

UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DEPARTMENT OF INFORMATION ENGINEERING
MASTER'S DEGREE IN ICT FOR INTERNET AND MULTIMEDIA

AI DEVELOPMENT ON RECTAL CANCER BY USING MEDICAL IMAGING

Supervisor

Prof. LORIS NANNI

Graduating

AMIRMOHAMMAD FARMAN

Co-supervisor

Prof. FILIPPO CRIMI

ACADEMIC YEAR 2024-2025

Date of graduation 27/02/2025

I would like to convey my deep gratitude to Prof. Loris Nanni for his exceptional guidance and steadfast support throughout my academic journey. His mentorship has played a pivotal role in shaping my development and knowledge. I am also profoundly thankful to Prof. Filippo Crimi, my co-supervisor, for his valuable insights, and to Marco Benozzi, whose mentorship during my internship and assistance in securing this project were truly significant.

I express my sincere thanks to the University of Padova for fostering an enriching academic environment that has enabled me to grow both personally and professionally.

From the depths of my heart, I wish to extend my warmest appreciation to my partner and family for their unwavering encouragement, patience, and faith in me. Lastly, I am immensely thankful to my friends, whose support and companionship have brightened my path.

This is not the END...

Abstract

Colorectal cancer (CRC) is a major global health issue and a leading cause of cancer-related deaths. Early and accurate diagnosis is crucial for improving patient outcomes, yet current methods for analyzing MRI scans—a key tool for tumor detection and staging—often rely on manual segmentation, a process that is time-consuming and prone to variability between observers.

This thesis presents a new, AI-driven framework that integrates advanced deep learning techniques with radiomics analysis to enhance the identification and characterization of CRC tumors from MRI scans. We introduce and evaluate three innovative deep learning models, 2D-UNet, 2D-UMamba, and 2D-Swin-UMamba, that automatically segment tumors on MRI images. The models are trained and tested on a carefully curated dataset of rectal cancer patients, ensuring robust performance across diverse imaging conditions.

After segmentation, radiomics is used to extract quantitative features that describe tumor shape, texture, and intensity. These features are analyzed for correlations with clinical factors such as tumor stage, microsatellite instability (MSI), and genetic mutations (KRAS/NRAS/BRAF), offering a more comprehensive and quantitative assessment of CRC tumors. This combined approach has the potential to support more personalized treatment planning and improved patient care.

Our experimental results demonstrate that the proposed AI models outperform traditional segmentation methods, achieving higher accuracy as measured by the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU). Furthermore, the radiomics analysis indicates that AI-driven feature extraction can identify imaging biomarkers predictive of clinical outcomes, potentially reducing the need for invasive diagnostic procedures. In general, this research contributes to a scalable and clinically relevant framework for automated MRI-based segmentation and radiomic analysis in the management of colorectal cancer.

Keywords: Colorectal cancer, MRI, Deep Learning, Segmentation, Radiomics, 2D-U-Net, 2D-UMamba, 2D-Swin-UMamba, PVTv2, HSN, Tumor Characterization, AI in Medical Imaging.

Contents

1	Introduction	1
1.1	Overview	1
1.2	Colorectal Cancer Background	2
1.2.1	Epidemiology	2
1.2.2	Etiology and Risk Factors	3
1.2.3	Pathogenesis	4
1.2.4	Clinical Presentation and Diagnosis	5
1.2.5	Staging	6
1.2.6	Therapeutic Approaches	7
1.3	Medical Imaging in Colorectal Cancer	9
1.3.1	MRI for Colorectal Cancer	9
1.3.2	Segmentation Challenges	10
1.4	Radiomics in Colorectal Cancer	11
1.4.1	Definition and Process	11
1.4.2	Applications	13
1.5	Deep Learning for Automated MRI Segmentation	15
1.5.1	Convolutional Neural Networks (CNNs)	15
1.5.2	The UNet Architecture	15
1.5.3	2D-UNet, 2D-UMamba, and 2D-Swin-UMamba	15
1.6	Objectives	16
1.7	Literature Review	17
2	Mathematical Foundations	23
2.1	Introduction	23
2.1.1	Convolutional Neural Networks (CNNs)	24
2.1.2	Pooling Layers	25
2.1.3	Batch Normalization	26
2.2	Segmentation Architectures	26

2.2.1	U-Net	27
2.2.2	UMamba	29
2.2.3	Swin-UMamba	31
2.2.4	PVTv2-Based Encoder Architecture	33
2.2.5	Hybrid Segmentation Network (HSN)	35
2.3	Loss Function: BCE With Logits	36
2.4	Evaluation Metrics: Dice and IoU	37
2.4.1	Dice Coefficient	38
2.4.2	Intersection-over-Union (IoU)	38
2.5	Data Augmentation and Preprocessing	39
2.6	Optimization: Adam	40
2.6.1	Algorithm	40
2.6.2	Training Configuration	41
2.7	Summary	42
3	Dataset, MRI Acquisition, and Preprocessing	43
3.1	Introduction	43
3.2	Dataset Description	43
3.3	Radiomics-Based Feature Extraction	45
3.4	MRI Acquisition Protocols	46
3.5	Data Preprocessing and Ground Truth Labeling	46
3.5.1	Initial Segmentation Using 3D Slicer	46
3.5.2	Visualization of MRI Segmentation and Preprocessing	46
3.6	Data Augmentation and Preprocessing	48
3.7	Slice Extraction and Dataset Partitioning	49
3.8	Summary	50
4	Deep Learning Models for Medical Image Segmentation	53
4.1	Introduction	53
4.1.1	Implementation Details Consistent with the Code	54
4.2	UNet with Hybrid Segmentation Network (HSN)	54
4.2.1	Key Components	54
4.2.2	Impact on Colorectal Cancer Segmentation	55
4.3	UMamba with PVTv2 (Pyramid Vision Transformer v2)	55
4.3.1	Model Architecture	56
4.4	Swin with PVTv2: Merging Local Windowed Attention and Multi-Scale Analysis	57
4.5	Deep Learning and Radiomics Synergy	58

4.6	Summary	58
5	Results and Experiments	61
5.1	Introduction	61
5.2	Experimental Setup	61
5.2.1	Dataset and Preprocessing	61
5.2.2	Training Details	62
5.3	Performance Evaluation	63
5.3.1	Quantitative Analysis	63
5.3.2	Loss and Convergence Trends	63
5.4	Qualitative Analysis	64
5.5	Radiomics Feature Analysis	65
5.5.1	Radiomics Feature Extraction Results	65
5.5.2	Clinical Relevance of Radiomics Features	66
5.6	Discussion	67
5.6.1	Model Performance Interpretation	67
5.6.2	Synergy Between Radiomics and Deep Learning	68
5.6.3	Future Directions	68
5.7	Limitations and Challenges	68
5.8	Summary	69

List of Figures

2.1	Block diagram of 2D CNN architecture	25
2.2	Max Pooling process	25
2.3	The U-net architecture	28
2.4	U-Mamba architecture.	31
2.5	The overarching framework of Swin-UMamba.	32
2.6	Overall architecture of Pyramid Vision Transformer (PVT).	34
2.7	Architecture of HSNet.	35
3.1	Segmented MRI scan highlighting the mesorectum (beige) and tumor (green) to improve colorectal cancer detection using deep learning techniques.	44
3.2	MRI-based tumor and mesorectum segmentation, essential for colorectal cancer analysis, facilitating automated medical image interpretation with AI models.	44
3.3	MRI images with corresponding segmentation masks overlaying tumor and mesorectum regions, illustrating deep learning-based colorectal cancer segmentation	48
5.1	UNet-HSN training and validation loss (left) and accuracy (right) trends.	63
5.2	UMamba-PVTv2 training and validation loss (left) and accuracy (right) trends.	64
5.3	Swin-PVTv2 training and validation loss (left) and accuracy (right) trends.	64

List of Tables

3.1	Dataset Overview and Key Variables	45
5.1	Final Model Performance	63
5.2	Intensity-Based Features	65
5.3	Texture-Based Features (GLCM)	65
5.4	Morphological and Shape-Based Features	66
5.5	Wavelet-Based Features	66

Chapter 1

Introduction

1.1 Overview

Colorectal cancer (CRC) is among the most common and deadly malignancies worldwide. The Global Cancer Observatory data indicate that CRC accounts for approximately 10% of all diagnosed cancers and 9.4% of all cancer-related deaths [29]. While there have been improvements in screening and treatment, CRC remains a major health challenge, partly due to its heterogeneous nature and potentially asymptomatic early stages.

Imaging plays a crucial role in the management of CRC, from early detection and screening to staging and follow-up. Recent advancements in medical imaging and artificial intelligence (AI) have expanded opportunities to analyze these images more quantitatively. In particular, *radiomics* and *deep learning-based segmentation* have shown great potential for characterizing both the visible and latent features of tumors.

This thesis focuses on the development and validation of deep learning models—2D-UNet, 2D-UMamba, and 2D-Swin-UMamba—to segment colorectal cancer lesions on Magnetic Resonance Imaging (MRI) and then apply radiomics analysis on the segmented regions. Additionally, the study incorporates advanced architectural components, including Pyramid Vision Transformer v2 (PVTv2) for enhanced feature extraction and Hybrid Spatial Normalization (HSN) for improved spatial representation, both of which contribute to more precise and robust segmentation results. This chapter provides an overview of CRC’s epidemiology, pathogenesis, imaging approach, and how radiomics plus deep learning segmentation fit into the modern workflow of CRC research.

Over the last decade, the rising global burden of CRC has prompted the medical community to seek more efficient and accurate diagnostic tools. Traditional detection methods—such as colonoscopies, biopsy for histopathological confirmation, and standard imaging techniques—remain central to clinical practice. However, these approaches often rely heavily on the subjec-

tive judgment of clinicians, which can lead to variations in diagnosis and staging accuracy. The integration of AI-driven analysis has emerged as a means to reduce subjectivity, standardize evaluations, and improve diagnostic confidence.

From a technical standpoint, deep learning architectures, particularly convolutional neural networks (CNNs), have excelled at image segmentation tasks by automatically learning hierarchical feature representations. However, challenges such as inter-patient variability, motion artifacts in MRI, and differences in scanner settings can complicate segmentation outcomes. Advanced transformer-based models, which incorporate attention mechanisms, address some of these issues by focusing on relevant regions within the image and capturing long-range dependencies. This is where PVTv2 comes into play, offering a more effective way of extracting multi-scale features without overburdening computational resources. Meanwhile, Hybrid Spatial Normalization techniques aim to stabilize feature distribution across varying spatial domains, enhancing the model's generalizability.

Once lesions are accurately segmented, radiomics can be applied to quantify a wide array of features—such as texture, shape, and intensity—that go beyond what the naked eye can discern. These features can then be correlated with histological subtypes, clinical outcomes, or treatment responses. In doing so, radiomics bridges the gap between imaging and personalized medicine, enabling more nuanced decision-making that factors in the unique characteristics of each patient's tumor. Ultimately, the synergy of deep learning and radiomics has the potential to streamline diagnostic workflows, reduce human error, and enable the discovery of novel biomarkers that may guide therapeutic strategies.

1.2 Colorectal Cancer Background

1.2.1 Epidemiology

Colorectal cancer is the third most common cancer in men and the second in women worldwide [29, 25]. According to GLOBOCAN 2020 estimates, there are nearly 1.9 million new CRC cases diagnosed annually and more than 900,000 deaths globally [31]. The incidence of CRC increases with age, particularly after the fifth decade of life. However, there is a concerning rise in early-onset CRC (patients under 50) in many countries [28].

Geographical variation in CRC incidence and mortality is influenced by socio-economic factors, lifestyle, and access to screening. Developed regions such as North America, Europe, and parts of Oceania show high CRC incidence, whereas many developing areas have seen rising rates following the adoption of Western lifestyles [2].

An important aspect of CRC epidemiology lies in its changing patterns across different socio-economic strata. While traditionally higher in wealthier nations, improvements in public health

measures, dietary guidelines, and routine screening programs have led to relatively stable or even decreasing CRC rates in some developed countries. In contrast, low- and middle-income regions are experiencing a surge in CRC incidence. The rise is partly attributed to increased life expectancies, the infiltration of Western eating habits (high in processed sugars, fats, and red meats), and lower engagement with preventive healthcare.

Healthcare systems are also grappling with the economic impact of CRC. The costs associated with chemotherapy, targeted therapies, surgical interventions, and long-term follow-up can be substantial. As patients live longer due to therapeutic advances, the burden of supportive care—both financial and emotional—also grows. This underscores the importance of well-structured national screening programs, which can detect polyps early or even prevent malignant transformation, thereby reducing overall treatment costs and improving patient quality of life.

The trend of early-onset CRC is of particular concern. Young patients often present with more advanced disease at diagnosis, possibly because they do not qualify for routine screenings at earlier ages and may ignore initial symptoms. Raising public awareness about risk factors and warning signs, along with updating screening guidelines, could mitigate the upward trend in early-onset cases. Ultimately, continued research in epidemiological patterns helps public health authorities refine guidelines, focusing on the most vulnerable populations and optimizing resource allocation.

1.2.2 Etiology and Risk Factors

CRC is a multifactorial disease shaped by genetic predisposition, personal history, and lifestyle choices. Up to 25% of CRC cases have a familial component; of these, about 3–5% are driven by hereditary CRC syndromes such as Lynch syndrome (Hereditary Non-Polyposis Colorectal Cancer, HNPCC) and Familial Adenomatous Polyposis (FAP) [23, 4].

Environmental and lifestyle factors that increase the risk of CRC include:

- **Dietary habits:** High consumption of red/processed meats and low dietary fiber
- **Obesity and sedentary lifestyle**
- **Smoking and heavy alcohol use**
- **Long-standing inflammatory bowel disease**

Understanding the intricate balance between genetic and environmental influences is essential to formulating targeted preventive strategies. Although individuals with hereditary syndromes like FAP or Lynch syndrome have a well-defined molecular basis for disease onset, the larger group of sporadic CRC cases often arises through the cumulative effect of multiple

risk factors. This reality highlights the importance of modifiable lifestyle components—diet, physical activity, and body weight management—as they provide a tangible opportunity for intervention.

The gut microbiome has also emerged as a noteworthy area of investigation. Diets rich in fiber tend to support a healthy gut flora, producing short-chain fatty acids that can exert protective effects against malignant transformation. Conversely, excessive intake of red or processed meats can alter the gut microbiota in ways that promote inflammation and carcinogenesis. Smoking and heavy alcohol consumption introduce additional mutagenic factors, compounding genetic mutations that may already be present due to inherited predisposition or sporadic errors in DNA replication.

Individuals with chronic inflammatory conditions such as ulcerative colitis or Crohn's disease live with persistent mucosal irritation, which can accelerate the adenoma-to-carcinoma sequence. For these patients, strict surveillance protocols are often recommended to catch dysplastic changes at an earlier stage. Ultimately, a holistic approach that integrates genetic counseling, lifestyle modifications, and vigilant surveillance can greatly reduce the overall disease burden. As healthcare systems move toward precision medicine, identifying high-risk profiles based on a combination of genetic markers and lifestyle factors will be key in customizing prevention and screening strategies.

1.2.3 Pathogenesis

Most CRCs arise from precursor lesions over a prolonged period (10–15 years). The transformation from normal colonic epithelium to carcinoma typically follows one of the following molecular pathways [17]:

1. **Chromosomal Instability (CIN) Pathway:** Characteristic of sporadic CRC and familial adenomatous polyposis (FAP). This pathway is initiated frequently by *APC* tumor suppressor loss, followed by activating mutations in *KRAS*, loss of other tumor suppressors (e.g., *TP53*), leading to polyp progression and carcinoma.
2. **Microsatellite Instability (MSI) Pathway:** Mutation or methylation-based silencing of DNA mismatch repair (MMR) genes (e.g., *MLH1*, *MSH2*). Lynch syndrome exemplifies a germline-mediated MSI. Many MSI CRCs show better prognosis and may respond differently to immunotherapies.
3. **Hypermethylation/CpG Island Methylator Phenotype (CIMP) Pathway:** Commonly associated with serrated polyps, with *BRAF* mutations frequently involved. This pathway involves widespread hypermethylation of promoter CpG islands, leading to silencing of key genes.

The multi-step nature of CRC development underscores why routine screening and early polyp removal are so effective. Adenomatous polyps, which may initially be benign, can accumulate further mutations over time, eventually breaching the basement membrane to become invasive carcinomas. In the CIN pathway, the sequential loss or mutation of key genes like APC, KRAS, and TP53 exemplifies how normal regulatory mechanisms collapse in favor of unchecked cellular proliferation. Detecting and removing these polyps before they acquire additional mutations can disrupt the malignancy cascade.

Microsatellite instability (MSI) provides an intriguing window into the genetic maintenance machinery. Tumors characterized by MSI exhibit a high mutational burden because of impaired DNA mismatch repair. This high burden can be both a challenge and an opportunity. On one hand, it can drive rapid tumor progression; on the other, it often elicits a strong immune response, making MSI tumors more susceptible to immunotherapeutic agents that harness the body's own defenses against cancer cells. Such insights have changed clinical practice, leading to the recommendation that many CRC patients be tested for MSI status to guide therapy choices.

In the CIMP pathway, epigenetic alterations like promoter hypermethylation inactivate tumor suppressor genes, effectively silencing their ability to regulate cell growth. These epigenetic events can occur earlier in the carcinogenic process than some genetic mutations, raising the possibility that interventions targeting aberrant methylation patterns might hold promise in preventing or slowing tumor progression. The serrated polyp, common in this pathway, presents a different morphological route to malignancy compared to classic adenomas. Recognizing such phenotypic differences is crucial for pathologists and gastroenterologists in tailoring screening and surveillance programs.

Overall, the diverse molecular underpinnings of CRC illustrate why the disease presents such varied clinical outcomes. Some tumors progress slowly and respond well to certain therapies, while others have an aggressive course and require more intensive treatment strategies. By studying these molecular pathways, researchers and clinicians can continue refining diagnostic markers, prognostic indicators, and therapeutic regimens to improve patient survival and quality of life.

1.2.4 Clinical Presentation and Diagnosis

Early stages of CRC often present no or minimal symptoms, which is why screening programs (colonoscopy, stool-based tests, or imaging-based approaches) are crucial to detect precursor lesions and early cancers [5]. Symptomatic disease may manifest as changes in bowel habits, rectal bleeding, iron-deficiency anemia, or abdominal pain.

Diagnosis commonly involves colonoscopic evaluation, biopsy, and imaging studies (e.g., computed tomography, MRI) to determine local extension and metastases [3].

In many patients, the subtlety of early CRC symptoms often leads to a delay in clinical detection, underlining the importance of robust screening measures. Although colonoscopy remains the gold standard for visualizing the mucosal surface and identifying precancerous lesions (polyps), adjunct methods have grown increasingly sophisticated. Stool-based tests, such as fecal immunochemical tests, continue to improve in sensitivity and specificity, making them viable primary screening tools in some healthcare settings. Imaging-based approaches, including CT colonography (often called “virtual colonoscopy”), offer an alternative for patients unwilling or unable to undergo standard colonoscopy.

Recent technological advances in endoscopic equipment have further enhanced diagnostic accuracy. High-definition endoscopes and electronic chromoendoscopy allow for better visualization of subtle mucosal changes, whereas novel imaging techniques can characterize vascular patterns and tissue architecture in real time. Some centers employ artificial intelligence (AI) algorithms that highlight suspicious areas during colonoscopy, potentially improving polyp detection rates. Though these algorithms are still being refined, preliminary data suggest they could standardize and streamline the diagnostic process.

Alongside endoscopic methods, MRI and CT scans help clinicians stage local disease and search for metastatic spread. MRI is particularly useful for assessing rectal tumors, as it can provide detailed images of soft tissue planes and the relationship of the tumor to surrounding structures. In certain cases, additional imaging modalities—such as positron emission tomography (PET)—may be used to identify distant metastases or recurrent disease post-treatment.

Biopsy confirmation remains the definitive step for diagnosis. Tissue samples help determine tumor histology, grade, and molecular markers, which increasingly guide individualized treatment. For instance, some centers incorporate liquid biopsy techniques, analyzing circulating tumor DNA to detect specific mutations or predict therapy resistance. Although not universally adopted, these emerging tools show promise in refining early diagnosis and monitoring disease progression or relapse.

Collectively, these advancements in screening and diagnostic protocols aim to identify CRC at an earlier stage, when curative interventions are more likely. Increased public awareness, coupled with accessible and less invasive testing options, could significantly reduce the burden of late-stage diagnoses that contribute to higher morbidity and mortality rates.

1.2.5 Staging

The *Tumor-Node-Metastasis* (TNM) staging provided by the American Joint Committee on Cancer (AJCC) is the gold standard. T describes the local infiltration depth, N the extent of regional lymph node involvement, and M the presence of distant metastases [1]. Staging guides therapeutic decisions and carries prognostic information.

Accurate staging is paramount in managing CRC because it helps tailor treatment strategies and predict patient outcomes. In practice, clinicians combine endoscopic findings, radiological imaging, and pathological evaluation of resected tissues to assign a TNM stage. This classification system is continually refined to incorporate new findings about tumor biology, local extension patterns, and the significance of micrometastases in regional lymph nodes.

Imaging techniques play a central role in staging. High-resolution CT scans of the chest, abdomen, and pelvis are routinely performed to evaluate potential spread to the liver or lungs—common metastatic sites for CRC. MRI is especially valuable for rectal tumors, as it can delineate mesorectal fascia involvement and vascular invasion with high accuracy. Such detail is crucial for deciding on neoadjuvant therapy and surgical planning, particularly when sphincter preservation is a consideration.

Recent efforts focus on identifying additional prognostic markers beyond the TNM classification. For instance, some centers measure the depth of tumor invasion in millimeters rather than grouping it into broad T categories, potentially revealing more nuances in disease aggressiveness. Similarly, the number of lymph nodes examined during surgery has been linked to survival outcomes, emphasizing the importance of adequate surgical sampling. In rectal cancer, the circumferential resection margin (CRM) and extramural vascular invasion are other key factors that guide treatment decisions.

Furthermore, molecular and genomic data are being integrated to refine prognosis within traditional staging tiers. Although TNM remains the foundational framework, tumors exhibiting certain molecular characteristics—like microsatellite instability or specific gene mutations—may behave differently or respond more favorably to targeted treatments. Incorporating such details into a more holistic staging approach could improve risk stratification and personalize therapy.

Ultimately, the accuracy and depth of staging significantly influence long-term survival and quality of life. By combining traditional pathological examination with cutting-edge imaging and molecular diagnostics, clinicians can tailor therapies more precisely, leading to better patient outcomes and more efficient healthcare resource utilization.

1.2.6 Therapeutic Approaches

Treatment for early-stage CRC often involves surgical resection. Adjuvant or neoadjuvant chemotherapy (and radiotherapy in rectal cancer) is given based on the tumor stage and risk features [24]. Targeted therapies (e.g., anti-EGFR) and immunotherapy (e.g., anti-PD-1 antibodies) are increasingly used, particularly guided by mutational profiles such as KRAS/NRAS/BRAF and MMR status [8].

The management of CRC has evolved considerably, with treatments becoming more precise

and tailored to the individual patient's tumor biology. In early-stage disease, surgical resection alone can be curative, especially for localized colon cancers. Advances in minimally invasive surgery, including laparoscopic and robotic-assisted procedures, have also made it possible to achieve similar oncological outcomes with reduced patient morbidity. Enhanced recovery after surgery (ERAS) protocols aim to shorten hospital stays and improve postoperative recovery, using strategies such as optimized pain control and early mobilization.

When tumors extend beyond the bowel wall or involve regional lymph nodes, adjuvant chemotherapy is generally recommended to eradicate microscopic disease and lower recurrence rates. In rectal cancer, neoadjuvant chemoradiotherapy has become standard practice for locally advanced lesions, allowing for tumor downstaging and potential sphincter-preserving surgery. Recent approaches explore total neoadjuvant therapy, where both chemotherapy and chemoradiotherapy are administered before surgery to maximize tumor response.

Targeted therapies mark a significant leap in CRC treatment by interfering with specific molecules responsible for tumor growth and progression. Anti-EGFR agents, for instance, can be effective in tumors without KRAS/NRAS mutations, helping to control disease in advanced settings. Other targeted drugs may inhibit vascular endothelial growth factor (VEGF), restricting the tumor's ability to form new blood vessels. These specialized treatments often extend survival, though resistance mechanisms can emerge, prompting ongoing research into combination regimens or next-generation inhibitors.

Immunotherapy represents another crucial frontier in CRC therapy, particularly for those with microsatellite instability. Immune checkpoint inhibitors, such as anti-PD-1 antibodies, can harness the body's immune system to recognize and attack cancer cells. In a subset of highly mutated tumors, this approach has demonstrated promising outcomes, occasionally leading to durable responses. Ongoing trials investigate immunotherapy's potential in earlier disease stages, as well as in combination with chemotherapy or radiation.

For metastatic CRC, multidisciplinary care is essential. Hepatic or pulmonary metastases may be surgically resectable in selected cases, offering a chance for prolonged survival or even cure. In other patients, a palliative strategy combining systemic therapies, local ablation techniques, and supportive care may enhance quality of life and extend survival. As scientists unravel more about CRC's molecular heterogeneity, personalized treatment regimens are expected to become standard, optimizing efficacy while minimizing unnecessary toxicity.

1.3 Medical Imaging in Colorectal Cancer

1.3.1 MRI for Colorectal Cancer

While computed tomography (CT) is often used for overall staging and detection of distant metastases, magnetic resonance imaging (MRI) has emerged as a powerful tool for local staging—especially in rectal cancer. MRI can provide excellent soft-tissue contrast, detailed visualization of rectal wall layers, mesorectal fascia, and sphincter complex, all of which are essential for surgical planning [14].

In addition, recent improvements in MRI protocols—such as high-resolution T2-weighted sequences and diffusion-weighted imaging (DWI)—enable better delineation of tumor margins, infiltration of mesorectal fat, and detection of involved lymph nodes. These findings guide precision therapy, such as total mesorectal excision or neoadjuvant chemoradiation [9].

MRI's ability to distinguish soft-tissue interfaces with high clarity makes it indispensable for assessing local tumor invasion. When evaluating rectal tumors, for instance, radiologists can inspect the depth of tumor penetration into or beyond the muscularis propria, helping to stratify patients for neoadjuvant treatments. High-resolution T2-weighted images provide a sharp view of the different layers of the rectal wall, which is critical for determining if the tumor is confined or has extended into the perirectal fat. Meanwhile, DWI offers functional insights into tissue cellularity and water diffusion, often revealing tumor boundaries that are not readily apparent on traditional sequences.

Beyond these core MRI sequences, specialized techniques such as contrast-enhanced studies can highlight vascularity and perfusion patterns in the tumor microenvironment, thereby indicating aggressiveness and potential response to therapy. In some centers, multiparametric MRI protocols combine T2-weighted, DWI, and dynamic contrast-enhanced imaging to yield a more comprehensive tumor assessment. This integrated approach can improve detection of smaller lesions and lymph node metastases, which are pivotal factors for treatment planning.

Moreover, MRI is increasingly utilized in the follow-up setting to evaluate treatment response or detect early recurrence. Post-therapy changes, such as fibrosis or edema, can sometimes mimic residual disease on other imaging modalities. However, advanced MRI sequences can better differentiate these tissue characteristics, guiding clinicians in deciding whether further intervention is necessary. As more research investigates the correlation between MRI-defined response criteria and histopathological outcomes, there is potential for MRI to become the mainstay imaging modality not just for diagnosis but also for ongoing disease monitoring.

In the context of research, MRI datasets offer rich opportunities for quantitative analyses. Texture features, volumetric measurements, and advanced radiomic signatures derived from MRI have shown promise in correlating with molecular biomarkers and patient prognoses.

These developments dovetail with the growing interest in personalized medicine, where an individual’s imaging phenotype might guide targeted therapy. Overall, MRI’s specificity and versatility make it a cornerstone in CRC management, particularly for rectal lesions, where precise anatomical delineation can be the difference between organ preservation and more extensive surgical procedures.

1.3.2 Segmentation Challenges

Accurate segmentation of the tumor region on MRI is crucial for:

1. Precise volumetric measurement of tumor burden.
2. Extraction of *radiomic* features that may correlate with histopathology or genetic profiles.
3. Guiding radiation therapy and surgical planning.

However, MR images are prone to intensity inhomogeneity, noise, and artifacts. Additionally, CRC tumors can vary considerably in appearance, location, and shape, which makes manual delineation tedious, time-consuming, and subject to inter-observer variability.

Automated or semi-automated segmentation methods using deep learning have gained traction to address these limitations. Convolutional Neural Networks (CNNs), in particular, have yielded state-of-the-art results in many medical imaging applications [26].

One of the key difficulties in tumor segmentation on MRI arises from the modality’s inherent sensitivity to tissue relaxation properties, which can manifest as inconsistent intensities across a single image or between different scans. This intensity inhomogeneity is especially pronounced if imaging protocols vary among institutions or if patient-specific factors—such as body composition—differ significantly. As a result, a segmentation model trained on one dataset might struggle to generalize to another unless techniques like normalization and data augmentation are carefully applied.

Another layer of complexity is introduced by the highly heterogeneous nature of CRC lesions themselves. Tumors may present with irregular borders, necrotic cores, or varying degrees of contrast enhancement, making them difficult to distinguish from surrounding tissues. In rectal cancers, for example, local structures like the mesorectal fascia or neighboring organs can appear similar to tumor tissue on certain MRI sequences. Such nuances can confuse both human observers and automated algorithms, leading to inaccuracies in boundary delineation.

Manual segmentation by expert radiologists, while often considered the gold standard, is time-consuming and can lead to variability based on individual experience or fatigue. Moreover, as volumetric imaging becomes more detailed, the sheer number of slices to review adds to the workload and the potential for human error. This is where CNN-based models show immense

promise. By learning hierarchical feature representations directly from the data, deep networks can identify subtle edges or texture patterns that may elude manual examination. However, these models also require large, well-annotated datasets to capture the full spectrum of possible tumor appearances and to generalize effectively.

In recent years, attention-based and transformer-based architectures have been introduced to complement or replace standard CNNs. These approaches allow the network to weight specific regions of the image more heavily, potentially improving segmentation in cases with complex backgrounds or variable tumor shapes. Coupled with advanced normalization techniques—like Hybrid Spatial Normalization mentioned in your work—these models aim to reduce artifacts and intensity discrepancies, leading to more uniform segmentation outcomes.

Ultimately, overcoming segmentation challenges requires a synergistic blend of high-quality imaging, robust data preprocessing, and sophisticated network architectures. As more research validates these methods in multi-center cohorts, automated segmentation is poised to become a cornerstone for both clinical decision-making and large-scale radiomic investigations, unlocking more precise and personalized approaches to CRC management.

1.4 Radiomics in Colorectal Cancer

1.4.1 Definition and Process

Radiomics is a field that extracts large numbers of quantitative features from standard-of-care images (CT, MRI, PET, etc.) through high-throughput data-mining algorithms [22]. These features, such as shape descriptors, intensity histograms, or higher-order textural patterns, may provide information on the tumor microenvironment and heterogeneity beyond what can be appreciated visually.

A typical radiomics pipeline involves:

1. Image acquisition and reconstruction.
2. Image segmentation (manually or using AI-based methods).
3. Feature extraction (first-order statistics, shape-based, texture-based, etc.).
4. Model building (machine learning or deep learning) to predict clinical endpoints.

At its core, radiomics seeks to convert medical images into high-dimensional, mineable data that can offer deeper insights into the biology of tumors. The principle hinges on the idea that visually subtle or invisible features of tumor morphology and texture may reflect underlying genomic or proteomic states. For instance, what appears as a uniformly enhanced tumor region

in a typical radiological view may, upon quantitative analysis, reveal patterns of pixel intensity variations that correlate with certain mutations or levels of aggressiveness.

An essential step in the radiomics process is image standardization. Different scanning parameters, variations in contrast administration, and diverse hardware settings can lead to inconsistencies in image intensities. Before extracting radiomic features, many pipelines require an image normalization phase to mitigate these discrepancies. This ensures that when features such as histogram metrics or textural patterns are calculated, they remain comparable across patients and institutions.

Another critical consideration is region of interest (ROI) definition. Since the accuracy of radiomics heavily depends on the precision of tumor segmentation, automated or semi-automated methods must be validated against expert annotations. Errors in ROI delineation can propagate through the entire pipeline, leading to incorrect feature values and flawed downstream analyses. In some workflows, multiple experts may annotate each scan, and consensus or majority voting is used to create a more reliable “ground truth.”

Once segmentation is complete, feature extraction software calculates diverse metrics. **First-order** features describe basic statistics of intensity values, such as mean, median, and standard deviation, without considering spatial relationships. **Shape-based** features quantify geometric aspects like sphericity, compactness, or surface area. **Texture-based** or **higher-order** features—derived from matrices such as the Gray-Level Co-occurrence Matrix (GLCM)—offer insights into spatial distribution and repetitive patterns in the image. These features are potentially linked to tumor vascularization, necrosis, and cellular density, among other histopathological attributes.

The final step involves building predictive models using these extracted features. Traditional machine learning techniques (e.g., random forests, support vector machines) have been widely employed; however, deep learning approaches can also ingest radiomic features directly or learn new, more complex features through convolutional layers. Regardless of the modeling approach, systematic feature selection and cross-validation are necessary to avoid overfitting. Model performance is typically evaluated using metrics such as accuracy, sensitivity, specificity, or area under the ROC curve (AUC), depending on the clinical question being addressed (e.g., classification vs. survival prediction).

One emerging trend is the fusion of radiomics with other *omics* data, such as genomics or proteomics, creating integrative models that can capture both phenotypic and molecular tumor characteristics. While this approach holds great promise, it also presents significant challenges in terms of data handling and model interpretability. Moreover, ensuring reproducibility across different institutions remains a pivotal issue, highlighting the necessity for standardized radiomics protocols and open-source toolkits.

In sum, the radiomics process is both data-intensive and methodologically intricate. Its ultimate goal—turning routine clinical images into objective, quantifiable biomarkers—has the potential to revolutionize personalized medicine, particularly in colorectal cancer. As technology evolves and collaborative data-sharing initiatives expand, the reliability and clinical impact of radiomics are expected to grow substantially.

1.4.2 Applications

Radiomics has been investigated in CRC to predict:

- Tumor staging and grading [22].
- Genetic mutations (*e.g.*, KRAS, NRAS, BRAF status) [30, 13].
- Microsatellite instability (MSI) [7, 11].
- Response to therapy and prognosis [32].

By discovering imaging biomarkers that correlate with pathology or *omics* data, radiomics can contribute to non-invasive, repeatable “virtual biopsies,” facilitating personalized therapy.

Radiomics in colorectal cancer (CRC) addresses a broad range of clinical challenges, aiming to optimize patient care at every stage of disease management. One of the earliest applications is in the realm of **tumor staging and grading**, where quantitative imaging features can complement conventional staging techniques. For instance, certain texture features have shown associations with tumor aggressiveness, potentially aiding clinicians in refining T-staging or even distinguishing between well-differentiated and poorly differentiated tumors. Such insights can be particularly valuable in borderline cases where standard imaging lacks sufficient clarity.

Another rapidly expanding area focuses on predicting **genetic mutations**, such as KRAS, NRAS, and BRAF [30, 13]. Mutational status often dictates the efficacy of targeted therapies (*e.g.*, anti-EGFR antibodies), and identifying these mutations through a non-invasive method could spare patients from repeated tissue biopsies. Studies employing CT or MRI-based radiomics have reported promising results in discriminating between wild-type and mutant KRAS or BRAF, revealing that specific intensity or textural metrics correlate with molecular subtypes. Although these predictive models require further validation in larger, multi-center cohorts, they illustrate the potential for radiomics to function as a surrogate for genetic testing.

Microsatellite instability (MSI) is another clinically important biomarker in CRC, associated with prognosis and response to immunotherapy. Traditionally, MSI status is determined through molecular assays on tumor tissue. However, radiomics-based approaches [7, 11] suggest that certain imaging signatures may indicate MSI-positive disease, thus opening the door

to non-invasive, image-based screening for immunotherapy eligibility. If validated at scale, this could accelerate treatment decisions and reduce the need for invasive procedures.

Perhaps the most immediately impactful application is in predicting **therapy response and prognosis** [32]. For patients undergoing neoadjuvant chemoradiotherapy or targeted treatments, early imaging biomarkers can provide feedback on therapeutic effectiveness. Radiomics metrics, monitored over time, may detect minute changes in tumor structure or vascularization before they become radiologically apparent. This could allow clinicians to adjust therapy regimens promptly or switch to alternative treatments when the current plan proves ineffective. Furthermore, models predicting long-term survival or disease recurrence based on pre- or post-therapy scans have shown potential to refine conventional risk stratification methods (e.g., TNM staging).

Outside of these primary domains, radiomics features have also been explored for **postoperative follow-up and recurrence detection**. Since CT or MRI imaging is commonly conducted during surveillance, the radiomics approach can be easily integrated into existing clinical workflows. Patients with abnormal radiomic signatures on surveillance imaging could be flagged for more frequent follow-up, additional imaging, or prompt intervention. This personalized monitoring could ultimately improve overall survival rates by catching recurrences at an earlier, more treatable stage.

Looking ahead, a notable area of growth lies in combining radiomics with clinical data, such as laboratory test results (e.g., CEA levels) and patient demographics. Machine learning models that incorporate these multi-dimensional inputs could yield more robust predictions for patient outcomes. Additionally, the integration of AI-driven segmentation in the radiomics pipeline can further streamline the process, reducing inter-observer variability and labor-intensive manual annotation.

Nevertheless, challenges remain. **Standardization** of image acquisition protocols and feature calculation methods is critical for reproducibility across institutions. Large-scale, validated databases are required to ensure that radiomics-based models apply broadly and do not overfit to a particular patient population. Despite these obstacles, the future of radiomics in CRC is promising, with ongoing research exploring refined methodologies and expanding the spectrum of clinically relevant applications. As computational tools evolve and data-sharing initiatives become more common, radiomics has the potential to become an integral part of precision oncology in colorectal cancer management.

1.5 Deep Learning for Automated MRI Segmentation

1.5.1 Convolutional Neural Networks (CNNs)

Deep learning, particularly with CNNs, has revolutionized image recognition tasks in both natural and medical images. CNNs automatically learn hierarchical representations from large datasets, capturing important spatial features for segmentation, classification, or detection tasks [18].

1.5.2 The UNet Architecture

U-Net is a popular CNN-based architecture designed for biomedical image segmentation [26]. It uses an encoder-decoder structure with skip connections to preserve spatial details while progressively abstracting and recovering semantic information. U-Net has shown high performance in segmenting organ boundaries and lesions with a limited amount of training data—a common constraint in medical imaging tasks.

1.5.3 2D-UNet, 2D-UMamba, and 2D-Swin-UMamba

- **2D-UNet:** A baseline variant of U-Net operating on 2D slices (rather than 3D volumes). It is simpler to train and requires less computational resources but may forgo volumetric context.
- **2D-UMamba:** A modified version of U-Net that integrates *Mamba* blocks (multi-scale attention modules, or other specialized attention/residual connections) to better capture local and global context in the 2D domain.
- **2D-Swin-UMamba:** Incorporates *Swin Transformers* in place of or alongside convolutional layers in some parts of the network to further enhance feature representation and capture long-range dependencies [20].

Such advanced hybrid CNN-Transformer architectures could provide more robust segmentation in the presence of highly variable tumor shapes and intensities. By combining the domain knowledge of typical CNNs with the global context modeling of Transformers, these architectures aim to achieve high segmentation accuracy on colorectal MRI scans.

Accurate and consistent tumor segmentation is a prerequisite for meaningful radiomics analysis. In this thesis, we propose and evaluate the performance of the aforementioned deep learning models for 2D slice-based MRI segmentation of CRC. Once the best-performing segmentation masks are generated, we extract a series of radiomics features (e.g., histograms,

co-occurrence matrices, and advanced textural features) to explore potential correlations with pathological markers and patient outcomes, including:

- Stage (TNM).
- MMR, KRAS/NRAS/BRAF status.
- Treatment response.

Thus, the synergy of deep learning-based segmentation and radiomics analysis can lay a foundation for developing *in vivo* biomarkers of CRC aggressiveness and prognosis.

Several studies have explored the combination of segmentation and radiomics in CRC:

- **Taguchi et al.** [30] used CT texture features to predict KRAS mutation in CRC, showing good performance but requiring manual segmentation.
- **Fan et al.** [7] demonstrated that CT radiomics could predict MSI status, although manual ROI selection was a bottleneck.
- **Zhang et al.** [34] examined MRI-based radiomics for rectal cancer MSI detection. They likewise stressed the need for accurate and standardized segmentation.

Deep learning-based methods for automated segmentation of rectal tumors on MRI or CT have shown promise [10], but differences in protocols and data availability hinder direct comparisons. The proposed 2D-UNet, 2D-UMamba, and 2D-Swin-UMamba architectures aim to address these issues by leveraging multi-scale attention or Transformer mechanisms to handle variability in MRI appearances.

1.6 Objectives

The overarching goal of this thesis is to develop and evaluate robust, automated segmentation methods for colorectal tumors on MRI, then integrate the segmentation results into a radiomics pipeline to derive predictive and/or prognostic features. Specifically, the objectives are:

1. **Segmentation Model Development:** Implement three segmentation models (2D-UNet, 2D-UMamba, and 2D-Swin-UMamba) and compare their accuracy and robustness on CRC MRI data.
2. **Radiomic Feature Extraction:** Conduct radiomics analysis on automatically segmented tumor regions, extracting first-order, textural, and higher-order features.

3. **Evaluation and Validation:** Assess the predictive potential of these extracted features for key biomarkers (e.g., MSI or KRAS mutation), staging, or response to therapy.
4. **Performance Benchmarking:** Compare results against manual segmentation baselines and discuss generalizability to different MRI protocols.

1.7 Literature Review

Artificial intelligence (AI) and radiomics are reshaping how rectal cancer (RC) is managed. By leveraging advanced imaging techniques such as magnetic resonance imaging (MRI), diagnostic accuracy, staging precision, and treatment outcomes can be enhanced. Rectal cancer constitutes approximately one-third of all colorectal cancers and poses significant treatment challenges due to its complexity and elevated risk of recurrence. Recent progress in AI—particularly in machine learning (ML) and deep learning (DL)—enables the extraction and analysis of high-dimensional imaging features, known as radiomics, which provide more comprehensive insights than traditional diagnostic methods [6].

High-resolution MRI (HR-MRI) is considered the optimal imaging modality for RC, especially for assessing T, N, and M stages, evaluating mucin content, and identifying lymphovascular invasion (LVI) and extramural vascular invasion (EMVI). Nonetheless, conventional imaging approaches often struggle to provide sufficient guidance for treatment planning. AI-based techniques can address these limitations by automating image analyses, uncovering tumor-specific patterns, and predicting treatment efficacy. For instance, ML models that incorporate MRI data have been shown to outperform traditional methods in predicting pathological complete response (pCR) and distant metastases (DM) [6].

Furthermore, combining AI and radiogenomics—linking imaging phenotypes to genetic profiles—enables identification of tumor genotypes such as KRAS, NRAS, BRAF, and microsatellite instability (MSI). These molecular markers are crucial for devising personalized treatment strategies. Despite these advances, challenges persist in standardizing AI workflows, ensuring image quality, and validating models in diverse clinical environments. Successful transition from research to practical use requires radiologists, oncologists, and data scientists to work collaboratively on reliable, robust systems [6].

Di Costanzo et al. highlight the potential of AI and radiomics in the management of rectal cancer, as well as the need for further research to mitigate current limitations and foster clinical implementation of AI-based diagnostic and prognostic tools.

The article “*Radiomics and Machine Learning Applications in Rectal Cancer: Current Update and Future Perspectives*” emphasizes the transformative role of artificial intelligence (AI), particularly radiomics and machine learning (ML), in improving the diagnosis and treatment of

rectal cancer (RC). Magnetic resonance imaging (MRI) is established as the most valuable imaging modality for local staging and restaging in RC, offering insights into tumor characteristics like circumferential resection margins and extramural venous invasion. However, conventional imaging techniques face limitations in addressing tumor heterogeneity and accurately predicting patient outcomes. AI, through radiomics, extracts quantitative imaging features that reflect tumor biology, paving the way for precision medicine. The study emphasizes that ML algorithms, trained on such features, can build predictive models to support clinical decision-making in various contexts, such as tumor staging, response evaluation, and risk stratification [27].

A significant focus of the article is the use of AI and radiomics to improve staging and prediction of therapeutic response. Radiomics-driven ML models have shown promise in differentiating tumor stages (T1–T4) and predicting pathological complete responses (pCR) to neoadjuvant chemoradiotherapy (nCRT). For example, ML classifiers that use radiomic features of T2-weighted and diffusion-weighted MRI sequences have achieved high sensitivity and specificity for these tasks, demonstrating their potential to guide treatment strategies. Furthermore, radiogenomics, an integration of radiomics with genetic profiling, has been explored to predict genetic mutations such as KRAS, which have critical prognostic and therapeutic implications. Such innovations address the unmet need for noninvasive tools that complement conventional histopathological evaluation [27].

Despite these advancements, the article underscores the challenges in translating these AI-based approaches into clinical practice. Issues such as standardizing radiomics pipelines, managing data variability, and ensuring model generalizability across institutions must be resolved. The study advocates for robust multicenter trials and the adoption of shared guidelines for AI research design to overcome these barriers. In conclusion, the integration of AI into medical imaging has the potential to revolutionize RC management by enabling personalized care. This aligns with the thesis objective of using artificial intelligence for rectal cancer, as it provides a solid foundation of current applications and a path for future improvements [27].

The article, “*Predicting Tumor Deposits in Rectal Cancer: A Combined Deep Learning Model Using T2-MR Imaging and Clinical Features*,” looks at how artificial intelligence (AI) can help doctors identify tumor deposits (TDs) in rectal cancer (RC) before surgery. Tumor deposits are important because they are signs of a worse outlook for patients and affect treatment choices [16].

The study created a hybrid deep learning (DL) model that combines T2-weighted MRI images and clinical risk factors to improve diagnosis. The researchers used data from 327 RC patients and compared four models: clinical, single-channel DL, multi-channel DL, and hybrid DL. The hybrid DL model worked best, with an area under the curve (AUC) score of 0.857 in the development dataset and 0.839 in a separate testing dataset [16].

A major contribution of this study is the inclusion of peri-tumoral regions in the DL models. The multi-channel DL model, which used images from both the tumor and surrounding areas, performed better than the single-channel model. This improvement comes from the importance of the fat around the tumor in understanding the tumor's behavior, which is often related to how aggressive and varied it is. The hybrid DL model improved its predictions further by adding clinical factors like extramural vascular invasion (EMVI), circumferential resection margin (CRM), and serum CA19-9 levels. These results show how combining imaging and clinical data can help doctors evaluate RC patients non-invasively and accurately before surgery [16].

The study also discusses the challenges and limitations of using DL in clinical settings. Problems such as making sure the model works well in different situations, variations in data, and needing manual segmentation of regions are obstacles to wider use. The authors suggest conducting multicenter studies to confirm their results and strengthen the model. Overall, this research provides a helpful framework for using AI in RC management and fits well with the goal of using AI in rectal cancer diagnosis through medical imaging [16].

The article [21] investigates the use of a deep learning-based approach, the DeepPCR model, for predicting pathological complete response (pCR) in locally advanced rectal cancer (LARC) patients. The study used hematoxylin and eosin (H&E)-stained biopsy slides from 842 LARC patients, which were retrospectively collected across three medical centers. The DeepPCR model achieved an AUC-ROC of 0.710 in the testing cohort and 0.723 in the external validation cohort, significantly outperforming other methods. Its integration of multi-instance learning (MIL) allowed for robust feature extraction from whole-slide images, showcasing its utility as an independent predictive factor for treatment response. This aligns with the thesis objective of using AI for RC diagnostics by demonstrating how advanced image analysis can aid in non-invasive prediction and clinical decision-making [21].

The article, [33], presents a review of the application of deep learning (DL) techniques in magnetic resonance imaging (MRI) for lesion segmentation in rectal cancer (RC). Highlighting the challenges of traditional diagnostic methods, such as reliance on radiologist expertise and inefficiencies in manual segmentation, the article explores how DL-based segmentation algorithms like U-Net and its variants (e.g., 3D U-Net and U-Net++) have revolutionized medical image analysis. By leveraging convolutional neural networks (CNNs) and innovative architectures, these methods have achieved remarkable performance in identifying tumors and surrounding tissues, enhancing accuracy in preoperative staging and radiotherapy planning. This review emphasizes the potential of DL-based segmentation to advance the efficiency and precision of RC diagnosis, directly supporting the integration of AI in RC imaging for clinical use [33].

References [33] and [21] collectively highlight the transformative potential of AI in rectal cancer management, from predictive modeling to advanced image segmentation, forming a

robust foundation for the thesis.

Rectal cancer is one of the most common and lethal malignancies, affecting both men and women. Accurate staging and therapeutic planning are crucial for optimizing treatment outcomes. Standard imaging approaches, including magnetic resonance imaging (MRI) and computed tomography (CT), provide vital information on tumor extent, lymph node involvement, and metastases. However, conventional qualitative assessment alone may fail to capture the full complexity of the disease.

This study, [15], evaluates the diagnostic accuracy of artificial intelligence (AI) models based on magnetic resonance imaging (MRI) in predicting pathological complete response (pCR) to neoadjuvant chemoradiotherapy (nCRT) in rectal cancer. A systematic review and meta-analysis were conducted by searching PubMed, Embase, the Cochrane Library, and Web of Science for relevant studies published before June 21, 2022. A total of 21 studies, including 1,873 patients in the validation cohorts, met the inclusion criteria. The methodological quality of these studies was assessed using the QUADAS-2 and Radiomics Quality Score (RQS) tools. A bivariate random-effects model was used to compute pooled estimates of sensitivity, specificity, and area under the curve (AUC).

The meta-analysis revealed that AI-based MRI models demonstrated strong predictive capability, with a pooled AUC of 0.91 (95% CI: 0.88–0.93), sensitivity of 0.82 (95% CI: 0.71–0.90), and specificity of 0.86 (95% CI: 0.80–0.91). Subgroup analyses indicated that deep learning (DL) models outperformed traditional radiomics models, achieving a higher AUC of 0.97 compared to 0.85. Moreover, combined models incorporating clinical factors had an improved diagnostic accuracy, with an AUC of 0.92, in contrast to radiomics-only models, which had an AUC of 0.87. The mean RQS score of included studies was 10.95, reflecting a need for higher methodological quality in future research.

In conclusion, AI-based MRI models, particularly deep learning approaches and models that integrate clinical data, offer a promising noninvasive tool for predicting pCR in rectal cancer patients undergoing nCRT. These findings highlight the potential role of AI in guiding personalized treatment strategies, reducing unnecessary surgeries, and improving patient outcomes. However, further prospective, large-scale, multicenter studies are needed to validate these models and enhance their clinical applicability.

Endoscopic imaging plays a crucial role in diagnosing and monitoring rectal cancer throughout its treatment process, [12], including initial screening, treatment response assessment, and follow-up for potential regrowth. However, visual assessment by clinicians is highly variable, leading to misdiagnosis that may result in unnecessary surgeries or missed detections of cancer recurrence. This study explores the effectiveness of deep learning-based classifiers, particularly the Swin Transformer, in providing robust and consistent detection of rectal cancer from

endoscopic images. The research compares Swin Transformer with convolutional neural networks such as ResNet-50 and WideResNet-50, as well as Vision Transformer (ViT), across in-distribution (ID) and out-of-distribution (OOD) datasets. Results indicate that Swin Transformer demonstrates superior accuracy (ID: 0.84, OOD: 0.83) and robustness, particularly under color-shift perturbations, making it a promising tool for longitudinal rectal cancer assessment.

The study also investigates the impact of distribution and concept drift on automated cancer assessment models. To simulate distribution shifts, color perturbations were applied using optimal transport methods, while concept drift was evaluated through the analysis of tumor regrowth and unrelated conditions like colitis and adenomas. Despite variations in endoscopic imaging conditions, Swin Transformer outperformed other models, maintaining high sensitivity and specificity even when tested on datasets containing significant confounders. Additionally, the study highlights the importance of deep learning methods that can generalize well across diverse datasets without the need for manual segmentation. The research underscores the potential of hierarchical vision transformers in medical imaging, suggesting that Swin Transformer can improve the accuracy of rectal cancer diagnosis and reduce inter-observer variability.

The results of this study pave the way for future research in applying transformer-based architectures to medical imaging tasks beyond rectal cancer assessment. The authors propose further validation with multi-institutional datasets and an in-depth evaluation of confounding factors such as blood, stool, and treatment-induced changes. These findings emphasize the need for AI-driven solutions that enhance clinical decision-making, offering more reliable tools for cancer diagnosis and follow-up monitoring. By leveraging the strengths of deep learning, particularly the Swin Transformer, this study contributes to advancing the field of gastrointestinal endoscopic imaging and improving patient outcomes.

Volumetric medical images, [19], such as Magnetic Resonance Imaging (MRI), play a crucial role in rectal cancer staging by distinguishing between T2 and T3 stages, which is essential for determining appropriate treatment strategies. The study proposes a volumetric convolutional neural network (CNN) that classifies rectal MRI volumes using a ResNet-based volume encoder with 3D convolution at the last layer, bilinear feature aggregation, and joint optimization of triplet and focal loss. Compared to traditional 2D CNN models that require radiologists to manually select representative slices, the proposed 3D CNN model effectively captures inter-slice relationships, improving classification accuracy. The experimental results demonstrated that the model achieved an AUC of 0.831, surpassing professional radiologists' reported accuracy, thus highlighting the potential of deep learning in automated rectal cancer staging.

The study also explored different network architectures and aggregation functions for volumetric image analysis. It compared pure 2D CNN, 3D CNN, and hybrid models with varying levels of inter-slice fusion, concluding that late fusion (3D convolution at the final layer) per-

formed best. Additionally, the research examined different depth aggregation functions, such as max pooling, attention weighting, and bilinear encoding, finding that bilinear encoding improved performance by capturing fine-grained details relevant to tumor staging. By integrating these advancements, the study presents a robust deep learning approach for preoperative rectal cancer staging, which could be extended to other volumetric medical imaging tasks.

Chapter 2

Mathematical Foundations

2.1 Introduction

This chapter presents the core mathematical and methodological background underlying our MRI segmentation framework. We begin by reviewing the fundamental operations of Convolutional Neural Networks (CNNs)—including convolutions, pooling, and batch normalization—which are indispensable for feature extraction in medical imaging. Next, we describe the encoder–decoder architectures that form the backbone of our segmentation pipeline: U-Net, UMamba, and SwinUMamba.

In addition, we introduce two advanced modules that significantly enhance segmentation performance:

PVTv2-Based Encoder: A state-of-the-art backbone that leverages the Pyramid Vision Transformer v2 (PVTv2) to capture multi-scale, long-range contextual information. This module uses a hierarchical transformer architecture with patch embedding and spatial reduction techniques to efficiently process high-resolution images.

Hybrid Segmentation Network (HSN): An architecture that integrates a pre-trained CNN encoder (e.g., ResNet18) with a decoder that employs skip connections. This design fuses high-level semantic features with low-level spatial details, ensuring that fine structural boundaries are accurately preserved in the segmentation process.

Finally, we detail our loss function—binary cross-entropy with logits—and describe the evaluation metrics (Dice coefficient and Intersection-over-Union) along with our data preprocessing, augmentation, and optimization strategies based on the Adam algorithm.

2.1.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) constitute the foundation of contemporary image analysis methodologies and have demonstrated exceptional efficacy in medical imaging applications, including but not limited to lesion detection and organ segmentation. The remarkable success of CNNs is largely attributable to their architectural design, which combines convolutional layers, responsible for learning localized features, and pooling layers, which progressively reduce the spatial resolution of the feature maps. This dual mechanism enables CNNs to efficiently capture and abstract intricate patterns within high-dimensional image data.

At the core of CNNs is the convolutional layer, where the principal operation is the two-dimensional (2D) convolution. This operation systematically extracts local features from an input image or feature map, serving as the critical process by which raw pixel data is transformed into meaningful representations. The mathematical formulation of the 2D convolution is expressed as:

$$(F * X)[i, j] = \sum_{a=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} \sum_{b=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} F[a, b] \cdot X[i - a, j - b], \quad (2.1)$$

where F denotes a learnable kernel of dimensions $k \times k$, and X represents the input feature map. In this equation, the kernel F is applied at each spatial location (i, j) of the input X . The summation indices a and b range symmetrically from $-\lfloor k/2 \rfloor$ to $\lfloor k/2 \rfloor$, ensuring that the kernel is centered on the target pixel. The operation computes a weighted sum of the local neighborhood in X , thereby capturing essential spatial information pertinent to the image structure.

Following the convolution operation, a bias term is incorporated into the resulting feature map. This bias term, which is also learnable, allows the network to adjust the activation levels and provides an additional degree of freedom during the training process. Subsequent to bias addition, a non-linear activation function—most commonly the Rectified Linear Unit (ReLU)—is applied to introduce non-linearity into the model. The inclusion of such an activation function is pivotal, as it endows the network with the ability to learn complex, non-linear mappings between the input and output, a characteristic that is indispensable for effective image analysis.

The architectural design of CNNs leverages the stacking of multiple convolutional layers to capture hierarchical patterns within the data. In the initial layers, the convolutional filters predominantly learn to detect simple, low-level features such as edges, lines, and basic textures. As the depth of the network increases, the successive layers integrate these elementary features to form more abstract and high-level representations that may correspond to specific shapes, object parts, or even entire anatomical structures. This hierarchical feature extraction is a principal factor underlying the robust performance of CNNs in diverse applications, particularly in

scenarios where subtle variations in texture and structure are diagnostically significant.

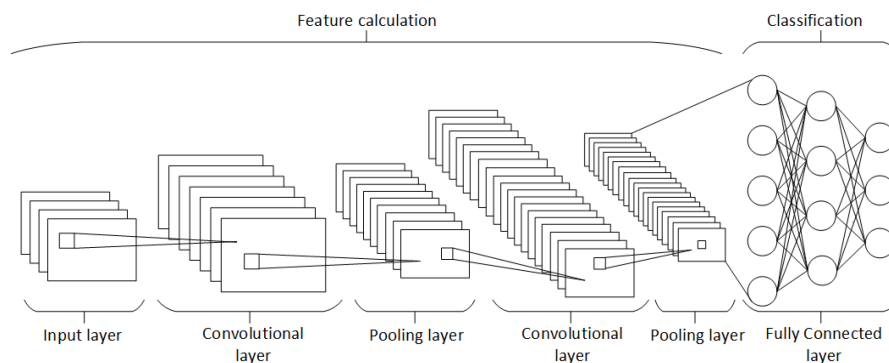


Figure 2.1: Block diagram of 2D CNN architecture

2.1.2 Pooling Layers

Pooling layers are essential components in CNNs that reduce the spatial dimensions of feature maps, thereby decreasing computational complexity and mitigating overfitting. A widely used pooling method is max-pooling, mathematically expressed as:

$$z_{(j)} = \max_{u \in \Omega(j)} \{a_u\} \quad (2.1)$$

In this expression, $\Omega(j)$ denotes the pooling window around the spatial index j , and a_u represents the activation at position u within that window. By selecting the maximum activation in each region, max-pooling preserves the most prominent features while discarding less relevant information. This operation not only reduces the number of parameters and computations in subsequent layers but also imparts translation invariance, ensuring that small shifts in the input do not significantly alter the output.

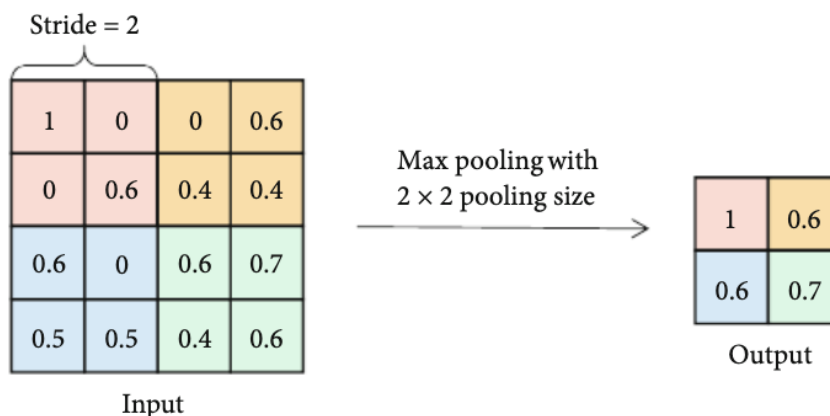


Figure 2.2: Max Pooling process

2.1.3 Batch Normalization

Deep neural networks often face problems like unstable gradients and internal covariate shift, which can slow down or even disrupt training. Batch Normalization (BN), introduced by Ioffe and Szegedy in 2015, helps solve these issues by normalizing the feature maps in each mini-batch during training.

For a mini-batch with m activations, denoted as z_i , BN first calculates the mean μ_b and variance σ_b^2 of these values. The calculations are as follows:

$$\mu_b = \frac{1}{m} \sum_i z_i,$$
$$\sigma_b^2 = \frac{1}{m} \sum_i (z_i - \mu_b)^2.$$

Next, each activation z_i is normalized by subtracting the batch mean and dividing by the square root of the variance plus a small constant ϵ (to prevent division by zero). After normalization, BN applies a learnable affine transformation to the result, giving the final output:

$$\text{BN}(z_i) = \gamma \times \left(\frac{z_i - \mu_b}{\sqrt{\sigma_b^2 + \epsilon}} \right) + \beta.$$

In this formula, γ and β are parameters learned during training, allowing the network to adjust the scaled and shifted values as needed.

Batch Normalization provides several benefits. It stabilizes the training process by reducing internal covariate shift, which allows the use of higher learning rates. Additionally, BN offers mild regularization, which can be helpful when working with medical imaging datasets that often have fewer images. Overall, Batch Normalization is essential for improving both the stability of training and the general performance of deep neural network models.

2.2 Segmentation Architectures

Our segmentation approach relies on *encoder-decoder* CNNs. The encoder progressively down-samples and learns deeper representations, while the decoder upsamples back to the original spatial resolution. Skip connections between corresponding encoder and decoder levels help preserve fine details.

2.2.1 U-Net

U-Net [26] is a fully convolutional neural network designed for biomedical image segmentation, distinguished by its symmetric encoder–decoder structure interconnected by long-range skip connections. This architecture models the segmentation process as a nonlinear mapping, where given an input image $I \in \mathbb{R}^{H \times W}$, the network learns a function

$$S = f_\theta(I),$$

with $S \in \mathbb{R}^{H \times W}$ representing the segmentation mask and f_θ the parameterized transformation defined by a series of convolutional operations.

The encoder, or contracting path, comprises L hierarchical levels that progressively reduce the spatial resolution while increasing the depth of feature representations. At each level l , the feature map X_l is computed by a double-convolution block:

$$X_l = \sigma\left(\text{BN}(\text{Conv}(X_{l-1}, W_l))\right),$$

where $\text{Conv}(\cdot, W_l)$ denotes the convolution with kernel weights W_l , $\text{BN}(\cdot)$ is batch normalization, and $\sigma(\cdot)$ is the ReLU activation function. Downsampling is achieved using a 2×2 max-pooling operation:

$$X_{l+1} = \text{MaxPool}(X_l).$$

At the network’s bottleneck—the deepest part of the encoder—a high-dimensional latent representation is obtained:

$$X_{\text{bottleneck}} = \sigma\left(\text{BN}(\text{Conv}(X_{\text{encoded}}, W_b))\right),$$

with X_{encoded} being the output of the final encoder layer and W_b the bottleneck weights.

The decoder, or expansive path, restores spatial resolution through transposed convolutions, effectively reversing the downsampling process. At each level, the upsampling is formulated as:

$$X_{l-1} = \text{UpConv}(X_l),$$

which doubles the spatial dimensions. Skip connections are then employed to concatenate the upsampled features with the corresponding encoder outputs:

$$X_{\text{merged}} = \text{Concat}(X_{\text{encoder}}, X_{\text{decoder}}),$$

ensuring that fine-grained spatial details are retained during reconstruction. The final segmentation mask is produced by applying a 1×1 convolution to X_{merged} followed by a softmax

activation:

$$S = \text{Softmax}\left(\text{Conv}_{1 \times 1}(X_{\text{merged}})\right).$$

The computational complexity of U-Net is approximately

$$\mathcal{O}(L \cdot k^2 \cdot C_{in} \cdot C_{out} \cdot H \cdot W),$$

where L is the number of levels, k is the kernel size, and C_{in} and C_{out} denote the number of input and output channels, respectively. Training is typically optimized using a hybrid loss function that combines Dice loss and binary cross-entropy (BCE) loss:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{Dice} + \lambda_2 \mathcal{L}_{BCE},$$

with λ_1 and λ_2 as weighting coefficients.

U-Net’s simple yet effective design, which combines an encoder and a decoder with skip connections, enables efficient feature extraction at multiple levels and precise localization of image details. These attributes make U-Net particularly well-suited to address the challenges of biomedical image segmentation.

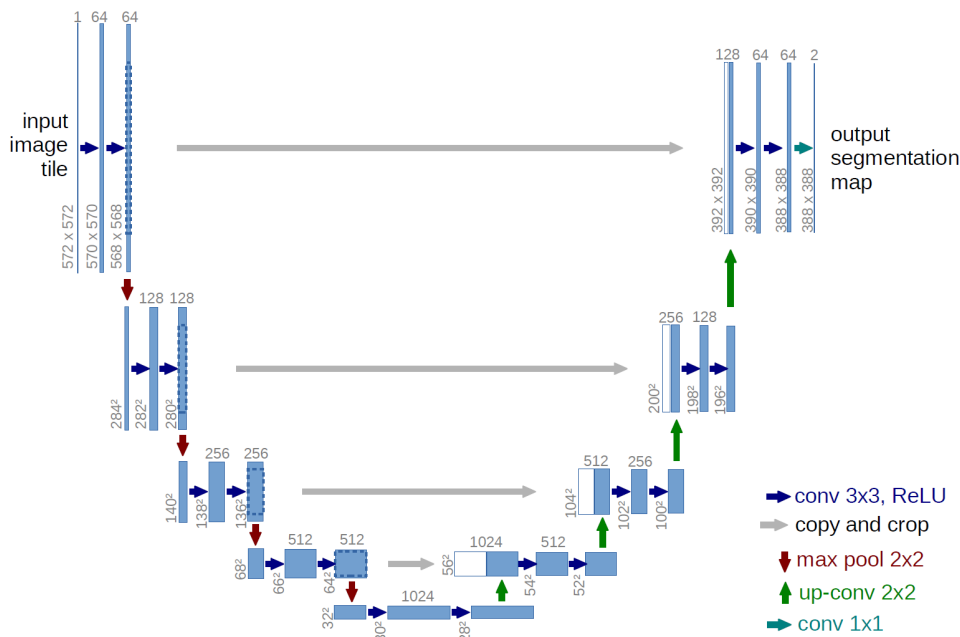


Figure 2.3: The U-net architecture

illustration for a resolution of 32x32 pixels is depicted. Each blue box indicates a multi-channel feature map, with the number of channels shown above the box. The dimensions of the x-y-size are indicated at the lower left corner of the box. White boxes signify feature maps that have been copied. The arrows represent the various operations involved.

2.2.2 UMamba

UMamba is an advanced medical image segmentation model that extends the traditional encoder–decoder framework by integrating structured state-space models (SSM) with convolutional mechanisms to enhance spatial feature extraction and representation learning. The architecture consists of a downsampling encoder, a bottleneck layer, and an upsampling decoder, with strategically placed residual connections ensuring information retention and gradient stability. Given an input MRI image

$$I \in \mathbb{R}^{H \times W \times C},$$

UMamba transforms it into a segmentation map

$$S \in \mathbb{R}^{H \times W}$$

via the nonlinear mapping

$$S = f_{\theta}(I),$$

where f_{θ} denotes the fully convolutional network parameterized by the trainable weights θ .

The feature encoding (downsampling path) involves hierarchical feature extraction using convolutional layers, batch normalization, and nonlinear activation functions. At each encoder layer l , the feature map X_l is computed as

$$X_l = \sigma\left(\text{BN}(\text{Conv}(X_{l-1}, W_l, k_l))\right),$$

where $\text{Conv}(\cdot, W_l, k_l)$ represents the convolution with learnable filter weights W_l and kernel size k_l , $\text{BN}(\cdot)$ applies batch normalization, and $\sigma(\cdot)$ is the nonlinear activation function (e.g., Leaky ReLU). Spatial downsampling is achieved through strided convolutions:

$$X_{l+1} = \text{Conv}_s(X_l, W_s, k_s),$$

where $\text{Conv}_s(\cdot)$ denotes a convolution with a stride greater than one.

At the bottleneck, UMamba incorporates structured state-space modeling to capture long-range dependencies while preserving local features. The latent representation is obtained by first flattening the output X_{enc} from the final encoder layer, applying layer normalization, and then processing the resulting sequence with SSM:

$$Z = \sigma\left(\text{SSM}(\text{LN}(\text{Flatten}(X_{\text{enc}})))\right),$$

where $\text{LN}(\cdot)$ denotes layer normalization and $\text{SSM}(\cdot)$ applies state-space modeling with learn-

able parameters. This latent representation is further refined via a one-dimensional convolution:

$$Z' = \text{Conv1D}(Z, W_{1D}, k_{1D}),$$

with W_{1D} and k_{1D} as the learnable parameters and kernel size for the 1D convolution, respectively.

In the decoding (upsampling) path, spatial resolution is progressively restored using transposed convolutions. At each decoder layer, the upsampling is formulated as

$$X_{l-1} = \sigma\left(\text{BN}(\text{UpConv}(X_l, W_l, k_l))\right),$$

where $\text{UpConv}(\cdot)$ denotes the transposed convolution operation. To effectively fuse the multi-scale features, UMamba utilizes adaptive skip connections that merge encoder and decoder features using a weighted sum:

$$X_{\text{merged}} = \alpha \cdot X_{\text{enc}} + \beta \cdot X_{\text{dec}},$$

with α and β being trainable fusion coefficients.

The final segmentation output is produced by applying a 1×1 convolution to the merged feature map, followed by a softmax activation:

$$S = \text{Softmax}\left(\text{Conv}_{1 \times 1}(X_{\text{merged}})\right).$$

UMamba is trained using a hybrid loss function that combines Dice loss and binary cross-entropy (BCE) loss:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{Dice}} + \lambda_2 \mathcal{L}_{\text{BCE}},$$

where λ_1 and λ_2 are weighting coefficients that balance the contributions of the two loss components.

Through the integration of hierarchical feature encoding, structured state-space modeling, adaptive skip connections, and transposed convolutions, UMamba achieves enhanced segmentation precision in complex medical imaging scenarios.

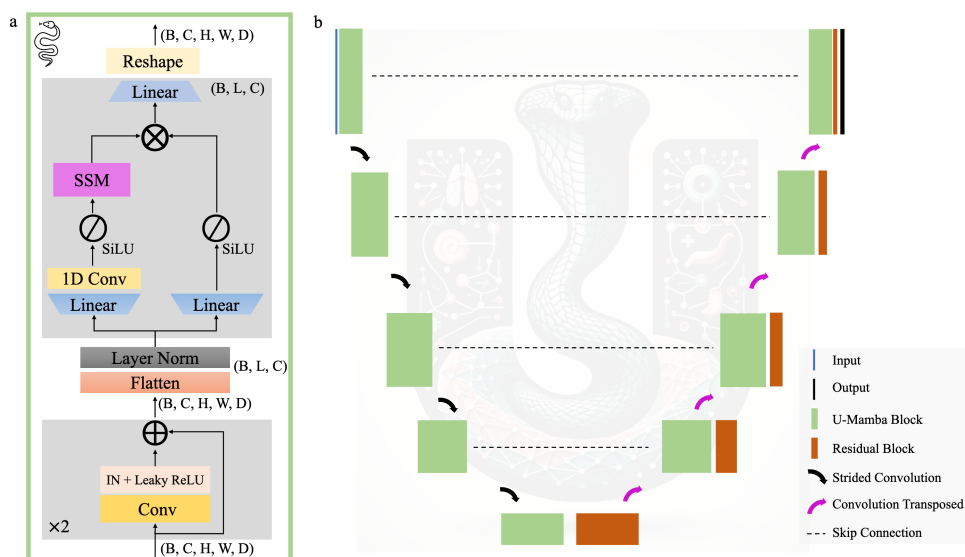


Figure 2.4: U-Mamba architecture.

a, The U-Mamba building block consists of two consecutive Residual blocks followed by the SSM-based Mamba block to improve the modeling of long-range dependencies. b, The U-Mamba utilizes the encoder-decoder framework featuring U-Mamba blocks in the encoder, Residual blocks in the decoder, along with skip connections.

2.2.3 Swin-UMamba

Swin-UMamba is an advanced medical image segmentation architecture that integrates hierarchical feature extraction with structured self-attention mechanisms, enhancing both local and global feature representations. The model follows an encoder-decoder structure but replaces conventional convolutional layers with Vision State-Space (VSS) blocks, inspired by the Swin Transformer paradigm, allowing for improved spatial-contextual awareness while maintaining computational efficiency.

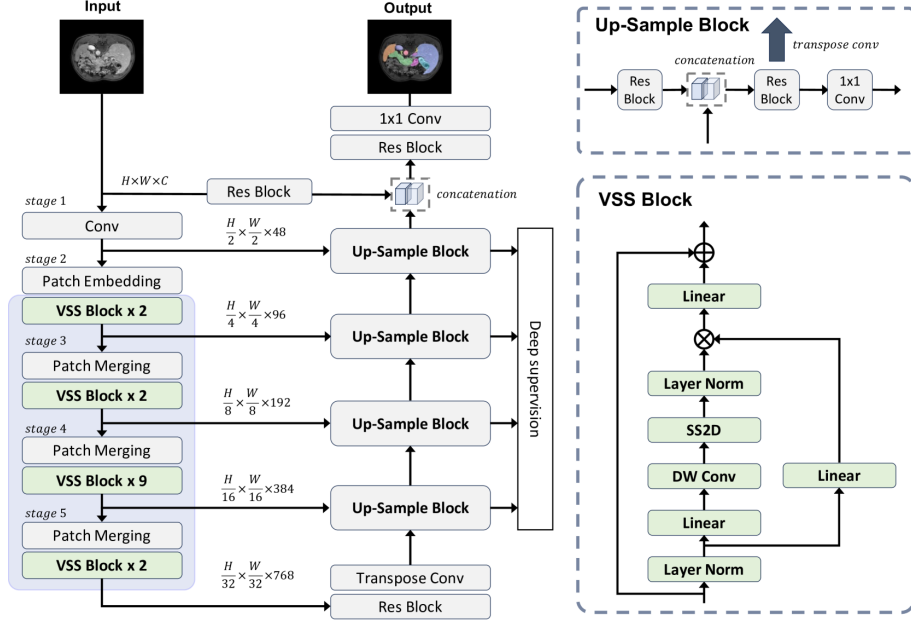


Figure 2.5: The overarching framework of Swin-UMamba.

Swin-UMamba is designed to utilize the effectiveness of vision foundation models by integrating weights from pretrained models. Each block encompassed within the blue box was initialized with weights derived from the ImageNet pretrained dataset.

The encoding process begins with an initial convolutional transformation that extracts low-level features from the input MRI image $I \in \mathbb{R}^{H \times W \times C}$. This is followed by hierarchical feature extraction through successive VSS blocks, where each stage involves a patch merging operation, reducing spatial dimensions while increasing feature representation depth. Mathematically, the initial convolutional layer is defined as:

$$X^1 = \sigma(\text{BN}(\text{Conv}(I, W^1, k^1))) \quad (2.2)$$

where W^1 and k^1 denote the convolutional filter weights and kernel size, respectively, while σ represents a non-linear activation function such as ReLU.

Each encoding stage consists of a patch embedding step followed by multiple VSS blocks. The transformation within a single VSS block can be expressed as:

$$X^l = \text{LN}(\text{SS2D}(\text{DWConv}(\text{LN}(\text{Linear}(X^{l-1})), W_{\text{lin}})))) \quad (2.3)$$

where LN represents layer normalization, Linear refers to a linear transformation, DWConv represents depth-wise convolution, and SS2D denotes a structured state-space transformation applied to extracted feature embeddings.

Patch merging is applied at each downsampling step, progressively reducing spatial dimen-

sions while increasing the number of feature channels:

$$X^{l+1} = \text{PatchMerge}(X^l) \quad (2.4)$$

where PatchMerge performs non-overlapping patch aggregation to improve representation learning. The deepest encoding layer transforms the spatially reduced features into a latent representation, ensuring global receptive field aggregation.

The decoder reconstructs the segmentation output through a series of upsampling blocks, where each upsampling step involves a transpose convolution:

$$X^{l-1} = \sigma(\text{BN}(\text{UpConv}(X^l, W^l, k^l))) \quad (2.5)$$

Skip connections merge features from the encoder with those in the decoder to retain spatial resolution and fine-grained details:

$$X_{\text{merged}} = \alpha \cdot X_{\text{enc}} + \beta \cdot X_{\text{dec}} \quad (2.6)$$

where α and β are learnable scaling parameters ensuring adaptive fusion.

The final segmentation mask is obtained via a 1×1 convolution followed by a softmax activation:

$$S = \text{Softmax}(\text{Conv}_{1 \times 1}(X_{\text{dec}})) \quad (2.7)$$

SwinUMamba is optimized using a composite loss function, combining Dice loss for overlap maximization and binary cross-entropy for pixel-wise classification:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{Dice}} + \lambda_2 \mathcal{L}_{\text{BCE}} \quad (2.8)$$

where λ_1 and λ_2 are weighting factors balancing the contributions of each loss term.

Through the integration of hierarchical VSS blocks, structured state-space models, and skip connections, SwinUMamba effectively balances local feature refinement with global contextual understanding, making it particularly suitable for segmentation tasks requiring precise boundary delineation in medical imaging.

2.2.4 PVTv2-Based Encoder Architecture

To further improve feature extraction, we integrate a Pyramid Vision Transformer v2 (PVTv2) as a backbone encoder. This transformer-based approach is particularly effective in scenarios where global context and multi-scale information are critical. The key components of the PVTv2-based encoder include:

Patch Embedding in the input image $I \in \mathbb{R}^{H \times W \times C}$ is divided into non-overlapping patches. Each patch is flattened and projected into an embedding space:

$$X_0 = \text{Flatten}(\text{Patch}(I)) \in \mathbb{R}^{N \times d_0}, \quad (2.9)$$

where N is the number of patches and d_0 is the embedding dimension.

In each transformer block, self-attention is computed over the patch embeddings. For an input matrix $X \in \mathbb{R}^{N \times d}$, the queries Q , keys K , and values V are generated as:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V, \quad (2.10)$$

where W_Q , W_K , and W_V are learnable weight matrices. The self-attention operation is given by:

$$\text{Attention}(X) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)V. \quad (2.11)$$

Spatial reduction mechanism efficiently handle high-resolution images, PVTv2 applies a spatial reduction strategy by downsampling the keys and values before computing the attention. This mechanism reduces the computational cost while still capturing essential long-range dependencies.

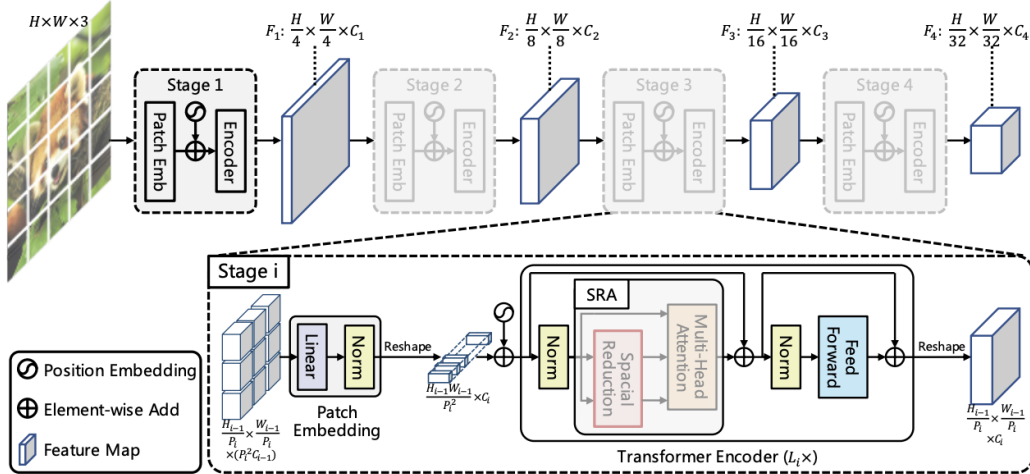


Figure 2.6: Overall architecture of Pyramid Vision Transformer (PVT).

Hierarchical feature extraction by stacking multiple transformer blocks with progressively reduced spatial dimensions, PVTv2 creates a pyramid of feature maps $\{F_1, F_2, \dots, F_L\}$ that encapsulate multi-scale information. These feature maps are subsequently passed to the decoder for segmentation mask reconstruction.

This architecture provides a global understanding of the image context, which is especially beneficial in medical imaging applications where anatomical structures can vary significantly in size and shape.

2.2.5 Hybrid Segmentation Network (HSN)

The Hybrid Segmentation Network (HSN) leverages the strengths of conventional CNN-based feature extraction along with a robust decoder that preserves spatial details via skip connections. The key components of HSN include:

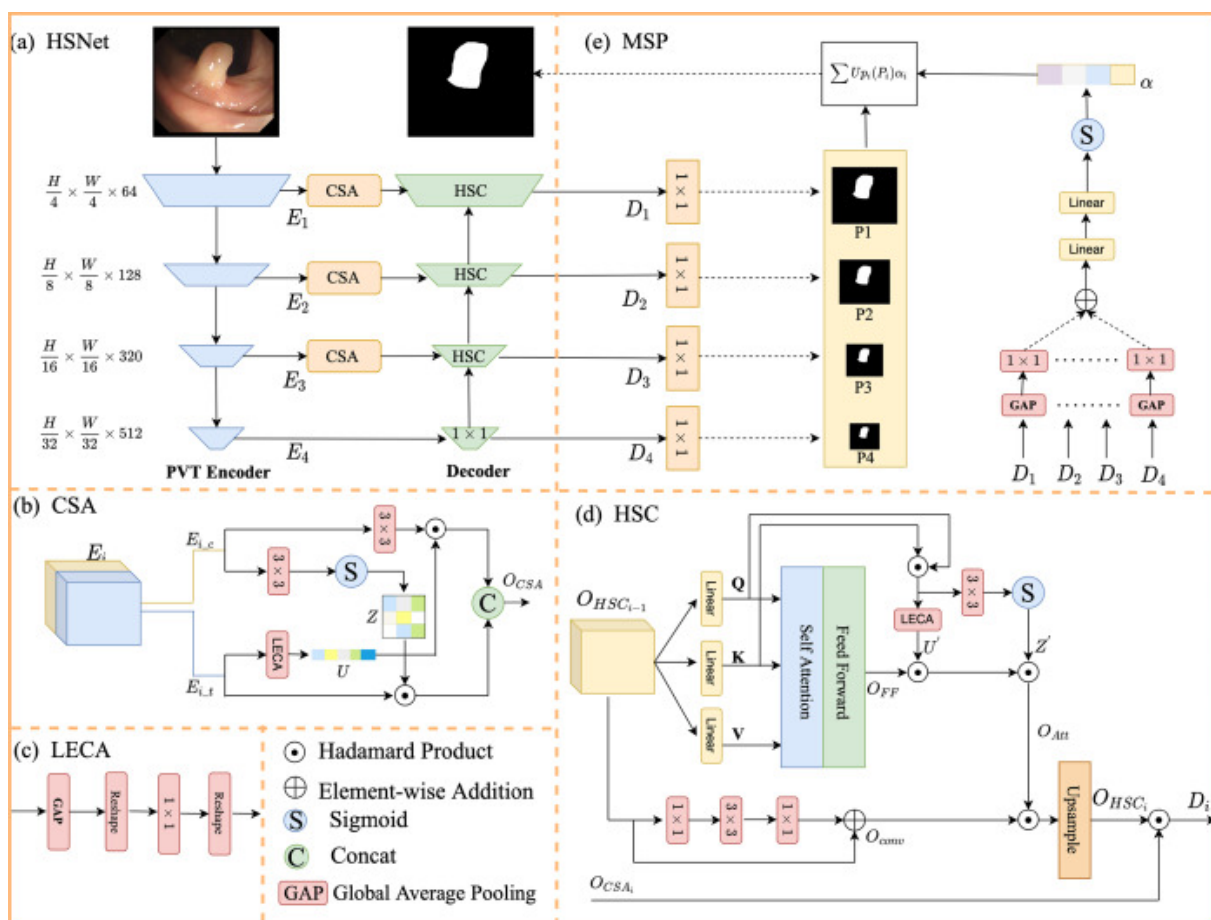


Figure 2.7: Architecture of HSNet.

integrating PVT Encoder, CSA, HSC, and MSP for medical image segmentation. The model utilizes LECA for feature enhancement, self-attention mechanisms for spatial refinement, and multi-scale processing (MSP) for improved segmentation accuracy, leveraging hierarchical deep learning techniques

- **Pre-trained CNN Encoder:** A pre-trained network (e.g., ResNet18) is employed to extract hierarchical features from the input image. These features capture complex patterns

inherent in medical images.

- **Skip Connections:** To ensure that fine-grained spatial information is retained, skip connections merge feature maps from the encoder with corresponding layers in the decoder. This fusion is mathematically expressed as:

$$F_{\text{dec}}^{(l)} = \text{Upsample}\left(F_{\text{dec}}^{(l+1)}\right) + F_{\text{enc}}^{(l)}, \quad (2.12)$$

where $F_{\text{enc}}^{(l)}$ denotes the feature map from the encoder at level l , $F_{\text{dec}}^{(l+1)}$ is the feature map from the subsequent decoder layer, and $\text{Upsample}(\cdot)$ is typically implemented via bilinear interpolation or transposed convolutions.

- **Final Output Generation:** The final segmentation mask is obtained by applying a 1×1 convolution to the fused decoder feature map, followed by a sigmoid activation function to constrain the output within the range $[0,1]$:

$$\hat{Y} = \sigma\left(\text{Conv}_{1 \times 1}\left(F_{\text{dec}}^{(0)}\right)\right). \quad (2.13)$$

This design effectively combines high-level semantic features from deeper layers with low-level spatial details from earlier layers, ensuring that fine anatomical boundaries are accurately delineated in medical image segmentation.

2.3 Loss Function: BCE With Logits

We model segmentation as a pixel-wise binary classification problem, where each pixel is classified as either part of the lesion (foreground) or the background. Let x_n be the logit output for pixel n , and $y_n \in \{0,1\}$ be the corresponding ground truth label. The Binary Cross-Entropy (BCE) loss is formulated as:

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{n=1}^N \left[y_n \log(\sigma(x_n)) + (1 - y_n) \log(1 - \sigma(x_n)) \right], \quad (2.14)$$

where $\sigma(\cdot)$ is the sigmoid activation function, and N is the total number of pixels in the image.

The BCE loss is particularly suitable for segmentation tasks due to its ability to penalize incorrect predictions at the pixel level. However, in practice, direct application of BCE on raw logits can lead to numerical instability, especially when logits take extreme values. To address this, we use the PyTorch implementation `BCEWithLogitsLoss`, which internally combines the

sigmoid activation with the cross-entropy calculation. This integration not only improves numerical stability but also simplifies our implementation by reducing the number of operations.

Moreover, in segmentation problems, class imbalance is a common challenge; for instance, the lesion region may occupy a much smaller area compared to the background. While BCE loss treats every pixel equally, its formulation can be adapted by weighting the positive and negative classes differently if needed. In our experiments, careful normalization of input intensities and balanced data augmentation help mitigate severe imbalances, ensuring that the BCE loss remains effective.

It is also common to combine BCE loss with overlap-based losses such as Dice loss, which directly measures the agreement between the predicted and ground truth masks. In our framework, we use a hybrid loss function that integrates both BCE and Dice losses:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{BCE}} + \lambda_2 \mathcal{L}_{\text{Dice}}, \quad (2.15)$$

where the Dice loss is defined as

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \sum_i p_i g_i}{\sum_i p_i + \sum_i g_i + \epsilon}, \quad (2.16)$$

with p_i and g_i representing the predicted and ground truth pixel values respectively, and ϵ a small constant to avoid division by zero. This combined loss function leverages the pixel-wise accuracy of BCE and the spatial overlap focus of Dice loss, resulting in improved segmentation performance particularly for small or irregular lesions.

the use of `BCEWithLogitsLoss`—along with the potential for a hybrid loss incorporating Dice loss—provides a robust and stable training objective. It ensures that our network learns accurate pixel-level classifications while maintaining numerical stability, making it highly suitable for the challenging task of colorectal tumor segmentation.

2.4 Evaluation Metrics: Dice and IoU

While pixel-level *accuracy* can measure the proportion of correctly classified pixels, it can be misleading in class-imbalanced cases (e.g., when the lesion occupies a very small portion of the image relative to the background). To overcome this issue, we adopt two overlap-based metrics that are more sensitive to the quality of the segmentation boundary: the Dice Similarity Coefficient (DSC) and the Intersection-over-Union (IoU).

2.4.1 Dice Coefficient

The Dice coefficient evaluates the overlap between predicted and ground-truth masks, quantifying how similar the two sets are:

$$\text{Dice}(P,G) = \frac{2 \sum_i (p_i \cdot g_i)}{\sum_i p_i + \sum_i g_i + \epsilon}, \quad (2.17)$$

where $p_i, g_i \in \{0,1\}$ denote the predicted and ground truth binary labels for pixel i , and ϵ is a small constant to avoid division by zero. The Dice coefficient ranges from 0 (no overlap) to 1 (perfect overlap). In practice, a high Dice score indicates that the segmentation model is accurately capturing the region of interest (ROI), which is particularly critical in medical applications where precise delineation of lesions can directly influence diagnosis and treatment planning.

The Dice metric is especially useful when dealing with class imbalances because it gives more weight to correctly segmented lesion pixels. However, it can sometimes be overly sensitive to slight misalignments or boundary errors, meaning that even small discrepancies between the predicted and true boundaries may result in a lower score.

2.4.2 Intersection-over-Union (IoU)

The Intersection-over-Union (IoU), also known as the Jaccard Index, measures the fraction of the union of the predicted and ground truth regions that is correctly predicted:

$$\text{IoU}(P,G) = \frac{\sum_i (p_i \cdot g_i)}{\sum_i p_i + \sum_i g_i - \sum_i (p_i \cdot g_i) + \epsilon}. \quad (2.18)$$

Like the Dice coefficient, IoU values range from 0 to 1, where 1 indicates a perfect match between the prediction and the ground truth. IoU provides a slightly different perspective by explicitly considering the size of the union, and it is generally considered to be more conservative than the Dice coefficient. In other words, for the same segmentation, the IoU is often lower than the Dice score.

The Dice coefficient and Intersection over Union (IoU) are commonly used together to assess how well a model performs in segmentation tasks. The Dice score is useful for measuring overlap, while IoU helps compare different methods, especially in competitive situations.

Both metrics require turning prediction outputs into binary results, typically using a threshold like 0.5 for sigmoid outputs. The choice of threshold affects the scores, so it is important to select one based on how the model performed during validation.

we calculate these metrics for the entire test set to help with hyperparameter tuning and model

selection. High Dice and IoU scores show strong segmentation, which improves the quality of radiomics analysis and the reliability of clinical predictions. By using both metrics, we achieve a thorough assessment of segmentation quality. This is essential for accurately defining lesions in colorectal cancer, thus enhancing diagnostic processes and personalized treatments.

2.5 Data Augmentation and Preprocessing

In medical imaging, datasets are often limited, and patient anatomies vary greatly, which can hinder a model's ability to generalize well to unseen data. To address these challenges, we employ a variety of data augmentation techniques to artificially expand the training set and simulate real-world variability. These augmentations help mitigate overfitting and improve the robustness of the segmentation models.

Augmentation Techniques:

- **Resizing:** Each 2D MRI slice is rescaled to a consistent size (e.g., 224×224) to standardize the input dimensions across the dataset. This uniformity is crucial for batching slices in deep networks and ensures that the learned features are not biased by differences in resolution.
- **Horizontal Flip:** Slices are randomly flipped left-to-right with a probability of 0.5. This augmentation simulates left-right anatomical variations and helps the model become invariant to lateral asymmetries, which is especially useful in clinical settings.
- **Normalization:** Each slice is normalized on a per-image basis by subtracting its mean intensity and dividing by its standard deviation. This z-score normalization reduces the impact of varying intensity scales across different scanners and acquisition protocols. Corresponding masks are scaled to ensure that the lesion voxels are confined within the range $[0,1]$, maintaining consistency between images and labels.
- **Brightness and Contrast Adjustments:** Although not explicitly listed in the initial bullet points, our implementation also includes mild brightness and contrast shifts (typically in the range of $\pm 20\%$). These perturbations simulate differences in scanner settings and patient conditions, further enhancing model generalization.
- **Random Crop/Zoom:** We apply random cropping or zooming to introduce small spatial variations. This augmentation helps the network learn to focus on both local and contextual information, which is particularly useful for detecting irregular tumor shapes and boundaries.

3D to 2D Slice Conversion: In addition to the augmentations, we convert the 3D MRI volumes into individual 2D axial slices. This approach offers several advantages:

- It significantly increases the total number of training samples by decomposing each 3D volume into many 2D slices.
- Training on 2D slices reduces computational complexity and memory requirements compared to full 3D models.
- The 2D approach allows us to leverage standard 2D CNN architectures and well-established data augmentation libraries, such as `albumentations`, which we use to apply the aforementioned augmentations consistently to both images and masks.

Our preprocessing pipeline is implemented using Python libraries such as `NiBabel` for loading NIfTI files and `SimpleITK` for image processing tasks. The `albumentations` library is used to apply the random transformations, ensuring that the same augmentation is applied to both the MRI slice and its corresponding binary mask. This synchronization is crucial for preserving the alignment of tumor boundaries in the augmented data.

Overall, these augmentation and preprocessing steps are essential for building a robust segmentation model. By standardizing image dimensions, intensities, and applying realistic perturbations, we create a diverse and representative training dataset that improves the model’s ability to generalize to new images from different centers and scanners.

2.6 Optimization: Adam

2.6.1 Algorithm

We optimize model parameters via Adam [[kingma2014adam](#)], which adaptively scales the learning rate for each parameter based on the first and second moments of the gradients. Let g_t be the gradient at time t ; Adam computes the exponential moving averages of the gradients and their squares as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2. \quad (2.19)$$

These moving averages serve as estimates of the first moment (mean) and the second moment (uncentered variance) of the gradients. The weight update rule is then given by:

$$w_{t+1} = w_t - \alpha \frac{m_t / (1 - \beta_1^t)}{\sqrt{v_t / (1 - \beta_2^t) + \epsilon}}, \quad (2.20)$$

where α is the base learning rate, β_1 and β_2 are hyperparameters controlling the decay rates for these moving averages, and ϵ is a small constant added for numerical stability. This formulation allows each parameter to have its own learning rate, which is especially beneficial when dealing with the heterogeneous and sparse gradients often encountered in segmentation tasks.

2.6.2 Training Configuration

For our experiments, we configure the training process based on extensive hyperparameter tuning and preliminary trials. The following configuration was used consistently across all architectures:

- **Optimizer:** We employ the Adam optimizer with the parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, along with a learning rate of $\alpha = 1 \times 10^{-4}$. This setting has proven effective in stabilizing training and ensuring smooth convergence.
- **Batch Size:** We use a batch size of 64 slices per mini-batch, which offers a balance between computational efficiency and sufficient gradient estimation.
- **Number of Epochs:** Training is conducted for 10 to 50 epochs, depending on the dataset size and convergence behavior observed during preliminary experiments.

During training, the Binary Cross-Entropy (BCE) loss is computed on each mini-batch, and periodic evaluations on the validation set are performed using overlap-based metrics such as the Dice coefficient and Intersection-over-Union (IoU). These metrics guide hyperparameter tuning and provide early feedback on model performance.

In addition to the basic configuration, we implement the following strategies to further enhance the training process:

- **Early Stopping:** An early stopping criterion is applied based on the validation loss with a patience of 3 epochs. If no improvement is observed, training halts and the best-performing model weights are restored.
- **Learning Rate Scheduling:** While our base learning rate is set to 1×10^{-4} , experiments with learning rate decay or scheduling (such as reducing the learning rate on plateau) were considered to fine-tune convergence, especially for deeper or more complex architectures.
- **Regularization and Data Augmentation:** Given the relatively small size of our dataset, regularization techniques and extensive data augmentation (e.g., horizontal flips, brightness/contrast shifts) are crucial in preventing overfitting and ensuring that the network generalizes well to unseen data.

These optimization strategies not only improve the convergence speed but also enhance the stability of the training process, ensuring that the models—whether incorporating a traditional CNN encoder or advanced transformer-based modules—are effectively optimized for the challenging task of colorectal cancer segmentation.

2.7 Summary

In summary, this chapter provided a comprehensive overview of our MRI segmentation framework. We reviewed the fundamental role of convolutional and pooling layers in learning hierarchical features, with batch normalization used to stabilize training and mitigate internal covariate shift. The chapter detailed several segmentation architectures, including the classic U-Net, the UMamba variant with a deeper bottleneck and modified filter sizes, and SwinUMamba, which integrates SwinBlocks to enhance local feature extraction while preserving the U-Net-like structure. Additionally, we introduced a PVTv2-based encoder that leverages patch embedding and self-attention mechanisms for global, multi-scale context, and a Hybrid Segmentation Network (HSN) that effectively combines a pre-trained CNN encoder with skip-connected decoder layers to merge high-level semantic information with fine-grained spatial details. The training strategy employs BCE with logits as the loss function, with evaluation based on Dice and IoU metrics. Furthermore, the framework includes data augmentation and preprocessing techniques—such as resizing, flipping, normalization, and careful 3D-to-2D conversion—to ensure robustness against variability in MRI intensities and scanning parameters. Finally, optimization is performed using the Adam optimizer with an adaptive learning rate of 1×10^{-4} , which enhances both convergence speed and training stability.

Chapter 3

Dataset, MRI Acquisition, and Preprocessing

3.1 Introduction

This chapter provides a comprehensive overview of our workflow for data acquisition, preprocessing, and ground truth labeling, which forms the foundation for our deep learning segmentation and radiomics analyses. We begin by describing the collection of 3D MRI scans from multiple clinical sites and the subsequent expert manual segmentation using 3D Slicer. The resulting segmentation masks are then exported and converted into 2D axial slices, which serve as inputs for our deep learning models and radiomics feature extraction. Additionally, we outline our systematic data augmentation and preprocessing strategies designed to enhance model robustness and ensure reliable radiomics measurements.

3.2 Dataset Description

Our dataset consists of 3D MRI scans collected from 108 patients across several clinical sites. The following points highlight the key aspects of the dataset:

- **Patient Demographics and Clinical Diversity:** The patients span a wide age range (approximately 27–85 years), ensuring that the dataset captures a diverse set of anatomical and pathological variations. This diversity is crucial for developing robust segmentation models.
- **MRI Scans and File Formats:** Each patient’s complete 3D MRI volume was acquired using standard clinical protocols and is stored in the NIfTI format. The MRI images are saved as `.nii` files, while the corresponding segmentation labels are stored as compressed

.nii.gz files. This file format is widely used in medical imaging because it preserves multidimensional data along with essential metadata, such as voxel dimensions and orientation information.

- **Segmentation with 3D Slicer:** Before any 2D work, we imported the raw 3D MRI scans into 3D Slicer, a top-notch, free tool for analyzing medical images. Experienced radiologists used 3D Slicer’s advanced tools to manually segment the mesorectum and tumor areas. They combined automated methods, like thresholding and region growing, with manual drawing to clearly define the areas of interest. We made sure the process was thorough with strict quality control, which included a second review by an independent radiologist. This ensured the segmentation labels were accurate and clinically valid.

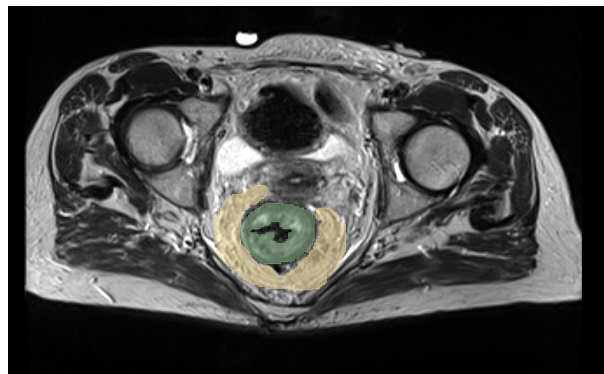


Figure 3.1: Segmented MRI scan highlighting the mesorectum (beige) and tumor (green) to improve colorectal cancer detection using deep learning techniques.

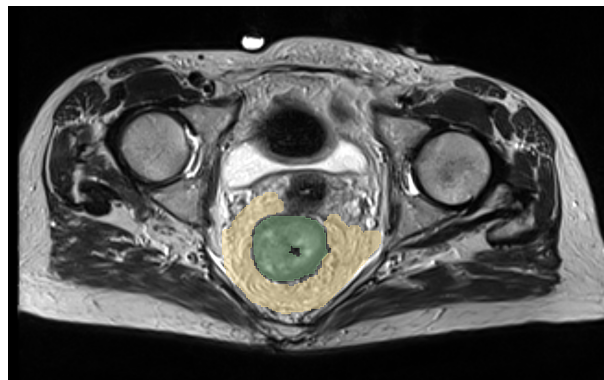


Figure 3.2: MRI-based tumor and mesorectum segmentation, essential for colorectal cancer analysis, facilitating automated medical image interpretation with AI models.

- **Slice Extraction:** To facilitate training with 2D convolutional neural networks (CNNs), axial slices were extracted from the 3D volumes. This process produced two synchronized sets of files:

- **slices/images:** Containing the individual 2D MRI slices.
- **slices/labels:** Containing the corresponding binary masks derived from the 3D Slicer segmentations.

Table 3.1: Dataset Overview and Key Variables

Variable	Value/Range	Description
Patients	108	Total patients in the dataset
Age Range	27–85	Corrected patient age range
Sequences	T1, T2, T2-FLAIR	MRI acquisition protocols
Brands	Siemens, GE, Philips, Hitachi	Scanner manufacturers
Field Strength	1.5T, 3.0T	Magnetic field strengths
Slices	40–160	Axial slices per 3D volume
Labels	Binary masks	Mesorectum and tumor masks
Data Split	70% / 15% / 15%	Train/Validation/Test split

3.3 Radiomics-Based Feature Extraction

Radiomics offers a systematic approach to extracting quantitative features from medical images, capturing detailed information about lesion morphology, texture, and intensity. In our study, radiomics plays a complementary role to deep learning by providing interpretable, handcrafted features that can be correlated with clinical endpoints.

- **First-Order Statistics:** Basic metrics such as mean intensity, variance, skewness, and kurtosis quantify the overall distribution of pixel intensities within the segmented region.
- **Higher-Order Texture Features:** Advanced descriptors are computed using gray-level co-occurrence matrices (GLCM) and gray-level run length matrices (GLRLM) to capture spatial relationships between pixels, texture patterns, and heterogeneity within lesions.
- **Feature Extraction Workflow:** Segmented slices are processed using standardized libraries (e.g., PyRadiomics) to ensure reproducibility. The resulting high-dimensional feature set can be combined with deep learning embeddings to enhance model performance.
- **Integration with Deep Learning:** By integrating radiomics features with CNN-derived representations, our approach benefits from both domain-specific handcrafted features and the hierarchical representations learned by neural networks.

3.4 MRI Acquisition Protocols

Since our dataset aggregates scans from multiple clinical sites, the MRI acquisition protocols vary slightly. Key characteristics include:

- **Scanner Brands:** Siemens, GE, Philips, Hitachi.
- **Field Strengths:** 1.5T and 3.0T.
- **Matrix Sizes:** Typically between 256×256 and 512×512 .
- **Slice Thickness:** Ranges from 1 mm to 5 mm.
- **MRI Sequences:** T1-weighted, T2-weighted, and T2-FLAIR.

Patients were scanned in the supine position using dedicated coils. Despite standardized protocols, minor variations in parameters (e.g., repetition time, echo time, flip angle) are inherent across different sites.

3.5 Data Preprocessing and Ground Truth Labeling

3.5.1 Initial Segmentation Using 3D Slicer

Each 3D MRI scan was imported into 3D Slicer, a state-of-the-art open-source platform for medical image analysis. Expert radiologists performed manual segmentation of the mesorectum and tumor regions using a combination of automated tools (e.g., thresholding and region growing) and manual contouring. This process ensured precise delineation of anatomical structures. Rigorous quality control was implemented by conducting a secondary review with an independent radiologist to verify the accuracy and consistency of the segmentations. The resulting high-quality segmentations serve as the ground truth for both deep learning training and radiomics analyses.

3.5.2 Visualization of MRI Segmentation and Preprocessing

To further validate and assess the segmentation process, we performed an extensive visualization of the segmented regions by overlaying the binary masks on the original MRI images. In these overlays, distinct colors were assigned to highlight critical anatomical structures—specifically, the tumor core, the mesorectum, and the surrounding tissues. This visualization not only serves to confirm the anatomical consistency of the segmentation but also provides insights into the inter-patient variability of the extracted features.

Several specialized libraries and tools were employed to facilitate the visualization and processing tasks.

NiBabel is utilized for loading and managing NIfTI (.nii) files, ensuring that the proper orientation and essential metadata (such as voxel dimensions and image orientation) are retained throughout the processing; OpenCV is applied for a range of image processing operations, including the generation of overlay images, intensity normalization, and general image enhancement; Matplotlib and Seaborn are employed to generate detailed plots of the segmentation results and to analyze and evaluate the distribution patterns of the radiomic features; and SimpleITK is leveraged for robust spatial transformations, image resampling, and the processing of multi-dimensional medical images, thereby ensuring that the spatial integrity of the images is maintained during the overlay process.

In parallel with the qualitative visualization, we conducted a rigorous quantitative assessment of the segmentation performance using several key metrics:

- **Dice Similarity Coefficient (DSC):** This metric quantifies the overlap between the predicted segmentation and the ground truth provided by manual delineations, with a value of 1 indicating perfect overlap and 0 indicating no overlap.
- **Hausdorff Distance:** This metric measures the maximum boundary discrepancy between the manual and model-generated segmentations, thereby providing an index of the worst-case error in boundary delineation.
- **Mean Intensity Difference (MID):** Employed to verify the consistency of intensity normalization across different MRI sequences, MID ensures that the statistical properties of the segmented regions are maintained.

Each segmentation mask was meticulously reviewed by independent radiologists, and any discrepancies were addressed through iterative refinement. The overlay visualization not only validates the anatomical accuracy of the segmentation but also reveals critical information regarding the variability of imaging features among different patients.

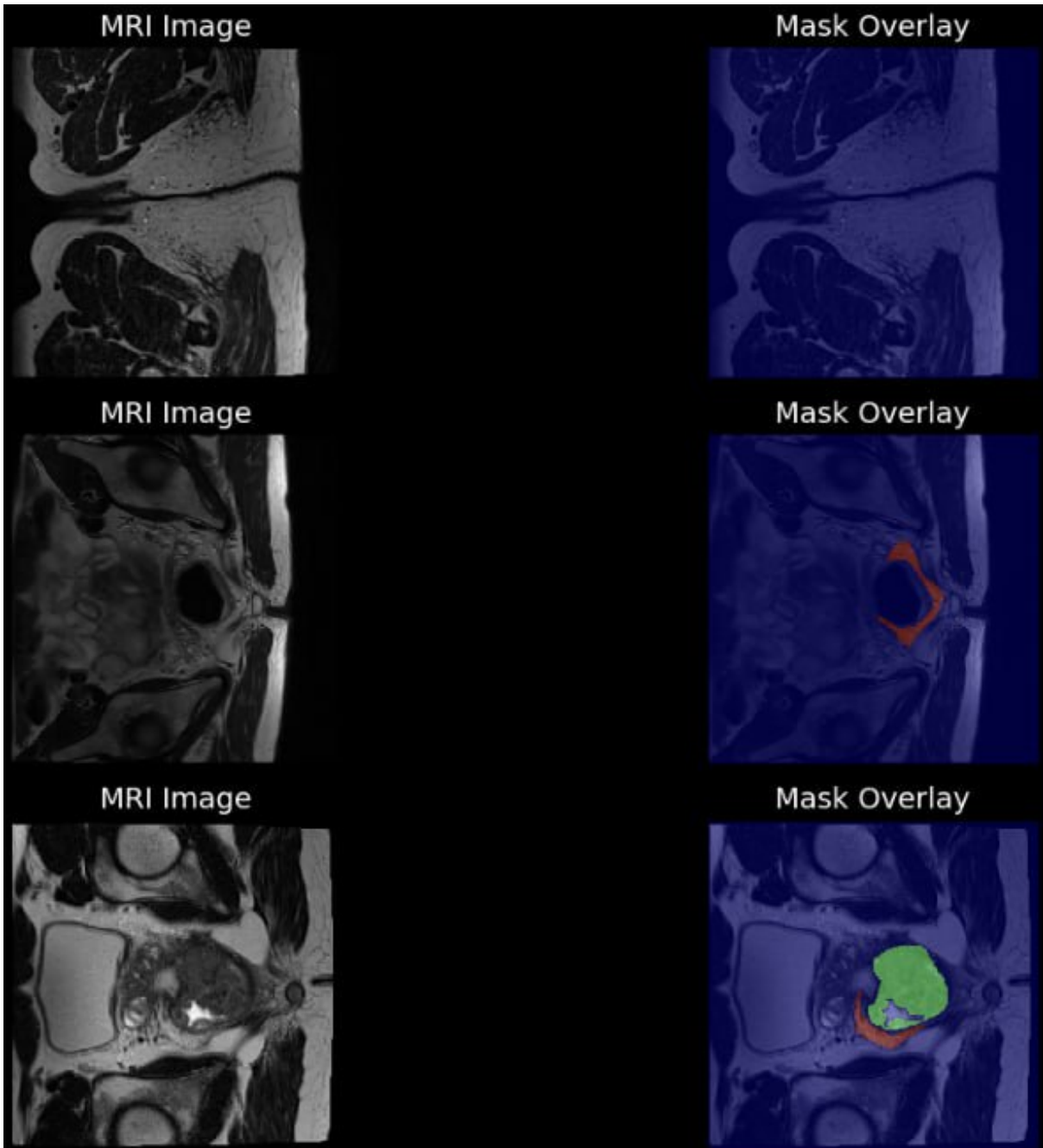


Figure 3.3: MRI images with corresponding segmentation masks overlaying tumor and mesorectum regions, illustrating deep learning-based colorectal cancer segmentation

3.6 Data Augmentation and Preprocessing

To improve the robustness and generalizability of our deep learning models in the face of scanner-induced variations and a limited dataset, we employ a comprehensive data augmentation and preprocessing strategy. First, each 2D MRI slice is resized to a uniform

resolution of 224×224 pixels, ensuring consistent input dimensions that facilitate efficient batching during training. We then apply random horizontal flips with a probability of 0.5 to simulate left-right positional variations, which helps account for natural anatomical asymmetries. In addition, we incorporate random brightness and contrast shifts—typically within a range of $\pm 20\%$ —to mimic differences in MRI contrast that can occur due to varying acquisition protocols or patient conditions. Random zoom and cropping operations, allowing scaling factors of up to $\pm 10\%$ followed by re-resizing to 224×224 , are also applied to simulate minor positional shifts and variations in the field-of-view. All these augmentations are applied consistently to both the image slices and their corresponding binary masks, ensuring that the alignment of tumor boundaries is maintained.

Each 2D slice undergoes per-slice z-score normalization, expressed as

$$I_{\text{norm}}(p) = \frac{I(p) - \mu}{\sigma},$$

where μ and σ are the mean and standard deviation of the pixel intensities in the slice. This normalization standardizes the intensity distribution across slices, reducing the impact of differences in scanner protocols or coil sensitivities, and thus simplifying the optimization process during training.

Although basic registration is typically performed in 3D Slicer, occasional residual misalignments may persist, particularly in multi-center datasets. To address this, we retain the option to apply rigid or affine registration during preprocessing in cases where motion artifacts or slight misalignments are observed. Post-registration, slices are cropped or padded as necessary to fit the 224×224 grid uniformly.

This augmentation and preprocessing pipeline increases our training set size by turning each 3D volume into several 2D slices. It also standardizes the data, which helps our deep learning models learn strong features even when images are captured in different ways. By using this approach, our models can better adapt to the various imaging conditions found in clinical practice.

3.7 Slice Extraction and Dataset Partitioning

Following the complete validation of segmentation masks, we decompose each 3D volume into axial slices. Each slice is saved as a `.nii` or `.png` file, accompanied by its binary mask. This step effectively transforms a single 3D MRI volume into multiple 2D samples, substantially increasing the total number of training images. In line with our multi-center

data approach, each patient’s collection of slices is stored together, ensuring that no single patient’s slices leak into multiple partitions.

We then finalize our dataset split at the **patient level**:

- **Training Set:** 70% of patients
- **Validation Set:** 15% of patients
- **Test Set:** 15% of patients

We confirm that slices from a single patient do not appear in more than one set, thereby preserving independence between subsets. Moreover, we aim for a balanced representation across centers in each split, mitigating overfitting to a single site’s protocol. This arrangement enables more realistic performance assessment, particularly when the final model is to be applied in broader clinical settings with mixed scanner types and field strengths.

3.8 Summary

In this chapter, we presented a comprehensive overview of our dataset and preprocessing strategy. We highlighted the multi-center nature of our data, which includes MRI scans from various sites and scanner manufacturers, ensuring that our training data captures a wide range of imaging variability. This diversity promotes model generalizability across different clinical environments. Additionally, we explained our choice of a 2D slice-based segmentation approach over full 3D methods, emphasizing its computational feasibility, the increased sample size obtained from decomposing 3D volumes into 2D slices, and its demonstrated success in rectal MRI tasks.

We then detailed our data augmentation techniques, which are essential for addressing scanner-induced variations and limited dataset sizes. These techniques include resizing images to a consistent 224×224 resolution, applying random horizontal flips, and incorporating brightness/contrast shifts and random zoom/cropping. All these augmentations are applied uniformly to both images and their corresponding masks to preserve accurate tumor boundaries. We also described the manual segmentation process conducted using 3D Slicer, followed by the extraction of axial slices and a careful patient-level partitioning into training, validation, and test sets.

Finally, we emphasized the importance of intensity normalization which standardize MRI data despite inter-scanner and inter-site variability. These preprocessing steps lay a solid

foundation for training deep learning models that can robustly handle different types of colorectal MRI scans.

Chapter 4

Deep Learning Models for Medical Image Segmentation

4.1 Introduction

Tumor segmentation lies at the heart of many critical tasks in colorectal cancer (CRC) management. Accurate and consistent delineation of tumor boundaries enables precise diagnosis, facilitates targeted therapy planning, and supports quantitative analyses such as radiomics. Recent advancements in deep learning have vastly improved segmentation accuracy in medical imaging, surpassing conventional methods that rely on handcrafted features or classical machine learning. Deep neural networks can automatically learn complex patterns from large volumes of data, offering robust performance even under varying scanner protocols and patient anatomies.

In this chapter, we focus on the architectural details and conceptual underpinnings of three primary convolutional neural network (CNN)-based approaches that we apply to 2D slices extracted from 3D MRI data. These models are:

1. **UNet with HSN (Hybrid Segmentation Network):** A traditional U-Net enhanced with a ResNet-based encoder and specialized normalization modules.
2. **UMamba with PVTv2 (Pyramid Vision Transformer v2):** A lightweight U-Net variant (*UMamba*) that integrates transformer-based global attention mechanisms through the PVTv2 backbone.
3. **Swin with PVTv2:** A more advanced design combining Swin Transformer concepts for local windowed attention with multi-scale feature extraction from PVTv2.

Each model represents a distinct trade-off between complexity and efficiency, targeting the unique challenges of CRC segmentation, such as irregular tumor shapes, subtle intensity contrasts, and inter-patient variability.

4.1.1 Implementation Details Consistent with the Code

Across all three architectures, our codebase (PyTorch-based) typically employs:

- **Optimizer:** Adam with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-4} .
- **Loss Function:** `BCEWithLogitsLoss`, which internally applies the sigmoid function and then computes binary cross-entropy, improving numerical stability.
- **Batch Size:** 64 slices per mini-batch (for both training and validation), reflecting the GPU memory constraints and the relatively modest spatial size (224×224) of each 2D slice.
- **Epochs and Early Stopping:** Up to 20 epochs, with an early stopping criterion (`patience = 3`) that reverts to the best-performing weights if the validation loss does not improve for 3 consecutive epochs.

4.2 UNet with Hybrid Segmentation Network (HSN)

The classical U-Net architecture by Ronneberger *et al.* [26] remains a cornerstone of medical image segmentation. It couples a contracting path for feature extraction with a symmetric expanding path for precise localization. Despite U-Net’s popularity and success, certain limitations arise when dealing with complex MRI intensity distributions and scanner-induced variations, which are common in multi-center colorectal studies. To mitigate these issues, we incorporate a **Hybrid Segmentation Network (HSN)** module and a ResNet-18 encoder from the *timm* library.

4.2.1 Key Components

1. **ResNet-18 Encoder:** We replace the traditional U-Net’s plain convolutional encoder with a ResNet-18 pretrained on a large natural image dataset. This encoder captures robust low-level edges and shapes, leveraging pre-learned weights that transfer effectively to medical imaging tasks. The `features_only` mode extracts intermediate layer outputs, which feed into the U-Net-like decoder.

2. **HSN Block:** While U-Net typically uses Batch Normalization (BN) or Instance Normalization (IN), HSN can adaptively handle the intensity shifts by combining spatial normalization with morphological awareness. This helps unify feature distributions across slices that may come from different scanners or vary widely in intensity.
3. **Lightweight Decoder:** Instead of a full multi-stage decoder, we utilize a single `nn.Upsample` layer in tandem with a final `Conv2d`. This minimalistic decoder reduces computational overhead, allowing focus on the powerful encoder features and hybrid normalization layers.

HSN improves resilience to inconsistent image characteristics, an important advantage in CRC imaging where protocols vary across sites. The pretrained encoder also accelerates convergence and mitigates overfitting, especially in scenarios with modest dataset sizes. However, because the decoder is simplified (fewer skip connections than classical U-Net), capturing very fine-grained details around lesion margins may prove more challenging.

4.2.2 Impact on Colorectal Cancer Segmentation

”UNet with HSN” is a reliable option for colorectal tumor segmentation, balancing speed and accuracy. It often serves as a starting point in multi-step processes, allowing for further improvements through techniques like morphological post-processing or radiomics analyses. Its moderate number of parameters makes it relatively easy to train, making it a good choice for clinical prototypes or settings with limited resources.

4.3 UMamba with PVTv2 (Pyramid Vision Transformer v2)

While CNN-based encoders excel at local feature extraction, they may struggle to capture long-range dependencies and global context, particularly in images with large anatomical variability. Transformer-based architectures introduce attention mechanisms that can learn complex relationships between distant regions. *Pyramid Vision Transformer v2 (PVTv2)* extends this concept by subdividing the feature maps at multiple scales, thus achieving better global context modeling with manageable computational costs.

4.3.1 Model Architecture

The **UMamba** network modifies the standard U-Net design by reducing the number of convolutional channels while retaining the essential encoder-decoder framework. This keeps the network lightweight. The **PVTv2 backbone**, imported via `timm`, then serves as the encoder. The final feature maps from PVTv2 are projected through a Conv2d layer to produce a single-channel segmentation map, and an upsampling operation restores the spatial resolution to 224×224 .

Important aspects include:

- **Attention Blocks:** Nested within the PVTv2, these blocks compute self-attention across tokens at each stage, enabling the model to weigh relationships between spatially separate tumor pixels and background tissues.
- **Multi-Scale Feature Extraction:** By progressively downsampling token embeddings, PVTv2 captures hierarchical representations from coarse to fine, which is especially beneficial in detecting multi-resolution features in MRI scans.
- **Reduced Decoder Complexity:** UMamba parallels the minimalistic decoder strategy described for UNet-HSN, thus leaning on the powerful multi-scale features from the encoder rather than a complex multi-layer upsampling path.

In colorectal cancer, tumors may invade or extend through mesorectal fat planes. PVTv2’s attention modules help track these elongated or spread-out regions. The global context awareness can improve the network’s ability to detect subtle boundaries, bridging separated tumor areas that might otherwise appear as discrete regions to a purely convolution-based encoder. In practice, UMamba with PVTv2 tends to surpass classical U-Net solutions in terms of Dice and IoU scores, especially for heterogeneous or irregular masses.

Transformer architectures, even lightweight ones, are generally more memory-intensive. Longer training times can be expected, particularly during the attention computations. While UMamba mitigates some of these concerns by restricting channel widths and decoder complexity, the memory footprint may still exceed that of simpler networks. Furthermore, if the dataset is small, careful regularization or data augmentation becomes crucial to prevent overfitting.

4.4 Swin with PVTv2: Merging Local Windowed Attention and Multi-Scale Analysis

Swin Transformers were introduced to handle larger images by splitting them into local windows where self-attention is calculated. This window-based approach reduces the quadratic complexity typical of global Transformers, allowing for efficient modeling of local neighborhoods. The *PVTv2* structure, on the other hand, progressively downsamples feature maps to achieve multi-scale learning. By uniting these concepts, the **Swin with PVTv2** architecture aspires to simultaneously capture local detail and large-scale context, aiming to address the notoriously complex boundary delineation problems in CRC imaging.

In this hybrid design, a Swin Base model pre-trained on natural images can serve as the main encoder, extracting high-dimensional feature maps from input slices. The final stage typically outputs a 1024-channel tensor. Meanwhile, PVTv2-style multi-scale processing ensures that feature maps are progressively compressed in spatial size and augmented in channel depth. A final Conv2d reduces channels to 1, and an `nn.Upsample` layer reverts the feature map to the original slice resolution.

Key highlights include:

- **Window-Based Attention:** Breaks the image into small overlapping patches or “windows,” performing local self-attention that is more computationally tractable than full global attention.
- **Shifted Windowing Strategy:** Over successive layers, the window positions shift, enabling cross-window interactions and effectively capturing a larger context without incurring the cost of full attention across the entire slice.
- **Hierarchical Representation:** Stacked blocks further downsample the embeddings, blending local and global signals to form a robust representation for the final segmentation step.

Colorectal tumors often contain a blend of subtle intensity variations, necrotic cores, and infiltration into adjacent tissues. Local window attention is particularly adept at identifying these subtle texture cues, while the overall multi-scale approach retains broad contextual awareness—critical for identifying tumor continuity across multiple image regions. Empirically, Swin with PVTv2 often yields higher IoU and Dice coefficients than simpler CNN-based models. The synergy of local-window attention and progressive feature

downsampling frequently proves advantageous when delineating fine tumor edges or capturing extended shapes.

4.5 Deep Learning and Radiomics Synergy

Deep learning segmentation models generate finely delineated boundaries that help isolate tumor regions. In parallel, **radiomics** quantifies morphological and textural attributes of those regions, facilitating correlation with clinical biomarkers or outcomes. This synergistic relationship yields:

- **High-Fidelity ROIs:** Automated segmentation replaces or augments the manual contouring step, saving time and reducing inter-observer variability. Such consistency is key for deriving robust radiomics features.
- **Quantitative Feature Extraction:** Once masks are finalized, radiomics can compute shape-based metrics (e.g., roundness, elongated contours), texture-based indices (e.g., GLCM, GLRLM), or wavelet-transformed features indicative of tumor heterogeneity.
- **Predictive Modeling:** Fusing deep representations with handcrafted radiomics can yield powerful predictive models for therapy response or survival. In CRC, understanding features such as infiltration depth or local texture variation can be critical for clinical decisions.
- **Explainability and Transparency:** Pure deep learning approaches are sometimes criticized for lacking interpretability. Radiomics features, however, can present more intuitive descriptors (e.g., “this lesion has high edge contrast”), aiding clinicians in understanding why certain predictions are made.

Notably, the success of this integration hinges on the precision of the segmentation masks produced by each architecture. Minor boundary inaccuracies may inflate or distort radiomics features (especially those tied to shape or boundary complexity), diminishing their clinical reliability. Therefore, the choice of a robust segmentation model, coupled with thorough validation, remains paramount when bridging deep learning and radiomics.

4.6 Summary

This chapter has explored three deep learning architectures tailored for 2D MRI slice segmentation in colorectal cancer:

- **UNet with HSN**, which builds upon a traditional U-Net structure and adds a specialized normalization approach plus a pretrained ResNet-18 encoder,
- **UMamba with PVTv2**, which integrates multi-scale transformer attention into a lightweight U-Net design, and
- **Swin with PVTv2**, combining local-window attention from the Swin Transformer with the hierarchical scaling of PVTv2 for more sophisticated feature representation.

Each network strikes a different balance of complexity and speed. **UNet with HSN** is relatively straightforward and economical in terms of GPU usage, yet it can miss intricate boundaries if the dataset is highly diverse. **UMamba with PVTv2** offers improved global context modeling via attention blocks, often outperforming basic CNN solutions, albeit at higher computational cost. Finally, **Swin with PVTv2** typically delivers the most accurate segmentation masks among the three architectures but demands more robust hardware and tuning.

Chapter 5

Results and Experiments

5.1 Introduction

This chapter presents a comprehensive evaluation of our deep learning-based segmentation framework for MRI-based colorectal cancer detection. We compare the performance of three models—UNet-HSN, UMamba-PVTv2, and Swin-PVTv2—using multiple evaluation metrics, including the **Dice Similarity Coefficient (DSC)**, **Intersection over Union (IoU)**, and **validation accuracy**, as well as an analysis of loss trends. Additionally, we investigate the role of radiomics-based feature extraction in enhancing segmentation quality and tumor delineation. The experiments address both the accuracy of tumor segmentation and the robustness of the training process, with an eye toward clinical applicability.

5.2 Experimental Setup

5.2.1 Dataset and Preprocessing

The MRI scans used in this study were obtained from a clinical database and preprocessed using tools such as NiBabel and SimpleITK. The preprocessing workflow consisted of the following steps:

- **Normalization:** Each 2D axial slice was normalized using z-score normalization (subtracting the mean and dividing by the standard deviation) to standardize pixel intensities.

- **Resizing:** All images were resized to a uniform resolution of 224×224 pixels to ensure consistency across the dataset.
- **Data Augmentation:** Techniques such as horizontal flipping, random cropping/zooming, contrast normalization, and intensity scaling were applied to both images and corresponding masks to prevent overfitting.
- **Segmentation Preprocessing:** Each 3D MRI scan was imported into 3D Slicer, where expert radiologists manually segmented the mesorectum and tumor regions. The resulting 3D segmentations were then converted into 2D axial slices.

The dataset was partitioned at the patient level into:

- **Training Set:** 70%
- **Validation Set:** 15%
- **Test Set:** 15%

This partitioning strategy helps avoid data leakage and ensures that the models are evaluated on unseen patient data.

5.2.2 Training Details

The training process was configured based on extensive hyperparameter tuning and preliminary experiments:

- **Loss Function:** A hybrid loss function combining Binary Cross-Entropy (BCE) and Dice loss was used to optimize both pixel-wise accuracy and overall overlap.
- **Optimizer:** Adam optimizer with a learning rate of 1×10^{-4} and a weight decay of 1×10^{-4} .
- **Batch Size and Epochs:** Training was performed with a batch size of 64 slices per mini-batch for up to 20 epochs. Early stopping based on validation loss was implemented to prevent overfitting.
- **Evaluation Metrics:** Performance was monitored using the Dice Score, IoU Score, and overall validation accuracy.

5.3 Performance Evaluation

5.3.1 Quantitative Analysis

Table 5.1 summarizes the final performance metrics for each model on the test set.

Table 5.1: Final Model Performance

Model	Dice Score	Validation Accuracy	IoU Score
UNet-HSN	0.6153	98.63%	0.6028
UMamba-PVTv2	0.6957	98.90%	0.6509
Swin-PVTv2	0.7452	98.94%	0.6961

As seen in Table 5.1, **Swin-PVTv2** outperformed the other models across all evaluation metrics. The higher Dice and IoU scores suggest that this model achieves a more accurate overlap between the predicted segmentation masks and the ground truth, especially in complex tumor regions. While **UNet-HSN** effectively preserves spatial details via skip connections, it falls short in capturing irregular tumor boundaries.

5.3.2 Loss and Convergence Trends

The training and validation loss curves for each model are shown in Figures 5.1, 5.2, and 5.3. These trends illustrate the learning stability and generalization capabilities of the models.

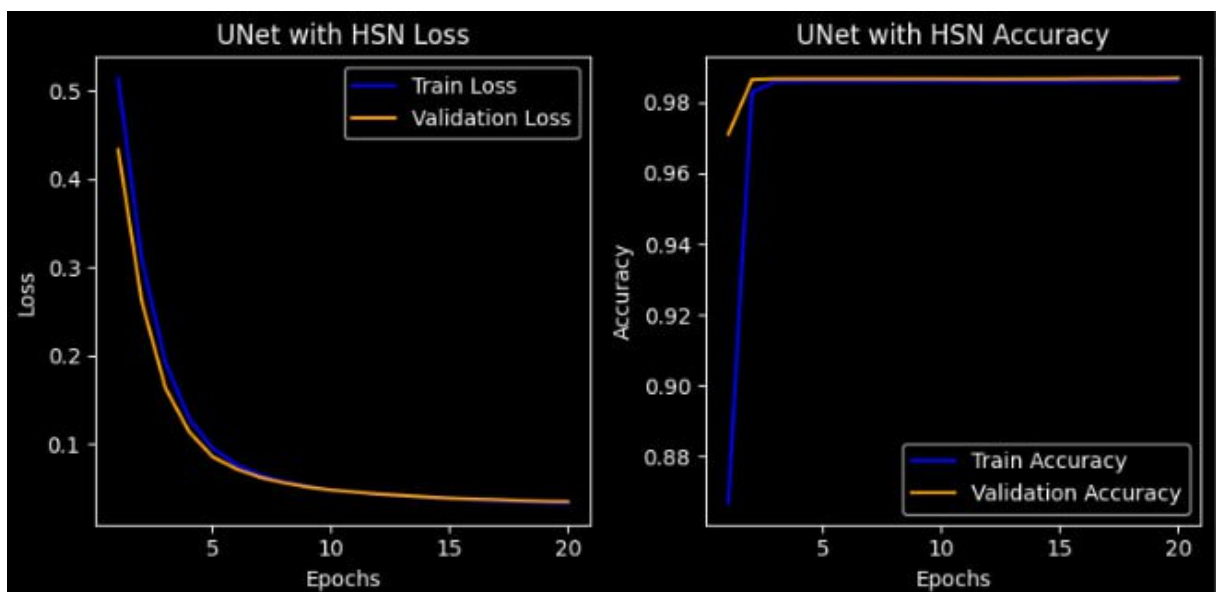


Figure 5.1: UNet-HSN training and validation loss (left) and accuracy (right) trends.

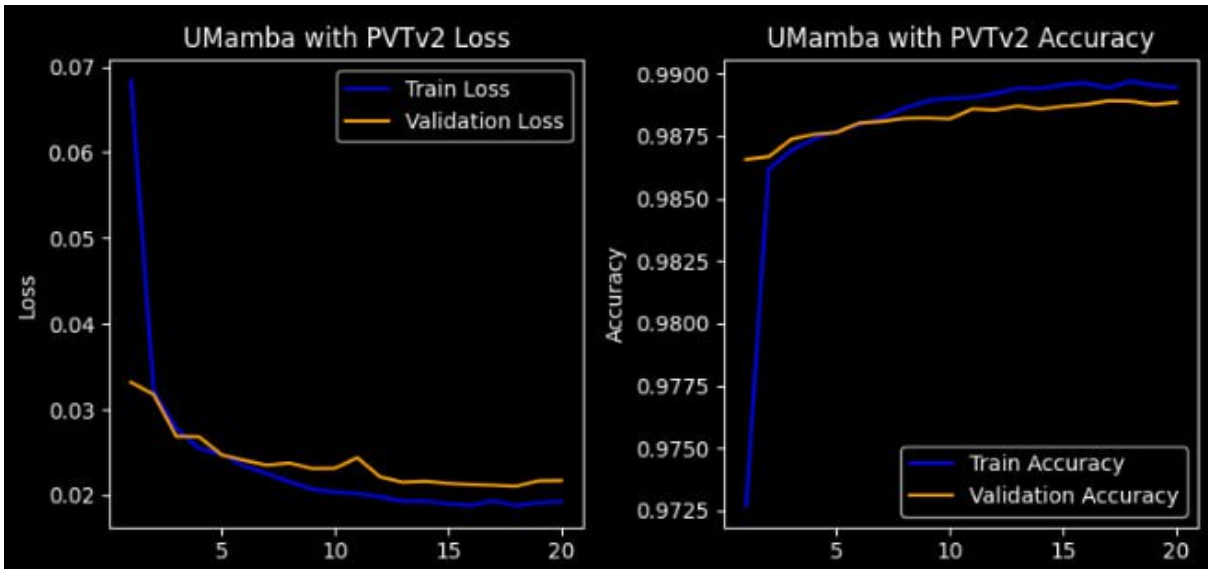


Figure 5.2: UMamba-PVTv2 training and validation loss (left) and accuracy (right) trends.

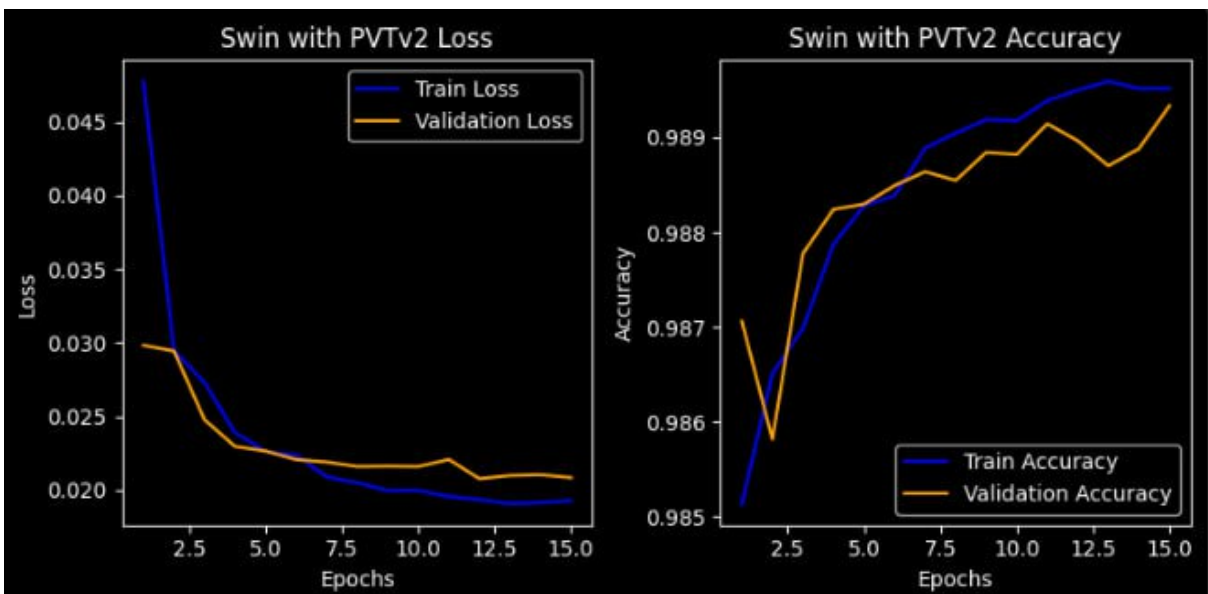


Figure 5.3: Swin-PVTv2 training and validation loss (left) and accuracy (right) trends.

5.4 Qualitative Analysis

Visual inspection of the segmentation outputs was performed to assess the qualitative performance of the models:

- **UNet-HSN:** Preserved spatial details well but occasionally failed to delineate complex tumor boundaries.

- **UMamba-PVTv2:** Improved the segmentation of tumor boundaries, though it sometimes produced spurious regions (false positives).
- **Swin-PVTv2:** Generated segmentation masks that closely resembled the ground truth, accurately delineating both overall tumor shape and fine boundary details.

5.5 Radiomics Feature Analysis

5.5.1 Radiomics Feature Extraction Results

Radiomics provides a quantitative framework for extracting features that characterize tumor morphology, texture, and intensity from segmented MRI data. In our study, the extracted features are grouped into four major categories: Intensity-Based, Texture-Based (GLCM), Morphological/Shape-Based, and Wavelet-Based features. The following tables summarize the statistical metrics for each category.

1. Intensity-Based Features

Table 5.2: Intensity-Based Features

Feature	Mean \pm SD	Max
Mean Intensity	11.30 \pm 37.57	192.44
Standard Deviation of Intensity	5.02 \pm 17.28	95.16

2. Texture-Based Features (GLCM)

Table 5.3: Texture-Based Features (GLCM)

Feature	Mean \pm SD	Max
Contrast	6.17 \pm 25.93	272.79
Correlation	0.07 \pm 0.23	0.99
Homogeneity	0.085 \pm 0.27	0.99
Energy	0.085 \pm 0.27	0.99

3. Morphological and Shape-Based Features

Table 5.4: Morphological and Shape-Based Features

Feature	Mean \pm SD	Max
Area	133.24 \pm 664.35	11,866 mm ²
Perimeter	17.10 \pm 65.37	752.82 mm
Eccentricity	0.058 \pm 0.19	0.978
Solidity	0.069 \pm 0.23	1.0

4. Wavelet-Based Features

Table 5.5: Wavelet-Based Features

Feature	Mean \pm SD	Max
Wavelet Mean	0.23 \pm 1.03	12.96
Wavelet Standard Deviation	2.18 \pm 7.90	61.93

5.5.2 Clinical Relevance of Radiomics Features

The radiomics features extracted in this study provide critical quantitative biomarkers that enhance our understanding of tumor characteristics and complement deep learning segmentation outputs. Their clinical relevance is discussed as follows:

Intensity-Based Features:

- **Mean Intensity** reflects the overall brightness of the tumor region. A high standard deviation indicates significant inter-patient variability in tumor density, which may be associated with differences in tissue composition, such as regions of necrosis or fibrosis.
- **Standard Deviation of Intensity** measures the dispersion of intensity values within the tumor, serving as a direct indicator of tumor heterogeneity.

Texture-Based Features (GLCM):

- **Contrast** quantifies local intensity variations. Elevated contrast values suggest that the tumor has a heterogeneous internal structure, which can be linked to aggressive tumor behavior.
- **Correlation** assesses spatial dependency between neighboring pixels; lower values indicate irregular tumor architecture, while higher values suggest more uniform tissue.

- **Homogeneity** and **Energy** provide complementary measures of texture uniformity. Lower homogeneity and energy values are typically associated with increased textural complexity, indicative of malignant tumor characteristics.

Morphological and Shape-Based Features:

- **Area** and **Perimeter** offer direct measurements of tumor size and boundary complexity, respectively. Variability in these features may reflect differences in tumor progression and stage.
- **Eccentricity** quantifies tumor elongation; higher values may indicate more irregular and invasive tumor shapes.
- **Solidity** describes the compactness of the tumor relative to its convex hull. Lower solidity values suggest a fragmented or infiltrative structure, which is often correlated with aggressive malignancies.

Wavelet-Based Features:

- **Wavelet Mean** captures multi-scale intensity variations, revealing structural details that are not apparent in the original image.
- **Wavelet Standard Deviation** measures the variability across different wavelet scales, which is indicative of complex microstructural patterns within the tumor.

Overall, integrating these radiomic features with deep learning segmentation outputs provides a comprehensive approach to tumor characterization. The quantitative biomarkers derived from these features can potentially improve the accuracy of diagnostic, prognostic, and therapeutic decision-making in colorectal cancer management.

5.6 Discussion

5.6.1 Model Performance Interpretation

The experimental results suggest that:

- **Swin-PVTv2** achieves the highest segmentation accuracy, likely due to its hierarchical attention mechanisms that capture both global context and fine local details.
- **UMamba-PVTv2** performs competitively by leveraging advanced multi-scale feature extraction, although its training exhibited slightly more variability.

- **UNet-HSN** effectively preserves local spatial details through skip connections but struggles with complex tumor boundaries.

5.6.2 Synergy Between Radiomics and Deep Learning

Integrating radiomics features with deep learning models provides several benefits:

- **Enhanced Interpretability:** Radiomics features offer quantitative descriptors that help explain the segmentation outcomes.
- **Improved Segmentation Accuracy:** The additional information from radiomics can refine segmentation predictions, particularly in heterogeneous tumor regions.
- **Clinical Relevance:** Quantitative radiomics metrics, such as texture and shape descriptors, can support clinical decision-making and prognosis.

5.6.3 Future Directions

Future work should focus on:

- **Multi-Center Validation:** Expanding the dataset to include more diverse patient populations to improve model generalizability.
- **Hybrid Architectures:** Further exploration of hybrid CNN-Transformer models to combine the strengths of both approaches.
- **Real-Time Deployment:** Optimizing models for integration into clinical workflows.
- **Advanced Post-Processing:** Developing techniques to reduce false positives and enhance segmentation boundary accuracy.

5.7 Limitations and Challenges

Despite the promising results, several limitations must be acknowledged:

- **Dataset Size:** The limited number of patients (108) may restrict the generalizability of the models.
- **Computational Complexity:** Transformer-based architectures, especially those with multi-scale attention, demand significant computational resources.

- **Segmentation Accuracy:** Occasional false positives and difficulties in delineating irregular tumor boundaries indicate the need for further model refinement.
- **Radiomics Integration:** The incorporation of radiomics features adds complexity and requires careful statistical validation.

5.8 Summary

In summary, our experiments showed that the Swin-PVTv2 model was better at finding tumors compared to other models. This is clear from its higher Dice scores, Intersection over Union (IoU), and validation accuracy. The model's ability to capture both global and local features helped it mark tumor boundaries more accurately. Additionally, using radiomics features gave us helpful details about the tumor's shape, texture, and intensity. These improvements make the results easier to understand and could lead to more accurate clinical predictions and tailored treatment plans.

However, there are still challenges to address. The small dataset might limit how widely our findings can be applied. The Swin-PVTv2 model also needs a lot of computer power, which can be a barrier in places with fewer resources. There were some segmentation errors, so we need to refine the model further. Future research should work on expanding the dataset to cover more cases, improving hybrid models that combine CNN and Transformer features, and making these models more suitable for real-world clinical use.

Bibliography

- [1] AJCC Cancer Staging Manual 8th Edition. *AJCC Cancer Staging Manual 8th Edition*. <https://cancerstaging.org>. 2018.
- [2] Melina Arnold et al. “Global patterns and trends in colorectal cancer incidence and mortality”. In: *Gut* 66.4 (2017), pp. 683–691.
- [3] Mitchell S. Cappell. “Pathophysiology, clinical presentation, and management of colon cancer”. In: *Gastroenterology Clinics of North America* 37.1 (2008), pp. 1–24.
- [4] Margherita De Rosa et al. “Genetics, diagnosis and management of colorectal cancer (Review)”. In: *Oncology Reports* 34.3 (2015), pp. 1087–1096.
- [5] Evelien Dekker et al. “Colorectal cancer”. In: *The Lancet* 394.10207 (2019), pp. 1467–1480.
- [6] G. Di Costanzo et al. “Artificial intelligence and radiomics in magnetic resonance imaging of rectal cancer: a review”. In: *Exploration of Targeted Antitumor Therapy* 4 (2023), pp. 406–421. doi: 10.37349/etat.2023.00142. url: <https://doi.org/10.37349/etat.2023.00142>.
- [7] S. Fan et al. “Computed tomography-based radiomic features could potentially predict microsatellite instability status in stage II colorectal cancer: A preliminary study”. In: *Academic Radiology* 26.12 (2019), pp. 1633–1640.
- [8] Karthik Ganesh et al. “Immunotherapy in colorectal cancer: rationale, challenges and potential”. In: *Nature Reviews Gastroenterology & Hepatology* 16.6 (2020), pp. 361–375.
- [9] Shifu Gao, Feng Li, and Yanan Dong. “Magnetic resonance imaging in the preoperative evaluation of rectal cancer: an overview”. In: *Clinical Imaging* 64 (2020), pp. 86–95.

- [10] David S. Garcia et al. “Deep learning-based segmentation and quantification of tumors in rectal cancer using MRI”. In: *Physics in Medicine & Biology* 65.21 (2020), p. 215001.
- [11] Jan S. Golia Pernicka et al. “Radiomic-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation”. In: *Abdominal Radiology* 44.11 (2019), pp. 3755–3763.
- [12] Jorge Tapias Gomez et al. “Swin transformers are robust to distribution and concept drift in endoscopy-based longitudinal rectal cancer assessment”. In: *arXiv preprint* 2405.03762 (2025). doi: 10.48550/arXiv.2405.03762. url: <https://arxiv.org/abs/2405.03762>.
- [13] V. González-Castro et al. “CT radiomics in colorectal cancer: Detection of KRAS mutation using texture analysis and machine learning”. In: *Applied Sciences* 10.18 (2020), p. 6214.
- [14] Richard M. Gore and Marc S. Levine. *Textbook of Gastrointestinal Radiology (5th Edition)*. Elsevier, 2015.
- [15] Lu-Lu Jia et al. “Artificial intelligence with magnetic resonance imaging for prediction of pathological complete response to neoadjuvant chemoradiotherapy in rectal cancer: A systematic review and meta-analysis”. In: *Frontiers in Oncology* 12 (2022), p. 1026216. doi: 10.3389/fonc.2022.1026216. url: <https://doi.org/10.3389/fonc.2022.1026216>.
- [16] Yumei Jin et al. “Predicting tumor deposits in rectal cancer: a combined deep learning model using T2-MR imaging and clinical features”. In: *Insights into Imaging* 14 (2023), p. 221. doi: 10.1186/s13244-023-01564-w. url: <https://doi.org/10.1186/s13244-023-01564-w>.
- [17] Vinay Kumar et al. *Robbins & Cotran Pathologic Basis of Disease*. 10th. Elsevier, Philadelphia, PA, 2020.
- [18] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *Nature* 521.7553 (2015), pp. 436–444.
- [19] Joohyung Lee et al. “Moving from 2D to 3D: Volumetric Medical Image Classification for Rectal Cancer Staging”. In: *arXiv preprint* 2209.05771 (2022). doi: 10.48550/arXiv.2209.05771. url: <https://arxiv.org/abs/2209.05771>.
- [20] Ze Liu et al. “Swin Transformer: Hierarchical vision transformer using shifted windows”. In: *IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021, pp. 10012–10022.

- [21] Xiaoying Lou et al. “Deep learning model for predicting the pathological complete response to neoadjuvant chemoradiotherapy of locally advanced rectal cancer”. In: *Frontiers in Oncology* 12 (2022), p. 807264. doi: 10.3389/fonc.2022.807264. url: <https://doi.org/10.3389/fonc.2022.807264>.
- [22] Michael G. Lubner et al. “CT Texture Analysis: Definitions, applications, biologic correlates, and challenges”. In: *RadioGraphics* 37.5 (2017), pp. 1483–1503.
- [23] Finlay A. Macrae. *Colorectal cancer: Epidemiology, risk factors, and protective factors*. UpToDate. 2021.
- [24] National Comprehensive Cancer Network. *NCCN Clinical Practice Guidelines in Oncology: Colon Cancer (Version 2.2021)*. https://www.nccn.org/professionals/physician_gls/pdf/colon.pdf. 2021.
- [25] Prashanth Rawla, Thandra Sunkara, and Artur Barsouk. “Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors”. In: *Przegląd Gastroenterologiczny* 14.2 (2019), pp. 89–103.
- [26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional networks for biomedical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer. 2015, pp. 234–241.
- [27] Arnaldo Stanzione et al. “Radiomics and machine learning applications in rectal cancer: Current update and future perspectives”. In: *World Journal of Gastroenterology* 27.32 (2021), pp. 5306–5321. doi: 10.3748/wjg.v27.i32.5306. url: <https://dx.doi.org/10.3748/wjg.v27.i32.5306>.
- [28] Elena M. Stoffel and Caitlin C. Murphy. “Epidemiology and mechanisms of the increasing incidence of colon and rectal cancers in young adults”. In: *Gastroenterology* 158.2 (2020), pp. 341–353.
- [29] Hyuna Sung et al. “Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. In: *CA: A Cancer Journal for Clinicians* 71.3 (2021), pp. 209–249.
- [30] N. Taguchi et al. “CT texture analysis for the prediction of KRAS mutation status in colorectal cancer via a machine learning approach”. In: *European Journal of Radiology* 118 (2019), pp. 38–43.
- [31] World Health Organization, International Agency for Research on Cancer (IARC). *Latest global cancer data: Cancer burden rises to 19.3 million new cases and 10.0 million cancer deaths in 2020*. https://www.iarc.fr/wp-content/uploads/2020/12/pr292_E.pdf. 2020.

- [32] C. Wu et al. “Radiomics analysis of iodine-based material decomposition images with dual-energy CT imaging for preoperative predicting microsatellite instability status in colorectal cancer”. In: *Frontiers in Oncology* 9 (2019), p. 1250.
- [33] Mingwei Yang et al. “Deep learning for MRI lesion segmentation in rectal cancer”. In: *Frontiers in Medicine* 11 (2024), p. 1394262. doi: 10 . 3389 / fmed . 2024 . 1394262. url: <https://doi.org/10.3389/fmed.2024.1394262>.
- [34] W. Zhang et al. “Development and validation of magnetic resonance imaging-based radiomics models for preoperative prediction of microsatellite instability in rectal cancer”. In: *Annals of Translational Medicine* 9.2 (2021), p. 134.