed hard-sharing branch in the segmentation network effectively enhance classification tasks by capturing localised tumour-related characteristics through segmentation. Incorporating the proposed pseudo sub-region labels and ground-truth sub-region labels further boosted performance compared to whole-tumour segmentation (Table 4.2). This suggests that pseudo labels can capture tumour heterogeneity more meaningfully for classification and act as a guide to the encoder branches in the modality-wise segmentation module, enabling the extraction of fine-grained local features from each sub-region. The increased performance indicates the efficacy of leveraging sub-region labels in obtaining heterogeneous features. The proposed method achieved AUC performance (0.671) comparable to the ground truth labels (0.681) in MGMT classification task, underscoring its potential to replace manually annotated segmentation (Table 4.2). The performance difference between the ground truth labels and the pseudo sub-region labels may be due to the alignment module merging regions with similar feature representations into a single area, as shown in the T1n mask in Figure 4.4. This can occur if the latent feature space learned by the model fails to capture the subtle variations between certain tissue types or sub-regions that human experts distinguish in the ground truth. These findings suggest that leveraging unsupervised semantic segmentation of GBM to generate sub-region inputs could

further enhance performance. However, the proposed method in previous chapter is limited to 2D input. Extending it to 3D and integrating it with the HA-DLC may alleviate the need for ground truth and achieve same level of performance.

The effect of the clustering method was evaluated based on the performance of the proposed framework by adjusting the parameters in the two algorithms, k-means and DFC. For the k-means, the number of centroids was changed to generate different granularity sub-region labels. The performance peaked at 5 cluster centres in LLD-MMRI2023 datasets (Table 4.3). This suggests that for the tumour-type classification task, a moderate level of granularity is necessary. For the BraTS 2021 Radiogenomic Classification task, the AUC peaks for 6 and 9 clusters, and the AUC maintained similar levels across 3 to 9 clusters (Table 4.3). This may be attributed to the focus on intra-tumour heterogeneity between two classes, where our alignment network consistently aggregates the generated sub-regions across cases. For the DFC method, the whole-image and tumour-ROI level clustering approaches were used to examine the impact of spatial constraints on the generated sub-regions. Under whole image level clustering, a lower continuity loss coefficient ($\mu$) resulted in an increased maximum number of clusters, with $\mu = 0.5$ yielding 17 clusters and a corresponding performance of 0.809 on LLD-MMRI2023. For MGMT classification, a further reduction to $\mu = 0.001$ resulted in 24 clusters and an improved AUC of 0.671. These results suggest that more sub-regions may capture finer local features that benefit molecular classification when considering the entire image. In contrast, when clustering was restricted to the tumour ROI, the maximum number of clusters was lower, reflecting the more constrained spatial domain. For the LLD-MMRI2023 dataset, the best performance was achieved with $\mu = 0.001$. In contrast, for the MGMT task, the differences across $\mu$ values were marginal, with the highest AUC occurring at $\mu = 2$ (Table 4.3). This indicates that, within the tumour ROI, overly fine segmentation does not necessarily translate into better performance, and a more continuous representation may be sufficient.

Figures 4.3 and 4.4 visualise each class's aligned pseudo-subregion labels for two selected patients in each dataset. The colour-coded cluster masks illustrate how our alignment strategy effectively maintains spatial correspondence across various contrast phases (e.g., arterial, venous, delayed) and mpMRI sequences (e.g., DWI, T2). Despite the inherent variability in lesion appearance across phases, the sub-regions remain well-aligned, indicating that our framework reliably captures meaningful anatomical structures. This underscores the efficacy of the heterogeneous sub-region pseudo label generation module in ensuring consistent delineation of each physical region of interest across different contrast weightings. In the BraTS 2021 Radiogenomic Classification dataset, although the exact cluster colours differ from the ground truth labels, the spatial distribution of clusters aligns closely with recognised tumour compartments in T1Gd images. Meanwhile, the sub-regions in FLAIR and T2 masks exhibit class-specific characteristics: for instance, the surrounding blue region may indicate a perfusion-related area potentially correlated with MGMT methylation status [38]. Such findings suggest that our framework can identify heterogeneous regions beyond standard human-defined labels. Nevertheless, given the performance gap between the generated masks and the ground-truth segmentation, further integration of ground-truth-based heterogeneity information may be beneficial.

Despite the promising results, the HA-DLC framework has several limitations. First, relying on unsupervised pseudo sub-region label generation can sometimes lead to a loss of granularity, as the alignment module may merge regions with similar feature representations, resulting in fewer sub-regions than those annotations delineated by experts. Future study will evaluate the impact from pseudo sub-regions alignment process. Additionally, the method’s performance is sensitive to the choice of clustering parameters (e.g., the number of centres in k-means or the continuity loss coefficient in DFC), which might not generalise well across different datasets or imaging modalities. This can potentially be overcome with automated parameter selection algorithms such

as Bayesian based optimisation. While effective in capturing both local and global features, the dual-stream architecture introduces increased computational complexity and higher memory requirements, which may limit its deployment in real-time clinical settings. This is mainly due to the modality-wise segmentation module having to extract local heterogeneity features from each sequence; exploring heterogeneity extraction approaches that leverage all modalities may help reduce the complexity of this module. Finally, although the proposed method has shown generalisability by working on two mpMRI datasets, its applicability to other types of medical imaging and tumour pathologies remains to be validated.

## 4.6. Conclusions

In this chapter, a novel heterogeneity-aware deep learning classification (HA-DLC) method was introduced that integrates unsupervised sub-region segmentation with global feature extraction to address both local and global aspects of tumour heterogeneity. By leveraging automatically generated pseudo sub-region labels, HA-DLC was able to capture the subtle intra-tumoral variations without relying on manual annotations. Extensive experiments on two distinct MRI datasets: LLD-MMRI2023 for inter-tumour heterogeneity and BraTS2021 Radiogenomic Classification for intra-tumour heterogeneity. The result demonstrated that HA-DLC consistently outperformed comparison methods. Further, ablation studies confirmed that incorporating sub-region segmentation significantly improved classification performance compared to using whole-tumour masks alone. Despite these promising results, the framework faces challenges such as parameter sensitivity, potential loss of fine-grained detail when merging similar sub-regions and increased computational overhead. Future work will investigate strategies to enhance sub-region fidelity, reduce

parameter dependencies, and extend the method to additional imaging modalities and tumour pathologies.

# Chapter 5. Conclusion and Future Works

## 5.1. Conclusion

This thesis addresses two main challenges in medical image tumour heterogeneity analysis: i) sub-region label acquisition depends on clinicians and ii) lack of a framework to capture local heterogeneity and global information in tumour image. The novel unsupervised semantic segmentation framework, characterised by its multi-phase training strategy and the incorporation of sparse spatial continuity and context-based hierarchical loss functions, has proven capable of delineating complex tumour sub-regions without requiring extensive manual annotation. In parallel, the heterogeneity-aware deep learning framework successfully combines local heterogeneity with global contextual information, enhancing tumour classification performance. The results show that deep learning-based approaches can effectively harness tumour heterogeneity from multiparametric MRI modalities to improve segmentation and classification tasks. Consequently, the proposed approaches offer a robust computational tool to improve diagnostic accuracy and support personalised clinical decision-making, ultimately advancing patient care.

## 5.2. Limitations

Despite these encouraging results, several limitations of the current study must be acknowledged. For the unsupervised semantic segmentation framework, the dependence on the whole tumour mask and 2D implementation constrains the ability to apply it in clinical settings. Also, the effectiveness of the unsupervised segmentation

framework relies on the quality of the pseudo-labels generated by clustering algorithms. In cases where these pseudo-labels fail to capture the subtle distinctions between tumour sub-regions, the segmentation accuracy may be adversely affected. For the heterogeneity-aware deep learning framework, the model's performance is influenced by the selection of clustering parameters, such as the number of cluster centres or the continuity loss coefficient. This sensitivity can restrict the method's robustness and generalisability across different datasets and imaging conditions. In addition, the pseudo-sub-region label generation is heavily based on clustering algorithms, limiting the segmentation result of its interpretability in a clinical context.

## 5.3. Future Works

Future research can focus on improving the clinical value of the two proposed methods. For the unsupervised semantic segmentation framework, adopting unsupervised whole tumour segmentation algorithms and extending the framework to operate on 3D volumetric directly would enable the framework to generate a 3D sub-region level mask without any human intervention. For the heterogeneity-aware deep learning framework, a more clinical context backgrounded pseudo-label generation should be considered to improve the framework's explainability and stability, for example, through acquiring a small amount of tissue-level labels for medical images and finetuning a foundation model that specifically segments sub-regions at the tissue level.

In addition, while we validated the proposed framework across multiple datasets, tumour heterogeneity remains challenging due to biological diversity and acquisition variability. Future work will (i) incorporate heterogeneity-aware objectives, including multi-instance/sub-region losses, diversity–consistency regularization, boundary- and uncertainty-aware terms, contrastive/consistency objectives across sequences, and distributionally robust optimization to better model intra-lesional variability and

domain shift; and (ii) broaden experimental validation to additional tumour types (e.g., brain, lung) and modalities (e.g., DCE-MRI, MR spectroscopy, CEUS, PET/CT), with multi-institutional external testing. We will also expand evaluation beyond aggregate accuracy to include calibration, subgroup, and sub-region stability metrics to more comprehensively assess generalization and clinical reliability."

# References


1. R. Fisher, L. Pusztai, and C. Swanton, “Cancer heterogeneity: implications for targeted therapeutics,” *British Journal of Cancer*, vol. 108, no. 3, pp. 479–485, Jan. 2013.
2. N. Upadhyay and A. D. Waldman, "Conventional MRI evaluation of gliomas," *British Journal of Radiology*, vol. 84, no. special_issue_2, pp. S107-S111, 2014.
3. Getaneh, A.M., Heijnsdijk, E.A.M. & de Koning, H.J. The comparative effectiveness of mpMRI and MRI-guided biopsy vs regular biopsy in a population-based PSA testing: a modeling study. *Sci Rep* **11**, 1801, 2021.
4. Wirsching, HG., Weller, M. (2017). Glioblastoma. In: Moliterno Gunel, J., Piepmeier, J., Baehring, J. (eds) *Malignant Brain Tumors*. *Springer, Cham*.
5. J. Fu and H. Wang, “Precision diagnosis and treatment of liver cancer in China,” *Cancer Letters*, vol. 412, pp. 283–288, Oct. 2017.
6. A. M. Rojiani and K. Dorovini-Zis, “Glomeruloid vascular structures in glioblastoma multiforme: an immunohistochemical and ultrastructural study,” *Journal of Neurosurgery*, vol. 85, no. 6, pp. 1078–1084, Dec. 1996.
7. Dagogo-Jack, I., Shaw, A. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **15**, 81–94, 2018.
8. R. Jain et al., ""“Real world” use of a highly reliable imaging sign: “T2-FLAIR mismatch” for identification of IDH mutant astrocytomas," *Neuro-Oncology*, vol. 22, no. 7, pp. 936-943, 2020.
9. D. Thi, M.-R. Yang, L. H. T. Lam, N. Q. K. Le, and Y. Wu, “Improving MGMT methylation status prediction of glioblastoma through optimizing radiomics features using genetic algorithm-based machine learning approach,” *Scientific Reports*, vol. 12, no. 1, Aug. 2022.

10. Li, L., Xiao, F., Wang, S. *et al.* Preoperative prediction of MGMT promoter methylation in glioblastoma based on multiregional and multi-sequence MRI radiomics analysis. *Sci Rep* **14**, 16031, 2024.
11. H. Zhang et al., "Sub-region based radiomics analysis for prediction of isocitrate dehydrogenase and telomerase reverse transcriptase promoter mutations in diffuse gliomas," *Clinical Radiology*, vol. 79, no. 5, pp. e682–e691, Feb. 2024.
12. K. He et al., "Transformers in medical image analysis," *Intelligent Medicine*, vol. 3, no. 1, pp. 59–78, Feb. 2023.
13. Alexey et al., "An image is worth 16x16 words: Transformers for image recognition at scale." *In International Conference on Learning Representations 2021 (ICLR)*, 2021
14. J. J. M. Van Griethuysen et al., "Computational Radiomics system to decode the radiographic phenotype," *Cancer Research*, vol. 77, no. 21, pp. e104–e107, Oct. 2017.
15. F. Ritter et al., "Medical image analysis," *IEEE Pulse*, vol. 2, no. 6, pp. 60–70, Nov. 2011.
16. H. Greenspan, "Super-Resolution in medical imaging," *The Computer Journal*, vol. 52, no. 1, pp. 43–63, Feb. 2008.
17. A. Elangovan and T. Jeyaseelan, "Medical imaging modalities: A survey," *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, pp. 1–4, Feb. 2016.
18. M. Rohrer, H. Bauer, J. Mintorovitch, M. Requardt, and H.-J. Weinmann, "Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths," *Investigative Radiolog*, vol. 40, no. 11, pp. 715–724, Oct. 2005.

19. W.-D. Heiss, P. Raab, and H. Lanfermann, "Multimodality assessment of brain tumors and tumor recurrence," *Journal of Nuclear Medicine*, vol. 52, no. 10, pp. 1585–1600, Aug. 2011.
20. S. Maniam, "Magnetic resonance imaging: Review of imaging techniques and overview of liver imaging," *World Journal of Radiology*, vol. 2, no. 8, p. 309, Jan. 2010.
21. D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 315–337, Aug. 2000.
22. L. Zhang et al., "Disentangling Human Error from Ground Truth in Segmentation of Medical Images," *Neural Information Processing Systems*, vol. 33, pp. 15750–15762, Jul. 2020.
23. K. K. D. Ramesh, G. Kumar, K. Swapna, D. Datta, and S. Rajest, "A review of Medical Image Segmentation Algorithms," *EAI Endorsed Transactions on Pervasive Health and Technology*, p. 169184, Jul. 2018.
24. O. Ronneberger, P. Fischer, and T. Brox, "U-NET: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Jan. 2015.
25. A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *International Conference on Learning Representations 2021 (ICLR)*, Jan. 2020.
26. S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.
27. M. E. Hegi et al., "MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma," *New England Journal of Medicine*, vol. 352, no. 10, pp. 997–1003, Mar. 2005.

28. A. A. Brandes et al., "Recurrence pattern after temozolomide concomitant with and adjuvant to radiotherapy in newly diagnosed patients with glioblastoma: correlation with MGMT promoter methylation status," *Journal of Clinical Oncology*, vol. 27, no. 8, pp. 1275–1279, Feb. 2009.
29. A. M. Tsimberidou, E. Fountzilas, M. Nikanjam, and R. Kurzrock, "Review of precision cancer medicine: Evolution of the treatment paradigm," *Cancer Treatment Reviews*, vol. 86, p. 102019, Mar. 2020.
30. F. Jozsa et al., "Safety and efficacy of brain biopsy: Results from a single institution retrospective cohort study," *Brain and Spine*, vol. 3, p. 101763, Jan. 2023.
31. H. B. Thomaides-Brears et al., "Incidence of Complications from Percutaneous Biopsy in Chronic Liver Disease: A Systematic Review and Meta-Analysis," *Digestive Diseases and Sciences*, vol. 67, no. 7, pp. 3366–3394, Jun. 2021.
32. Z. Makowska et al., "Gene expression analysis of biopsy samples reveals critical limitations of transcriptome‐based molecular classifications of hepatocellular carcinoma," *The Journal of Pathology Clinical Research*, vol. 2, no. 2, pp. 80–92, Jan. 2016.
33. F. Ritter et al., "Medical Image Analysis," *IEEE Pulse*, vol. 2, no. 6, pp. 60-70, Nov.-Dec. 2011.
34. K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Physics and Technology*, vol. 10, no. 3, pp. 257–273, Jul. 2017.
35. J. J. M. Van Griethuysen et al., "Computational Radiomics system to decode the radiographic phenotype," *Cancer Research*, vol. 77, no. 21, pp. e104–e107, Oct. 2017.
36. S. Rathore et al., "Non-invasive determination of the O6-methylguanine-DNA-methyltransferase (MGMT) promoter methylation status in glioblastoma (GBM) using magnetic resonance imaging (MRI).," *Journal of Clinical Oncology*, vol. 36, no. 15_suppl, p. 2051, May 2018.

37. P. Lambin et al., "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, Jan. 2012.
38. C. H. Suh, H. S. Kim, S. C. Jung, C. G. Choi, and S. J. Kim, "Clinically Relevant Imaging Features for MGMT Promoter Methylation in Multiple Glioblastoma Studies: A Systematic Review and Meta-Analysis," *American Journal of Neuroradiology*, Jul. 2018.
39. G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Jul. 2017.
40. C. Militello et al., "On Unsupervised Methods for Medical Image Segmentation: Investigating Classic Approaches in Breast Cancer DCE-MRI," *Applied Sciences*, vol. 12, no. 1, p. 162, Dec. 202.
41. C. N. Rodrigues, I. M. Nunes, M. B. Pereira, H. Oliveira, and J. a. D. Santos, "From superpixels to foundational models: An overview of unsupervised and generalizable image segmentation," *Computers & Graphics*, vol. 123, p. 104014, Jul. 2024.
42. M. Prakram, K. Rawal, and M. Prakram, "Unsupervised Learning Based Medical Image Segmentation: A Comparative Review of Algorithms with Issues and Challenges," *SSRN Electronic Journal*, Jan. 202.
43. K. Raza and N. K. Singh, "A tour of unsupervised deep learning for medical image analysis," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 17, no. 9, pp. 1059–1077, Jan. 2021.
44. M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," *European Conference on Computer Vision 2018 (ECCV)*, 2018.
45. W. Kim, A. Kanezaki, and M. Tanaka, "Unsupervised learning of image segmentation based on differentiable feature clustering," *IEEE Transactions on Image Processing*, vol. 29, pp. 8055–8068, Jan. 2020.

46. X. Jiang, Z. Hu, S. Wang, and Y. Zhang, "Deep Learning for Medical Image-Based Cancer diagnosis," *Cancers*, vol. 15, no. 14, p. 3608, Jul. 2023.
47. A. Hosny, H. J. Aerts, and R. H. Mak, "Handcrafted versus deep learning radiomics for prediction of cancer therapy response," *The Lancet Digital Health*, vol. 1, no. 3, pp. e106–e107, Jun. 2019.
48. M. Reyes et al., "On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities," *Radiology Artificial Intelligence,* vol. 2, no. 3, p. e190043, May 2020.
49. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Computer Vision and Pattern Recognition 2017 (CVPR)*, 2017.
50. A. Vaswani et al., "Attention is All you Need," *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017.
51. K. Li et al., " UniFormer: Unified Transformer for efficient Spatiotemporal Representation learning," *International Conference on Learning Representations 2022(ICLR)*, 2022.
52. M. Zhang and C. Desrosiers, "High-quality Image Restoration Using Low-Rank Patch Regularization and Global Structure Sparsity," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 868-879, 2019.
53. U. Baid et al., "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification," *arXiv preprint*, 2021.
54. Z. Yu et al., "Cross-grained Contrastive Representation for Unsupervised Lesion Segmentation in Medical Images," *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2023.
55. M. Lou, H. Ying, X. Liu, H.-Y. Zhou, Y. Zhang, and Y. Yu, "SDR-Former: A Siamese Dual-Resolution Transformer for liver lesion classification using 3D multi-phase imaging," *Neural Networks*, vol. 185, p. 107228, Feb. 2025.

56. J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, Jan. 2024, doi: 10.1038/s41467-024-44824-z.
57. F. BABA, "Leaderboard of RSNA-MICCAI Brain Tumor Radiogenomic Classification," *Kaggle*.
58. M. Hu et al., "MGMT promoter methylation prediction based on multiparametric MRI via vision graph neural network," *Journal of Medical Imaging*, vol. 11, no. 01, Feb. 202.
59. Ross Wightman. *PyTorch Image Models*. https://github.com/rwightman/pytorch-imagemodels. 2019.
60. K. He, X. Zhang, Xiangyu, S. Ren, and J. Sun, "Deep residual learning for image recognition," In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016.
61. M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," In *Proceedings of the International Conference on Machine Learning (ICML)*, 2019.