UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

DEVELOPING NOVEL IMAGING-BASED CLINICAL OUTPUT PROGNOSTICATION MODELS AND IDENTIFICATION OF IMAGING BIOMARKERS IN OVARIAN CANCER PATIENTS

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By
NEMAN ABDOLI
Norman, Oklahoma
2026

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/343206

DEVELOPING NOVEL IMAGING-BASED CLINICAL OUTPUT PROGNOSTICATION MODELS AND IDENTIFICATION OF IMAGING BIOMARKERS IN OVARIAN CANCER PATIENTS

A DISSERTATION APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Yuchen Qiu (Chair)

Dr. Javeir Jo

Dr. Hjalti Sigmarsson

Dr. Wei Chen

©Copyright by NEMAN ABDOLI 2026

All Rights Reserved.

Dedications

I dedicate this dissertation to my beloved parents, Nasrin and Manouchehr, whose unconditional love, endless support, and countless sacrifices have made this journey possible. Their encouragement, wisdom, and faith in me have been a constant source of strength throughout my life.

I would also like to dedicate this work to my beloved wife, Youka, whose love, patience, understanding, and unwavering support have sustained me through every challenge and achievement along the way. Her companionship has been my greatest blessing, and this accomplishment would not have been possible without her encouragement and belief in me.

To Mom, Dad, and Youka, I offer my deepest love, gratitude, and appreciation. This work is dedicated to you.

Acknowledgments

I extend my deepest gratitude to my Ph.D. advisor, Dr. Yuchen Qiu, for his unwavering dedication, mentorship, and support throughout my doctoral studies. His guidance in research design, scientific writing, and data analysis inspired me to surpass my own expectations and tackle challenges I once thought insurmountable. His commitment to fostering my scientific curiosity and professional growth has been instrumental to my success.

I am sincerely grateful to my committee members, Dr. Wei R. Chen, Dr. Javier A. Jo, and Dr. Hjalti Sigmarsson, for their invaluable guidance, encouragement, and insightful feedback. Their expertise and perspectives greatly enriched this work and were essential to the completion of my doctoral journey. I hope to honor their support and mentorship by paying it forward to future students and colleagues. I would like to express my sincere appreciation to Dr. Javier A. Jo for his kindness, generosity, and support during my doctoral studies. His encouragement and willingness to help during important moments of my academic journey are deeply appreciated. I am grateful for the confidence he placed in me and for the opportunities and support he provided along the way.

I am also sincerely thankful to my lab mates, collaborators, and research partners for their friendship, support, and collaboration throughout my Ph.D. studies. Their contributions to experiments, data collection, scientific discussions, and problem-solving enriched this research and enhanced my growth as a researcher. I greatly value the knowledge, encouragement, and camaraderie they shared, which made this journey both productive and memorable.

Finally, I would like to thank everyone who contributed to my academic, profes-

sional, and personal growth throughout these years. Your support, guidance, and encouragement have played an important role in making this achievement possible.

Table of Contents

Dedications . . . iv

Acknowledgments . . . iv

List of Tables . . . x

List of Figures . . . xii

Abstract . . . xvi

1 Introduction . . . 1
- 1.1 Clinical Background and Motivation . . . 1
- 1.2 Clinical Endpoints for Treatment Response Evaluation . . . 4
- 1.3 Biomarkers for Precision Oncology . . . 6
  - 1.3.1 Definition and Classification of Biomarkers . . . 7
  - 1.3.2 Established Biomarkers in Ovarian Cancer . . . 8
  - 1.3.3 Limitations of Current Biomarkers . . . 8
- 1.4 Medical Imaging and Imaging Biomarkers . . . 9
  - 1.4.1 Role of Medical Imaging in Cancer Research . . . 10
  - 1.4.2 Biological Basis of Imaging Biomarkers . . . 11
  - 1.4.3 Imaging Biomarkers in Oncology . . . 12
  - 1.4.4 Computed Tomography and Radiomics . . . 13
- 1.5 Machine Learning for Imaging Biomarker Discovery and Treatment Response Prediction . . . 15
- 1.6 Deep Learning and Foundation Models for Imaging Biomarker Discovery 19
  - 1.6.1 Deep Learning for Medical Image Analysis . . . 19
  - 1.6.2 Transformer-Based Representation Learning . . . 20
  - 1.6.3 Foundation Models . . . 21
  - 1.6.4 Toward Multimodal Foundation Models . . . 22
- 1.7 Research Gap and Dissertation Contributions . . . 23
- 1.8 Organization of the Dissertation . . . 25

2 Research Objective and Hypothesis . . . 29
- 2.1 Advances Toward Robust Imaging Biomarkers . . . 29

2.1.1 Feature robustness and data standardization 29
2.1.2 Transition from handcrafted to learned features 31
2.2 Multimodal Data Integration 33

3 Comparing the Effectiveness of 2D and 3D Features on Predicting the Response to Chemotherapy for Ovarian Cancer Patients 36
3.1 Introduction 36
3.2 Materials and Methods 37
3.3 Results 39
3.4 Discussion and Conclusion 40

4 Evaluating the Effectiveness of 2D and 3D CT Image Features for Predicting Tumor Response to Chemotherapy 44
4.1 Introduction 44
4.2 Materials and Methods 46
4.2.1 Database 46
4.2.2 Radiomics Feature Extraction 47
4.2.3 Developing Machine Learning-Based Models to Predict Tumor Response to chemotherapy 50
4.3 Results 52
4.4 Discussion 54

5 A vision transformer-radiomics approach for enhanced chemotherapy outcome prediction in ovarian cancer 60
5.1 Introduction 60
5.2 Materials and Methods 63
5.2.1 Database 63
5.2.2 Image Feature Computation 64
5.2.3 Machine Learning Model Development 68
5.3 Results 72
5.4 Discussion 78

6 Multimodal Representation Fusion for Chemotherapy Response Prediction in Ovarian Cancer 82
6.1 Introduction 82
6.2 Materials and Methods 85
6.2.1 Dataset 85
6.2.2 WSI Preprocessing and Feature Encoding 87
6.2.3 CT Feature Encoding 89
6.2.4 Feature Selection 89
6.2.5 Multimodal Fusion Strategies and Model Development 90
6.2.6 Model Evaluation and Statistical Analysis 94
6.3 Results 95

6.3.1 Correlation Analysis of Imaging Features . . . 95
6.3.2 Model Complexity . . . 96
6.3.3 Predictive Performance . . . 96
6.3.4 Receiver Operating Characteristic and Precision–Recall Analysis 98
6.3.5 Confusion Matrix Analysis . . . 98
6.3.6 Feature Stability Analysis . . . 98
6.3.7 Relative Contribution of CT and WSI Features . . . 99
6.4 Discussion . . . 99

7 Conclusions and Future Work . . . 108
7.1 Conclusions . . . 108
7.1.1 Journal Articles . . . 111
7.1.2 Conference Proceedings . . . 113
7.2 Future Studies . . . 114
7.2.1 Expanding Multimodal Biomarkers . . . 115
7.2.2 Improving Generalizability and Clinical Robustness . . . 116
7.2.3 Reasoning-Based Artificial Intelligence for Clinical Decision Support . . . 117
7.2.4 Translation Toward Precision Oncology . . . 119

Appendix A: Radiomics Features . . . 120
A.1 First-order Features (Density) . . . 121
A.2 Shape Features (2D) . . . 121
A.3 Shape Features (3D) . . . 122
A.4 Gray Level Co-occurrence Matrix Features (GLCM) . . . 122
A.5 Gray Level Run Length Matrix Features (GLRLM) . . . 123
A.6 Gray Level Size Zone Matrix Features (GLSZM) . . . 124
A.7 Neighboring Gray Tone Difference Matrix Features (NGTDM) . . . 125
A.8 Gray Level Dependence Matrix Features (GLDM) . . . 125

Appendix B: Supplementary Results for Integrated Radiomics and Deep Features . . . 126

Reference List . . . 128

List of Tables

3.1 Summary of the optimal feature clusters 40

3.2 Performance comparison of the 2D and 3D feature-based prediction models 41

3.3 Confusion matrix for 2D and 3D feature-based prediction models 41

4.1 Patient characteristics in the study cohort 47

4.2 Number of radiomics features extracted from the original and filtered images 48

4.3 Summary of selected features by radiomics category and filtering method 56

4.4 Performance comparison of prediction models 56

5.1 Demographic and clinical characteristics of the study cohort. 64

5.2 Prediction performance of the models with confidence interval. 76

6.1 Model complexity comparison for single-modality and fusion models. . 96

6.2 Prediction performance of the models with confidence intervals. 97

6.3 Confusion matrices for single-modality and fusion models. 99

A.1 Summary of extracted radiomics feature classes. 120

B.1 Performance metrics of the models trained with different feature sets. . 126

B.2 No-SMOTE ablation results for the best-performing integrated model (Radiomics + ViT + MedSAM) using the same stratified nested cross-validation protocol. 126

B.3 Features selected by the best-performing integrated model (Radiomics + ViT + MedSAM). Features within each source are ordered according to decreasing mean absolute SHAP value. 127

List of Figures

1.1 Overview of the standard clinical management of ovarian cancer. The treatment pathway consists of diagnosis and staging, cytoreductive surgery, platinum-based chemotherapy, and subsequent assessment of treatment response. 2

1.2 RECIST 1.1 criteria scheme. 6

1.3 Taxonomy of biomarkers and representative examples used for ovarian cancer outcome prediction. 10

1.4 General radiomics workflow for imaging biomarker discovery. Following medical image acquisition, the tumor is segmented and represented by a large set of handcrafted quantitative features describing tumor morphology, intensity, texture, and higher-order image characteristics. These features are subsequently analyzed using statistical and machine learning methods to identify imaging biomarkers associated with treatment response, prognosis, and other clinical outcomes. 16

1.5 Machine learning development and application models for treatment response classification tasks. 17

1.6 Graphical overview of the dissertation 28

3.1 An illustration of the tumor segmentation process. (a) Identified tumors on the CT slices. (b) Tumors outlined using our developed CAD software. (c) Segmented tumor masks. 40

3.2 The heatmap of the Pearson correlation coefficients for (a) 2D features and (b) 3D features, and the corresponding histograms for (c) 2D features and (d) 3D features. 41

3.3 ROC curves generated by the 2D and 3D feature-based prediction models 42

4.1 Overview of the feature extraction process. 49

4.2 Representative tumor images from (a) a responder patient and (b) a non-responder patient. 52

4.3 Example of the tumor segmentation process. The central slice is outlined in red. (a) Identified tumor on the CT slices. (b) Tumors outlined using the developed CAD software. (c) Segmented tumor masks. 53

4.4 Feature heatmaps of (a) 2D tumors, (b) incomplete 3D tumors, and (c) 3D tumors. The corresponding histograms are shown in (d)–(f) for 2D, incomplete 3D, and 3D features, respectively. 55

4.5 Performance comparison of the models for six-month progression-free survival prediction. (a) ROC curves of the prediction models. (b) Corresponding AUC values with 95% confidence intervals. 55

5.1 Flow diagram illustrating patient selection and cohort construction. . . 64

5.2 Vision Transformer-based feature extraction workflow. (a) CT image with highlighted tumor. (b) Tumor region of interest. (c) Schematic representation of ViT-based feature extraction. 68

5.3 Overview of the proposed framework for chemotherapy response prediction. (a) Feature extraction pipeline illustrating the computation of handcrafted radiomics features, vision transformer embeddings, and MedSAM embeddings from segmented tumor regions. (b) Machine learning workflow including feature preprocessing, correlation filtering, SMOTE oversampling, LASSO-based feature selection, support vector machine (SVM) classification, and nested cross-validation evaluation. 69

5.4 Sample CT images illustrate tumor segmentation. (a) Four images from the responder cohort, and (b) four images from the non-responder cohort. 72

5.5 (a) Heatmap of Pearson correlation coefficients among all extracted features from radiomics, ViT-based, and MedSAM-based groups. (b–d) Individual heatmaps showing intra-group correlations. 73

5.6 Normalized distributions of mutual information (MI) between feature pairs used in training multi-modal models. An MI threshold of 0.1 indicates minimal statistical dependency. (a) Radiomics and ViT features, (b) ViT and MedSAM features, and (c) Radiomics and MedSAM features. 75

5.7 Radar chart comparing the performance of models trained on different feature sets across multiple evaluation metrics, including accuracy, precision, sensitivity, F1-score, and specificity. 77

5.8 Calibration curve of the best-performing integrated model. The close agreement between predicted and observed probabilities indicates good calibration performance. 78

5.9 Performance of the six-month PFS prediction model achieved by the best performing classification model based on Radiomics, ViT and MedSAM feature integration. (a) ROC curve. (b) Precision-recall curve. (c) Confusion matrix. 79

5.10 SHAP summary plots for the features selected in the best-performing model (Radiomics + ViT + MedSAM), shown separately by source: (a) MedSAM, (b) ViT, and (c) radiomics. To enhance visual clarity, feature names are not shown in the plots and are instead listed in Supplementary table B.3 in the chapter B. 80

6.1 Overview of the tile filtration workflow for cancerous hotspot selection. 88

6.2 Structural block diagrams of the evaluated models: (a) CT-only baseline, (b) WSI-only baseline, (c) Early fusion, (d) Late fusion, (e) Feature attention fusion, and (f) Cross-modal attention fusion. 104

6.3 Intra- and inter-modality feature correlations across the CT and WSI feature sets. 105

6.4 Heatmap of model performance metrics. 106

6.5 Model performance curves. (a) Receiver operating characteristic (ROC) curves and (b) precision–recall (PR) curves for the single-modality and multimodal fusion models. 106

6.6 Stable predictive features for treatment response. Key features selected across cross-validation folds in the Feature Attention model, highlighting robust biomarkers across fusion strategies. 107

6.7 Relative modality contributions across fusion models. Distribution of learned fusion weights assigned to CT and WSI features. 107

B.1 Contribution of the top 20 optimal features to the best-performing classifier, as quantified by SHAP values. 127

Abstract

Selecting an effective treatment regimen is central to cancer management. Although clinicians consider numerous patient- and disease-specific factors, many biologically relevant characteristics remain inaccessible through conventional clinical assessment, leading to suboptimal treatment decisions and variable patient outcomes. Medical imaging offers a noninvasive means of capturing tumor phenotype and heterogeneity, enabling the development of quantitative imaging biomarkers that may improve personalized treatment selection. This dissertation investigates imaging biomarkers derived from computed tomography (CT) and digital pathology, spanning handcrafted radiomics, deep learning, and foundation model representations, for predicting six-month progression-free survival (PFS) following platinum-based chemotherapy in ovarian cancer.

The first study investigated whether incorporating volumetric information improves radiomics-based prediction by comparing handcrafted features extracted from a single two-dimensional (2D) tumor slice with those extracted from the full three-dimensional (3D) tumor volume. More than 1,200 radiomics features were extracted and evaluated within a machine learning framework incorporating feature selection and support vector machine classification. While 3D radiomics demonstrated improved predictive performance compared with conventional 2D analysis, the results also indicated that incorporating the entire tumor volume introduced additional feature redundancy and variability.

Building upon these findings, the second study proposed an incomplete-3D radiomics strategy designed to balance volumetric information with feature stability. Rather than analyzing the entire tumor volume, radiomics features were extracted from

the central tumor slice together with a limited number of adjacent slices. This compact volumetric representation consistently outperformed both conventional 2D and full 3D radiomics. Feature redundancy analysis further demonstrated that the most discriminative imaging biomarkers originated from the central tumor region, suggesting that limited volumetric context captures the majority of clinically relevant information while reducing redundant image characteristics.

The third study investigated complementary CT imaging biomarkers derived from both handcrafted radiomics and foundation models. Handcrafted radiomics features were combined with deep representations generated by Vision Transformer (ViT) and MedSAM models from metastatic tumor CT images. A nested cross-validation framework with in-fold feature selection and model optimization was employed to compare individual and fused imaging biomarkers. The results demonstrated that foundation model representations provided highly informative imaging biomarkers, while their integration with handcrafted radiomics produced robust and interpretable predictive models. Model interpretation further revealed that MedSAM-derived features contributed most strongly to treatment response prediction, whereas selected radiomics descriptors captured complementary measures of intratumoral heterogeneity.

Recognizing that chemotherapy response is influenced by both the intrinsic biological characteristics of the primary tumor and the phenotypic behavior of metastatic disease, the fourth study developed a multimodal imaging framework integrating digital pathology and radiology. Histopathological representations were extracted from primary tumor whole-slide images using the CONCH foundation model, while complementary imaging representations were obtained from metastatic CT lesions using MedSAM. Multiple multimodal fusion strategies, including early fusion, late fusion, and attention-based fusion, were investigated to learn complementary biomarkers spanning microscopic tissue architecture and macroscopic metastatic phenotypes. By integrat-

ing imaging biomarkers from both disease compartments, this biologically motivated framework captures complementary aspects of ovarian cancer biology that cannot be characterized by either modality alone and demonstrates the potential of multimodal foundation models for improving treatment response prediction.

Overall, this dissertation demonstrates the evolution of imaging biomarkers from handcrafted radiomics to multimodal foundation model representations for treatment response prediction in ovarian cancer. The findings show that compact volumetric radiomics improves robustness over conventional 3D analysis, foundation model-derived imaging representations provide powerful predictive biomarkers, and integrating pathology-derived information from the primary tumor with radiology-derived information from metastatic lesions offers a promising strategy for capturing complementary aspects of tumor biology associated with chemotherapy response. Together, these studies establish a scalable framework for multimodal imaging biomarker discovery and lay the foundation for future integration with genomic, molecular, and clinical information to advance precision oncology.

Chapter 1

Introduction

1.1 Clinical Background and Motivation

Ovarian cancer is the deadliest gynecological malignancy and remains one of the leading causes of cancer-related mortality among women worldwide [1]. In the United States alone, an estimated 20,890 new cases and 12,730 deaths are expected in 2025 [2]. The high mortality rate is largely attributed to the absence of effective screening strategies and the nonspecific nature of early symptoms, including abdominal pain, bloating, urinary urgency, and urinary frequency. Current screening approaches, including bimanual pelvic examination, serum cancer antigen 125 (CA125), and transvaginal ultrasound, lack sufficient sensitivity and specificity for reliable early detection [3]. Consequently, approximately 80% of patients are diagnosed with advanced-stage disease, at which point the cancer has frequently metastasized throughout the peritoneal cavity, substantially limiting therapeutic options and worsening prognosis [1].

Ovarian cancer is not a single disease but rather a heterogeneous group of malignancies with distinct molecular, histopathological, and clinical characteristics. The most common form, epithelial ovarian cancer, comprises five major histological subtypes that differ in their genomic alterations, biological behavior, and response to therapy [4]. This heterogeneity contributes to the considerable variability observed in treatment outcomes, even among patients with similar clinical presentations, and

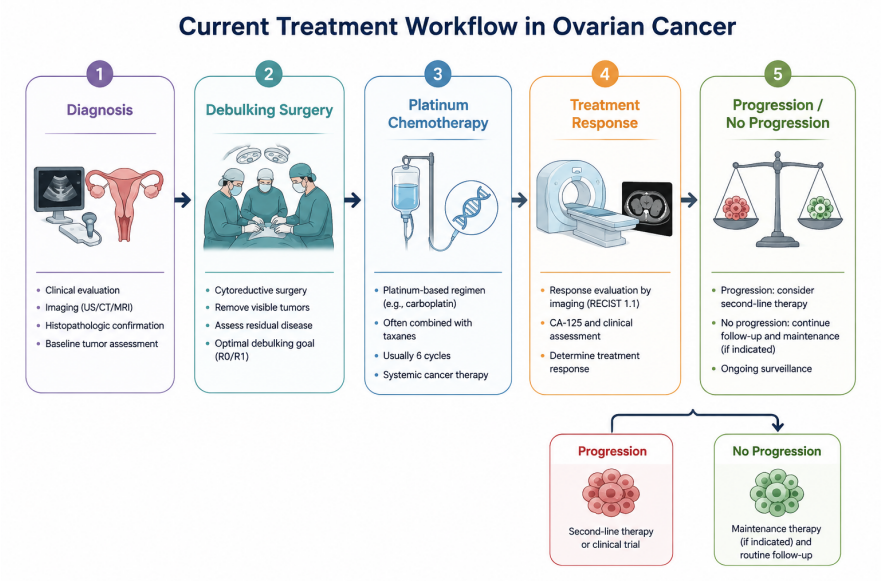


Figure 1.1: Overview of the standard clinical management of ovarian cancer. The treatment pathway consists of diagnosis and staging, cytoreductive surgery, platinum-based chemotherapy, and subsequent assessment of treatment response.

represents one of the major challenges in precision oncology.

Treatment strategies are primarily determined according to the International Federation of Gynecology and Obstetrics (FIGO) staging system and typically involve a multimodal approach consisting of maximal cytoreductive surgery (debulking) followed by platinum-based chemotherapy [3], as shown in fig. 1.1. Additional clinical factors, including residual tumor burden, histological subtype, tumor grade, patient performance status, age, and comorbidities, are also considered when selecting the optimal therapeutic regimen. Despite advances in surgical techniques and systemic therapies, treatment response remains highly variable across patients. While many individuals initially respond well to platinum-based chemotherapy, a substantial proportion eventually develop platinum-resistant disease or experience early recurrence, resulting in poor long-term survival.

The ability to identify patients who are unlikely to benefit from standard chemotherapy before or shortly after treatment initiation has therefore become a major objective

in ovarian cancer research. Inappropriate treatment selection exposes patients to unnecessary toxicity, delays the initiation of potentially more effective therapies, increases healthcare costs, and reduces opportunities for personalized treatment planning [5]. For example, the overall five-year survival rate for ovarian cancer remains approximately 47%, largely because of late-stage diagnosis and the development of platinum resistance in approximately 20–30% of patients [6]. Accurate early prediction of treatment response would enable clinicians to identify patients at high risk of treatment failure, facilitating earlier therapeutic intervention and individualized treatment strategies. Beyond its direct clinical impact, reliable treatment response prediction can also improve the efficiency of clinical trials by enriching patient cohorts most likely to benefit from investigational therapies while reducing unnecessary treatment costs. Recent studies have suggested that machine learning-assisted treatment selection may reduce healthcare expenditures by improving therapeutic decision-making and optimizing resource utilizationc̃itehormaty2025a.

These challenges highlight the growing need for robust biomarkers capable of characterizing tumor biology and predicting treatment response before disease progression becomes clinically apparent. Such biomarkers constitute a fundamental component of precision oncology by enabling patient stratification, individualized therapeutic decision-making, and more effective monitoring of treatment efficacy. The following section introduces the clinical endpoints commonly used to evaluate treatment response in ovarian cancer and provides the rationale for selecting six-month progression-free survival as the primary endpoint throughout this dissertation.

1.2 Clinical Endpoints for Treatment Response Evaluation

Evaluating treatment efficacy requires clinically meaningful outcome measures that accurately reflect disease status and patient prognosis. Depending on the clinical objective, outcomes may be represented as continuous variables, binary variables, or time-to-event endpoints. For example, postoperative mortality may be evaluated as a binary outcome at a predefined time point (e.g., 30 or 60 days after surgery) or as a survival outcome using time-to-event analysis. The selection of an appropriate endpoint is therefore critical because it directly influences both clinical interpretation and predictive model development.

Several clinical endpoints have been established for evaluating therapeutic efficacy in ovarian cancer, including response rate (RR), time to progression (TTP), progression-free survival (PFS), and overall survival (OS) [7]. Overall survival, defined as the time from treatment initiation until death from any cause, remains the gold standard endpoint because it directly reflects patient benefit. However, its use often requires prolonged follow-up and may be confounded by subsequent lines of therapy administered after disease progression.

For these reasons, progression-free survival has become one of the primary endpoints in randomized clinical trials involving patients with solid tumors. PFS is defined as the time from treatment initiation until the first documented disease progression or death from any cause, whichever occurs first. Patients who remain alive without documented progression at the time of analysis are considered censored observations [8]. Compared with overall survival, PFS can be assessed earlier during follow-up and more directly reflects the effectiveness of the initial therapeutic regimen, making it particularly valu-

able for evaluating treatment efficacy and accelerating clinical trials.

Although PFS is inherently a time-to-event endpoint, it can also be transformed into a binary outcome by evaluating progression status at a predefined landmark time. One widely adopted endpoint is 6-month progression-free survival (6-month PFS), which classifies patients according to whether disease progression or death occurs within six months following treatment initiation. This endpoint provides an early assessment of treatment efficacy while maintaining strong clinical relevance. In ovarian cancer, six-month PFS has been extensively used in both clinical research and pharmaceutical drug development as an early indicator of therapeutic benefit, enabling more rapid identification of patients likely to respond or develop resistance to treatment.

Determining progression-free survival requires standardized and reproducible criteria for evaluating changes in tumor burden. In clinical oncology, treatment response is primarily assessed by measuring changes in tumor size before, during, and after therapy. The Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 provides an internationally accepted framework for objectively assessing tumor response across clinical trials and routine patient management [9]. RECIST is independent of treatment modality and is applicable to chemotherapy, targeted therapy, immunotherapy, and selected local treatments.

According to the RECIST 1.1 guidelines, up to five measurable target lesions (with a maximum of two lesions per organ) are identified at baseline. Eligible non-nodal lesions must have a longest diameter of at least 10 mm, whereas pathological lymph nodes must have a short-axis diameter of at least 15 mm. The sum of diameters (SOD) of all target lesions serves as the quantitative reference for longitudinal assessment. During follow-up, changes in the SOD are used to classify treatment response as complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD). While treatment response (CR or PR) is evaluated relative to the baseline SOD, disease

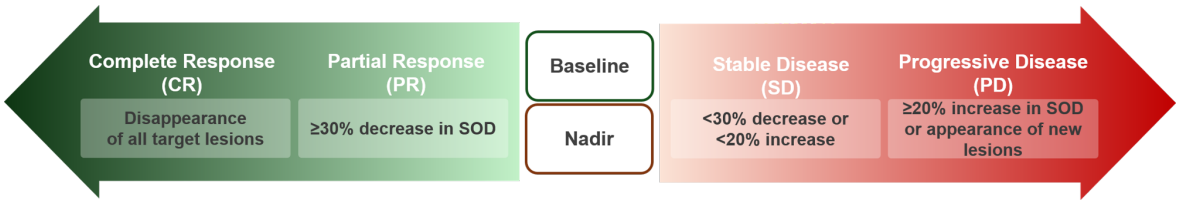


Figure 1.2: RECIST 1.1 criteria scheme.

progression is determined relative to the smallest previously recorded tumor burden (the nadir), as illustrated in fig. 1.2.

The widespread adoption of RECIST has substantially improved the reproducibility and comparability of oncology clinical trials by providing objective criteria for measuring tumor burden and treatment response. Because RECIST-based progression forms the basis for calculating progression-free survival, it serves as the clinical foundation for the treatment response prediction models developed throughout this dissertation. Consequently, six-month progression-free survival was selected as the primary endpoint for all studies presented in this work, providing a clinically meaningful and standardized measure for evaluating chemotherapy response in ovarian cancer.

1.3 Biomarkers for Precision Oncology

The substantial variability in treatment response among patients with ovarian cancer has driven increasing interest in biomarkers that can support personalized therapeutic decision-making. Rather than relying solely on clinical observations or anatomical measurements, biomarkers provide objective information regarding the biological characteristics of both the patient and the tumor. By identifying patients who are likely to respond or develop resistance to a particular therapy, biomarkers enable more individualized treatment selection, improve risk stratification, and facilitate the development of precision oncology.

1.3.1 Definition and Classification of Biomarkers

According to the U.S. Food and Drug Administration (FDA) and the National Institutes of Health (NIH), a biomarker is defined as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or responses to therapeutic interventions. Biomarkers have become indispensable throughout the cancer care continuum, contributing to disease screening, diagnosis, prognosis, treatment selection, response assessment, and surveillance following therapy.

Cancer biomarkers can be broadly classified into several categories depending on the biological information they represent. Diagnostic biomarkers assist in detecting disease, prognostic biomarkers estimate the natural course of the disease independent of treatment, predictive biomarkers identify patients most likely to benefit from a specific therapy, monitoring biomarkers evaluate treatment effectiveness over time, while pharmacodynamic biomarkers quantify biological responses following therapeutic intervention. Among these categories, predictive biomarkers are particularly important for treatment response prediction because they directly support individualized therapeutic decision-making.

Biomarkers may originate from multiple sources, including blood, tissue specimens, molecular assays, genomic sequencing, proteomic analysis, metabolomics, and medical imaging. Each biomarker modality provides complementary information regarding tumor biology, reflecting different spatial and biological scales of disease. Consequently, integrating multiple biomarker sources has become an increasingly important objective in precision oncology.

1.3.2 Established Biomarkers in Ovarian Cancer

Several biomarkers have already been incorporated into the clinical management of ovarian cancer. Serum cancer antigen 125 (CA125) remains the most widely used circulating biomarker for monitoring disease burden and evaluating recurrence, although its limited specificity restricts its utility for population screening [3]. Human epididymis protein 4 (HE4) has been introduced to improve diagnostic accuracy and is frequently combined with CA125 in clinical risk assessment models. At the molecular level, germline and somatic mutations in the *BRCA1* and *BRCA2* genes, together with homologous recombination deficiency (HRD), have become important predictive biomarkers for identifying patients who may benefit from poly(ADP-ribose) polymerase (PARP) inhibitor therapy. Additional genomic and molecular alterations continue to be investigated as potential indicators of prognosis and therapeutic response. fig. 1.3 presents an overview of biomarker taxonomy for ovarian cancer outcome prediction. The figure categorizes biomarkers into major classes and provides representative examples of each category used to predict patient outcomes.

Collectively, these biomarkers have substantially advanced precision oncology by enabling patient stratification and targeted therapeutic interventions. Nevertheless, they characterize only selected biological aspects of the disease and therefore provide an incomplete representation of the complex spatial and temporal heterogeneity exhibited by ovarian tumors.

1.3.3 Limitations of Current Biomarkers

Despite their clinical value, currently available biomarkers present several limitations that restrict their ability to accurately predict chemotherapy response. Many molecular and genomic biomarkers require invasive tissue sampling obtained through biopsy or

surgical resection, limiting opportunities for repeated longitudinal assessment throughout treatment. Furthermore, biopsy specimens represent only a small portion of the tumor and may therefore fail to capture the substantial intratumoral and intertumoral heterogeneity that characterizes advanced ovarian cancer. Tumors continuously evolve under therapeutic pressure, and a single tissue sample obtained at one time point may not adequately reflect the dynamic biological changes associated with treatment resistance and disease progression.

Circulating biomarkers such as CA125 provide valuable information regarding overall disease burden but offer limited insight into tumor morphology, spatial heterogeneity, vascularity, or characteristics of the tumor microenvironment. Likewise, molecular biomarkers frequently describe only one biological pathway, whereas treatment response is influenced by numerous interacting factors, including tumor cellularity, angiogenesis, hypoxia, immune infiltration, stromal composition, and metastatic burden.

These limitations have stimulated growing interest in non-invasive biomarkers capable of characterizing the entire tumor and its surrounding microenvironment while allowing repeated assessment throughout treatment. Medical imaging represents a unique opportunity to fulfill this role by providing quantitative measurements of tumor phenotype across the whole lesion without requiring destructive tissue sampling. Consequently, imaging biomarkers have emerged as an important complement to conventional molecular biomarkers and have become a major focus of research for predicting treatment response in precision oncology.

1.4 Medical Imaging and Imaging Biomarkers

The limitations of conventional molecular and tissue-based biomarkers have stimulated growing interest in medical imaging as a complementary source of quantitative biomark-

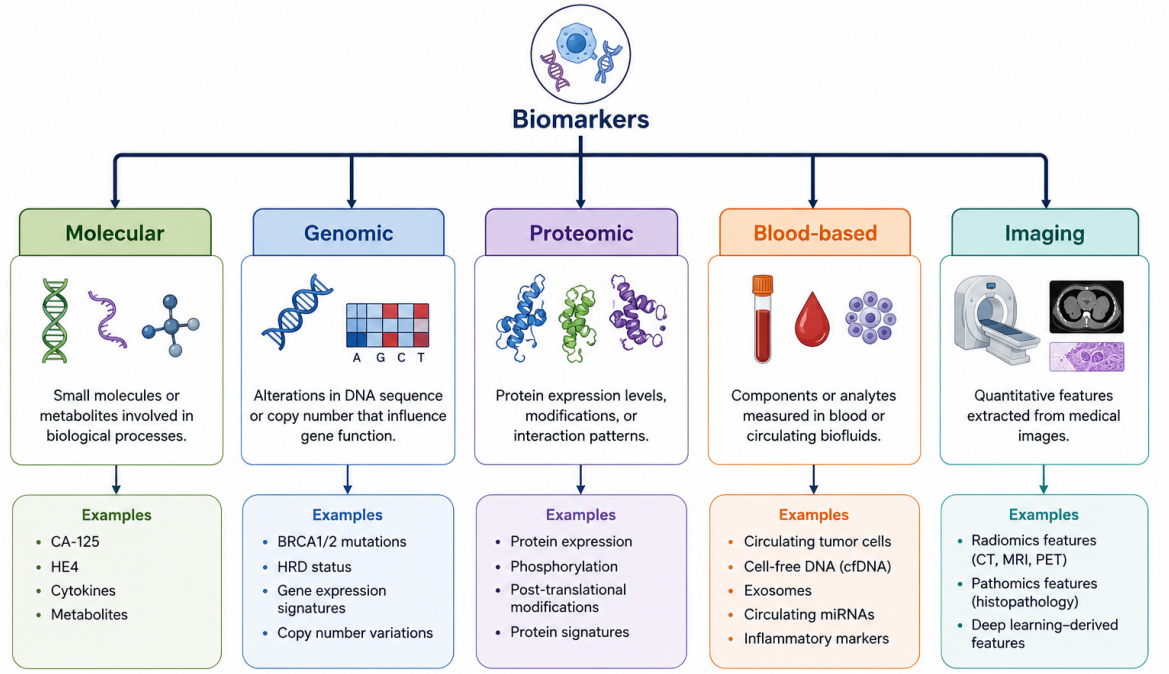


Figure 1.3: Taxonomy of biomarkers and representative examples used for ovarian cancer outcome prediction.

ers. Unlike tissue sampling, which characterizes only a limited portion of the tumor at a single time point, medical imaging provides a non-invasive means of repeatedly evaluating the entire tumor and its surrounding environment throughout the course of treatment. Consequently, imaging has become an indispensable component of modern precision oncology, supporting diagnosis, staging, treatment planning, therapeutic monitoring, and disease surveillance while simultaneously providing rich quantitative information regarding tumor phenotype and biology.

1.4.1 Role of Medical Imaging in Cancer Research

Biomedical imaging enables visualization of anatomical, physiological, and functional characteristics of tissues across multiple spatial scales without destructive sampling [10]. Modern imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasound, and digital

pathology, provide complementary information regarding tumor morphology, vascularity, cellular organization, and microenvironment. These imaging techniques have become integrated throughout the entire cancer care pathway, from early detection and diagnosis to biopsy guidance, disease staging, treatment planning, response evaluation, recurrence monitoring, and clinical research [10].

Each imaging modality provides unique information reflecting different aspects of tumor biology. CT primarily depicts tissue density and anatomical structure, MRI offers excellent soft tissue contrast and functional characterization, PET visualizes metabolic activity, while digital pathology provides microscopic assessment of tissue architecture, cellular morphology, and the tumor microenvironment. Rather than competing technologies, these modalities provide complementary perspectives of the same biological disease process.

Conventional histopathology remains the reference standard for characterizing tumor phenotype and microenvironment because it directly visualizes cellular organization and tissue morphology. However, pathology requires invasive tissue acquisition and is susceptible to sampling bias, variability in tissue preparation, staining procedures, and region-of-interest selection [11]. Furthermore, biopsy specimens often represent only a small fraction of a biologically heterogeneous tumor, limiting their ability to characterize the entire disease burden. These limitations have motivated increasing interest in imaging techniques capable of non-invasively assessing whole-tumor characteristics.

1.4.2 Biological Basis of Imaging Biomarkers

Tumor appearance on medical images is not random. Instead, imaging phenotypes arise from the underlying biological and physiological characteristics of the tumor. Genetic alterations influence cellular proliferation, angiogenesis, stromal remodeling,

immune infiltration, hypoxia, necrosis, and extracellular matrix organization. These biological processes modify tissue composition and architecture, which subsequently alter the physical properties measured by medical imaging modalities. Consequently, quantitative image characteristics can indirectly reflect the underlying molecular and cellular behavior of the tumor.

For example, highly aggressive tumors often exhibit irregular morphology, infiltrative growth patterns, heterogeneous cellular density, extensive necrosis, abnormal vascular networks, and increased intratumoral heterogeneity. These biological characteristics influence measurable image properties such as tumor shape, boundary irregularity, attenuation patterns, enhancement characteristics, and texture complexity. Likewise, hypoxic tumors frequently demonstrate increased necrosis and heterogeneous enhancement, while highly vascular tumors exhibit distinct perfusion characteristics associated with angiogenesis and drug delivery. Numerous studies have demonstrated that these imaging phenotypes are associated with clinically important outcomes including tumor aggressiveness, recurrence, survival, and response to chemotherapy.

Importantly, imaging biomarkers should not necessarily be interpreted as causal determinants of treatment response. Instead, they frequently act as surrogate markers that capture biological processes associated with therapeutic sensitivity or resistance. Even when the exact biological mechanisms remain incompletely understood, quantitative imaging biomarkers may still provide substantial predictive value by identifying complex imaging patterns correlated with clinical outcome.

1.4.3 Imaging Biomarkers in Oncology

An imaging biomarker is a quantitative characteristic extracted from medical images that objectively reflects normal biological processes, pathological changes, or responses to therapeutic intervention. Unlike conventional qualitative image inter-

pretation, imaging biomarkers transform visual information into measurable variables suitable for statistical analysis and predictive modeling.

Imaging biomarkers encompass a wide range of quantitative descriptors, including tumor size, volume, shape, margin irregularity, vascularity, enhancement kinetics, tissue density, texture heterogeneity, and temporal changes observed during treatment. Many of these biomarkers have demonstrated significant associations with important clinical endpoints such as overall survival, progression-free survival, metastatic potential, and treatment response across multiple cancer types.

One of the principal advantages of imaging biomarkers is their ability to characterize the entire tumor together with its surrounding microenvironment while allowing repeated measurements during therapy. Consequently, imaging biomarkers capture both spatial and temporal heterogeneity that cannot be adequately represented by localized tissue biopsy. Furthermore, imaging examinations such as CT are routinely acquired during standard clinical management, enabling quantitative biomarker extraction without additional invasive procedures or increased patient burden.

Recent studies have further demonstrated strong associations between imaging biomarkers and molecular characteristics traditionally measured using tissue-based assays. For example, radiomic features extracted from routine CT images have been shown to predict $CD8^+$ T-cell infiltration, a biomarker typically quantified using immunohistochemical analysis of tumor tissue [12, 13]. These findings suggest that imaging biomarkers can complement conventional molecular biomarkers while providing comprehensive whole-tumor characterization.

1.4.4 Computed Tomography and Radiomics

Computed tomography remains the primary imaging modality for staging, treatment monitoring, and response evaluation in patients with ovarian cancer because of its

rapid acquisition, high spatial resolution, widespread availability, and compatibility with standardized assessment criteria such as RECIST 1.1 [14]. Traditionally, CT has been used primarily for visual assessment of tumor burden and anatomical changes during therapy. However, advances in quantitative image analysis have demonstrated that considerably more biological information is contained within medical images than can be appreciated through visual interpretation alone.

Radiomics seeks to unlock this hidden information by converting standard medical images into large numbers of quantitative features describing tumor morphology, intensity distributions, texture, and higher-order spatial patterns [15]. These handcrafted imaging descriptors provide objective measurements of tumor heterogeneity and phenotype that may reflect underlying biological processes associated with treatment response, disease progression, and patient survival.

Unlike conventional pathological assessment, which samples only selected tissue regions, radiomics analyzes the entire imaged tumor volume, thereby providing a more comprehensive representation of intratumoral heterogeneity. Furthermore, radiomic features have demonstrated associations with genomic, transcriptomic, proteomic, and immunological characteristics, giving rise to the emerging field of radiogenomicsc̃itegillies2016a,yip2016a. Consequently, radiomics has become one of the most widely investigated approaches for discovering imaging biomarkers capable of supporting precision oncology through non-invasive prediction of prognosis and therapeutic response. The general radiomics workflow is illustrated in fig. 1.4.

Although radiomics has achieved encouraging results across numerous oncology applications, handcrafted feature engineering remains dependent on image quality, segmentation accuracy, and manually designed mathematical descriptors. These limitations have motivated the development of increasingly sophisticated machine learning and deep learning approaches capable of learning more discriminative imaging repre-

sentations directly from medical images. The following section introduces these computational methodologies and discusses their role in imaging biomarker discovery.

1.5 Machine Learning for Imaging Biomarker Discovery and Treatment Response Prediction

The quantitative characterization of tumors through radiomics results in hundreds or even thousands of imaging-derived features describing tumor morphology, intensity, texture, and higher-order spatial relationships. While these high-dimensional feature sets provide rich information regarding tumor phenotype, they also introduce substantial analytical challenges. Many radiomic features exhibit complex nonlinear relationships, redundancy, and strong inter-feature correlations that cannot be adequately modeled using conventional statistical techniques. Consequently, advanced computational methods are required to identify the imaging biomarkers most strongly associated with clinically relevant outcomes.

Machine learning (ML) has emerged as one of the principal computational frameworks for discovering predictive imaging biomarkers from high-dimensional medical imaging data. Unlike traditional statistical models, which often assume linear relationships and require predefined mathematical formulations, ML algorithms learn complex decision boundaries directly from data. By identifying nonlinear interactions among numerous imaging features, ML enables the construction of predictive models capable of estimating treatment response, disease recurrence, and patient survival with greater flexibility than conventional regression-based approaches [16, 17].

A typical machine learning pipeline for imaging biomarker discovery consists of several sequential stages, including image preprocessing, tumor segmentation, feature extraction, feature selection, model training, validation, and performance evaluation,

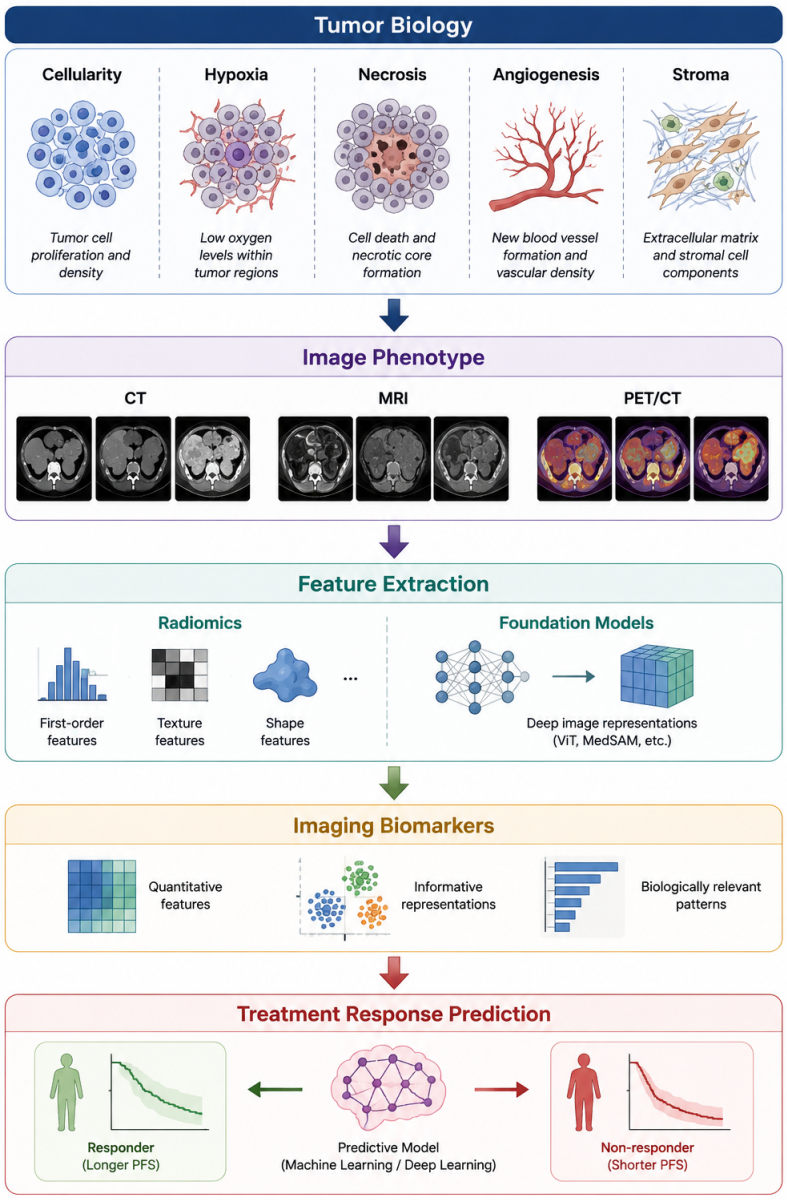


Figure 1.4: General radiomics workflow for imaging biomarker discovery. Following medical image acquisition, the tumor is segmented and represented by a large set of handcrafted quantitative features describing tumor morphology, intensity, texture, and higher-order image characteristics. These features are subsequently analyzed using statistical and machine learning methods to identify imaging biomarkers associated with treatment response, prognosis, and other clinical outcomes.

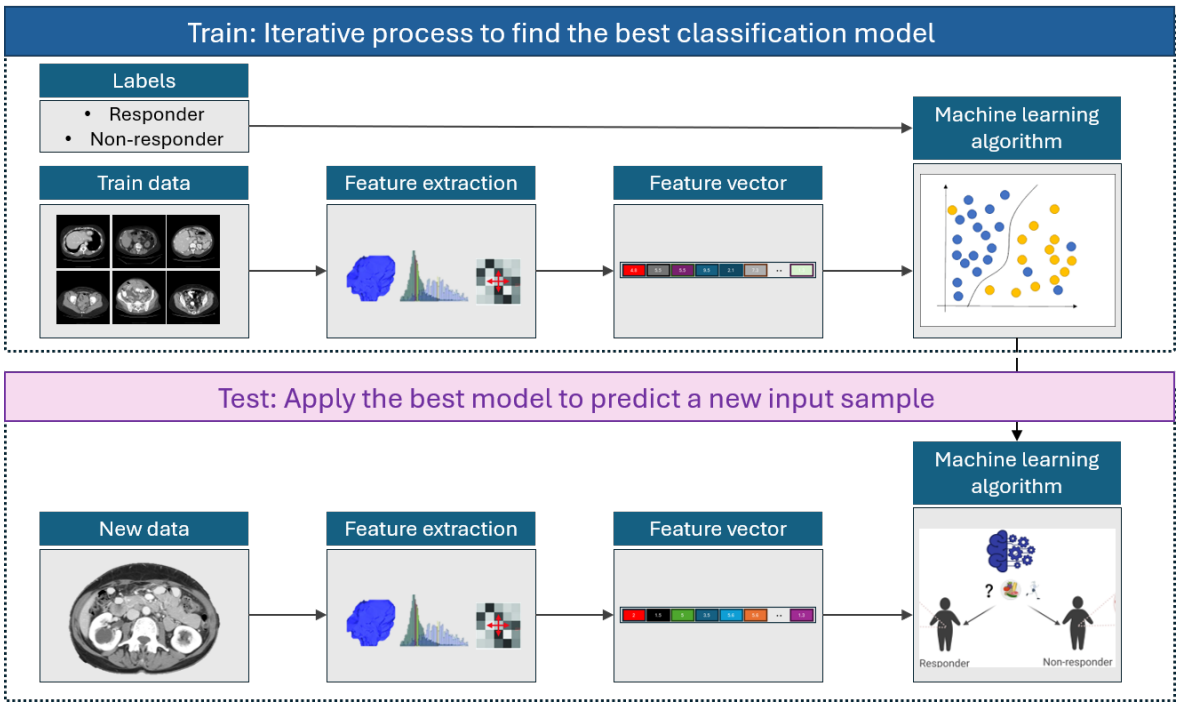


Figure 1.5: Machine learning development and application models for treatment response classification tasks.

as illustrated in fig. 1.5. During feature selection, the objective is not only to improve predictive accuracy but also to identify robust biomarkers that consistently contribute to clinical outcome prediction. Because many radiomic features are highly correlated or contain redundant information, feature selection plays a central role in reducing model complexity, improving interpretability, and minimizing overfitting.

Machine learning has been successfully applied to numerous oncological applications, including diagnosis, tumor grading, survival prediction, recurrence estimation, and prediction of response to chemotherapy, radiotherapy, targeted therapy, and immunotherapy [18]. In ovarian cancer, ML models have demonstrated encouraging performance for predicting progression-free survival, identifying platinum resistance, estimating recurrence risk, and discovering quantitative imaging biomarkers associated with treatment response. Nevertheless, compared with other major cancer types, the application of ML in gynecologic oncology remains relatively limited despite its con-

siderable clinical potential [19, 20].

Among imaging-based machine learning approaches, radiomics has become one of the most widely investigated frameworks for biomarker discovery. Handcrafted radiomic features are used as input variables for ML classifiers, enabling quantitative characterization of tumor heterogeneity and phenotype. However, the combination of high-dimensional feature spaces and relatively small clinical datasets presents significant methodological challenges. Medical imaging studies frequently contain hundreds or thousands of candidate predictors but only a limited number of patients, resulting in a high risk of overfitting and poor model generalizability. In addition, radiomic features are sensitive to variations in image acquisition, reconstruction protocols, segmentation accuracy, and feature extraction settings, potentially affecting reproducibility across institutions.

To address these challenges, robust machine learning pipelines incorporate standardized preprocessing, feature normalization, correlation analysis, dimensionality reduction, feature selection, hyperparameter optimization, and rigorous internal or external validation. Careful experimental design is particularly important in medical imaging studies because optimistic performance estimates may arise from information leakage, inappropriate validation strategies, or unstable feature selection procedures. Consequently, the development of reproducible machine learning pipelines has become as important as the choice of classification algorithm itself.

Although machine learning substantially improves the ability to analyze high-dimensional radiomic data, conventional ML frameworks remain dependent on handcrafted feature engineering. The quality of the final predictive model therefore depends on whether manually designed imaging features adequately represent the underlying biological characteristics of the tumor. This limitation has motivated the rapid development of deep learning approaches that automatically learn hierarchical imaging

representations directly from medical images, reducing dependence on handcrafted feature design while enabling increasingly powerful imaging biomarker discovery. The following section reviews these deep learning methodologies and discusses their role in modern precision oncology.

1.6 Deep Learning and Foundation Models for Imaging Biomarker Discovery

Although conventional machine learning has substantially advanced imaging biomarker discovery, its performance remains fundamentally dependent on handcrafted feature engineering. Radiomic features are manually designed mathematical descriptors that quantify predefined image characteristics such as tumor shape, intensity, and texture. While these features have demonstrated significant predictive value across numerous oncology applications, they may not fully capture the complex and hierarchical imaging patterns associated with tumor biology. This limitation has motivated the adoption of deep learning approaches capable of automatically learning discriminative imaging representations directly from medical images.

1.6.1 Deep Learning for Medical Image Analysis

Deep learning (DL) is a subset of machine learning based on artificial neural networks composed of multiple hierarchical processing layers. Unlike conventional ML algorithms, which rely on manually engineered input features, deep learning models automatically learn increasingly abstract feature representations directly from raw image data through end-to-end optimization. Lower network layers typically capture simple image characteristics such as edges and intensity gradients, while deeper layers

progressively encode complex anatomical structures, tissue organization, and high-level semantic information.

Convolutional neural networks (CNNs) have been the dominant deep learning architecture for medical image analysis over the past decade. Through hierarchical convolutional filters, CNNs automatically learn imaging representations without requiring handcrafted feature extraction. CNN-based models have demonstrated remarkable success in tumor detection, segmentation, classification, survival prediction, and treatment response prediction across numerous imaging modalities, including CT, MRI, PET, ultrasound, and digital pathology [21, 22]. Compared with handcrafted radiomics, CNN-derived representations often capture richer and more complex imaging characteristics while reducing dependence on manually designed feature descriptors.

Despite these advantages, conventional CNNs primarily model local image patterns because of their limited receptive fields. Capturing long-range spatial relationships frequently requires increasingly deeper network architectures, resulting in higher computational complexity and greater demand for large annotated datasets.

1.6.2 Transformer-Based Representation Learning

Inspired by the success of self-attention mechanisms in natural language processing, Vision Transformers (ViTs) have recently emerged as an alternative deep learning architecture for computer vision. Rather than processing local image neighborhoods through convolutional filters, transformer models divide an image into a sequence of patches and employ self-attention mechanisms to model relationships among all image regions simultaneously. This global representation learning enables transformers to capture long-range spatial dependencies that may be difficult for conventional CNNs to represent efficiently.

In medical imaging, transformer-based models have demonstrated competitive or

superior performance for numerous tasks, including lesion detection, image segmentation, disease classification, and treatment response prediction. Their ability to integrate information across the entire image is particularly advantageous for characterizing heterogeneous tumors, whose biologically relevant patterns may extend beyond local texture variations. Nevertheless, transformer architectures generally require substantially larger training datasets than CNNs because they contain fewer built-in assumptions regarding image structure.

1.6.3 Foundation Models

Recent advances in artificial intelligence have led to the emergence of foundation models, representing a major shift in medical image representation learning. Rather than training deep neural networks for a single downstream task, foundation models are first pretrained on extremely large and diverse datasets using self-supervised or weakly supervised learning strategies. During pretraining, these models learn highly generalizable visual representations that can subsequently be transferred to a wide variety of medical imaging applications with relatively limited task-specific data.

Foundation models offer several important advantages for biomedical imaging. First, they substantially reduce dependence on large annotated clinical datasets, which are often expensive and time-consuming to obtain. Second, they produce robust and transferable feature representations that generalize across imaging modalities, organs, institutions, and clinical tasks. Finally, the learned embeddings frequently encode complex semantic information beyond handcrafted radiomic descriptors, providing richer imaging biomarkers for downstream predictive modeling.

Several foundation models have recently demonstrated outstanding performance across medical imaging applications. For radiological imaging, models such as MedSAM leverage large-scale medical image pretraining to generate anatomically meaningful

image embeddings that support segmentation and quantitative image representation. Similarly, in computational pathology, vision-language foundation models including CONCH, UNI, and BiomedCLIP have shown remarkable capability for learning biologically informative representations from whole-slide histopathology images without requiring extensive task-specific annotation. These models have significantly advanced computational pathology by enabling robust feature extraction from gigapixel tissue images while improving generalization across multiple downstream applications.

An important characteristic of foundation models is that the learned feature embeddings are not explicitly handcrafted or manually interpretable. Instead, they represent high-dimensional latent imaging biomarkers learned directly from large-scale data. Although the precise biological meaning of individual embedding dimensions may remain unknown, numerous studies have demonstrated that these learned representations capture clinically relevant information associated with tumor phenotype, molecular alterations, prognosis, and treatment response.

1.6.4 Toward Multimodal Foundation Models

Cancer is inherently a multiscale disease involving biological processes that span molecular, cellular, tissue, organ, and systemic levels. Consequently, no single imaging modality can fully characterize all aspects of tumor biology. Radiological imaging provides macroscopic information regarding tumor morphology, spatial heterogeneity, and metastatic burden, whereas digital pathology captures microscopic tissue architecture, cellular organization, and tumor microenvironment. These modalities therefore provide complementary rather than redundant biological information.

Recent developments in foundation models have created new opportunities for integrating imaging biomarkers extracted from multiple modalities. Rather than relying solely on handcrafted feature fusion, multimodal foundation models enable biologi-

cally complementary image representations learned independently from radiological and pathological data to be combined within unified predictive frameworks. Such integration has the potential to provide a more comprehensive characterization of tumor biology than either modality alone.

This paradigm forms the central motivation of the present dissertation. Building upon handcrafted radiomics, conventional machine learning, and foundation-model representation learning, this work investigates increasingly informative imaging biomarkers for chemotherapy response prediction in ovarian cancer. The final stage of this progression explores multimodal integration of pathology-derived representations from primary tumors and radiology-derived representations from metastatic lesions, providing a biologically motivated framework for precision oncology.

1.7 Research Gap and Dissertation Contributions

The preceding sections have demonstrated that accurate prediction of chemotherapy response is essential for advancing precision oncology in ovarian cancer. Although considerable progress has been achieved through molecular biomarkers, radiomics, machine learning, and more recently deep learning, several important challenges remain unresolved.

First, many existing imaging biomarker studies continue to rely exclusively on either two-dimensional (2D) or full three-dimensional (3D) tumor characterization. While 2D analysis may fail to capture the complete spatial heterogeneity of the tumor, full 3D analysis frequently introduces redundant information, increased computational complexity, and greater sensitivity to segmentation variability. Consequently, the optimal balance between volumetric information and predictive robustness remains incompletely understood.

Second, despite the success of handcrafted radiomics, these approaches remain dependent on manually designed mathematical descriptors that may not fully represent the complex biological characteristics associated with treatment response. Recent advances in deep learning and foundation models have demonstrated remarkable capability for learning informative imaging representations directly from medical images. However, relatively few studies have systematically investigated how handcrafted radiomic biomarkers and foundation-model representations complement one another in predicting chemotherapy response in ovarian cancer.

Finally, the majority of existing predictive models rely on a single imaging modality, despite the inherently multiscale nature of cancer biology. Histopathological images characterize the microscopic architecture and cellular organization of the primary tumor, whereas computed tomography provides macroscopic information regarding metastatic disease, tumor burden, and anatomical heterogeneity. Because ovarian cancer treatment response is influenced by biological processes occurring across multiple spatial scales, integrating complementary imaging biomarkers from pathology and radiology represents a promising yet insufficiently explored direction for precision oncology.

Motivated by these challenges, this dissertation investigates the progressive development of increasingly informative imaging biomarkers for predicting six-month progression-free survival following platinum-based chemotherapy in ovarian cancer. Rather than proposing a single predictive model, the dissertation systematically evaluates successive generations of imaging biomarkers and representation learning strategies, beginning with conventional handcrafted radiomics and culminating in multimodal foundation-model representations.

The first study establishes a baseline by comparing conventional two-dimensional and three-dimensional handcrafted radiomic biomarkers extracted from metastatic tumor CT images. Building upon these observations, the second study proposes an

incomplete-three-dimensional radiomics framework designed to identify the optimal volumetric representation that balances predictive performance with feature robustness. The third study extends handcrafted radiomics by integrating quantitative imaging biomarkers extracted using modern foundation models, thereby investigating the complementary predictive value of manually engineered and automatically learned imaging representations. Finally, the fourth study introduces a multimodal imaging biomarker framework that integrates microscopic representations extracted from primary tumor whole-slide histopathology images with macroscopic representations extracted from metastatic CT lesions. Multiple feature fusion strategies are investigated to combine complementary biological information across imaging modalities for treatment response prediction.

Collectively, these studies demonstrate the evolution of imaging biomarkers from handcrafted radiomics to multimodal foundation-model representations. Beyond improving predictive performance, this progression provides a biologically motivated framework for integrating complementary information across multiple spatial scales of tumor biology. The resulting framework contributes to the development of more robust, interpretable, and clinically applicable imaging biomarkers capable of supporting personalized treatment selection and advancing precision oncology for ovarian cancer.

1.8 Organization of the Dissertation

This dissertation is organized into six chapters that collectively investigate the development of increasingly informative imaging biomarkers for predicting chemotherapy response in ovarian cancer. Figure 1.6 provides a graphical overview of the dissertation and summarizes the relationship among the individual studies.

Chapter 2 presents the first study, which investigates handcrafted CT radiomics

for treatment response prediction. Conventional two-dimensional (2D) and three-dimensional (3D) radiomic representations are systematically compared to evaluate the impact of volumetric tumor characterization on predicting six-month progression-free survival following platinum-based chemotherapy.

Chapter 3 builds upon the findings of the previous chapter by introducing an incomplete-three-dimensional radiomics framework. Rather than analyzing either a single image slice or the entire tumor volume, this study investigates compact volumetric representations centered on the largest tumor cross-section to identify the optimal balance between predictive performance, feature robustness, and computational efficiency.

Chapter 4 extends handcrafted radiomics by incorporating imaging biomarkers learned through modern foundation models. Quantitative radiomic features are integrated with deep image representations extracted using vision transformer (ViT) and MedSAM models from metastatic tumor CT images. The complementary predictive value of handcrafted and automatically learned imaging biomarkers is investigated through multiple fusion strategies and model interpretation analyses.

Chapter 5 presents the final and most comprehensive study of this dissertation by developing a multimodal imaging biomarker framework that integrates radiological and pathological foundation-model representations. Histopathological image embeddings extracted from primary tumor whole-slide images are combined with CT-derived representations obtained from metastatic lesions using multiple multimodal fusion strategies. This framework explores the complementary biological information contained within microscopic tissue architecture and macroscopic tumor morphology for improving chemotherapy response prediction.

Finally, Chapter 6 summarizes the principal findings of this dissertation, discusses their clinical significance, outlines the limitations of the current work, and presents

several directions for future research toward clinically translatable multimodal imaging biomarkers for precision oncology.

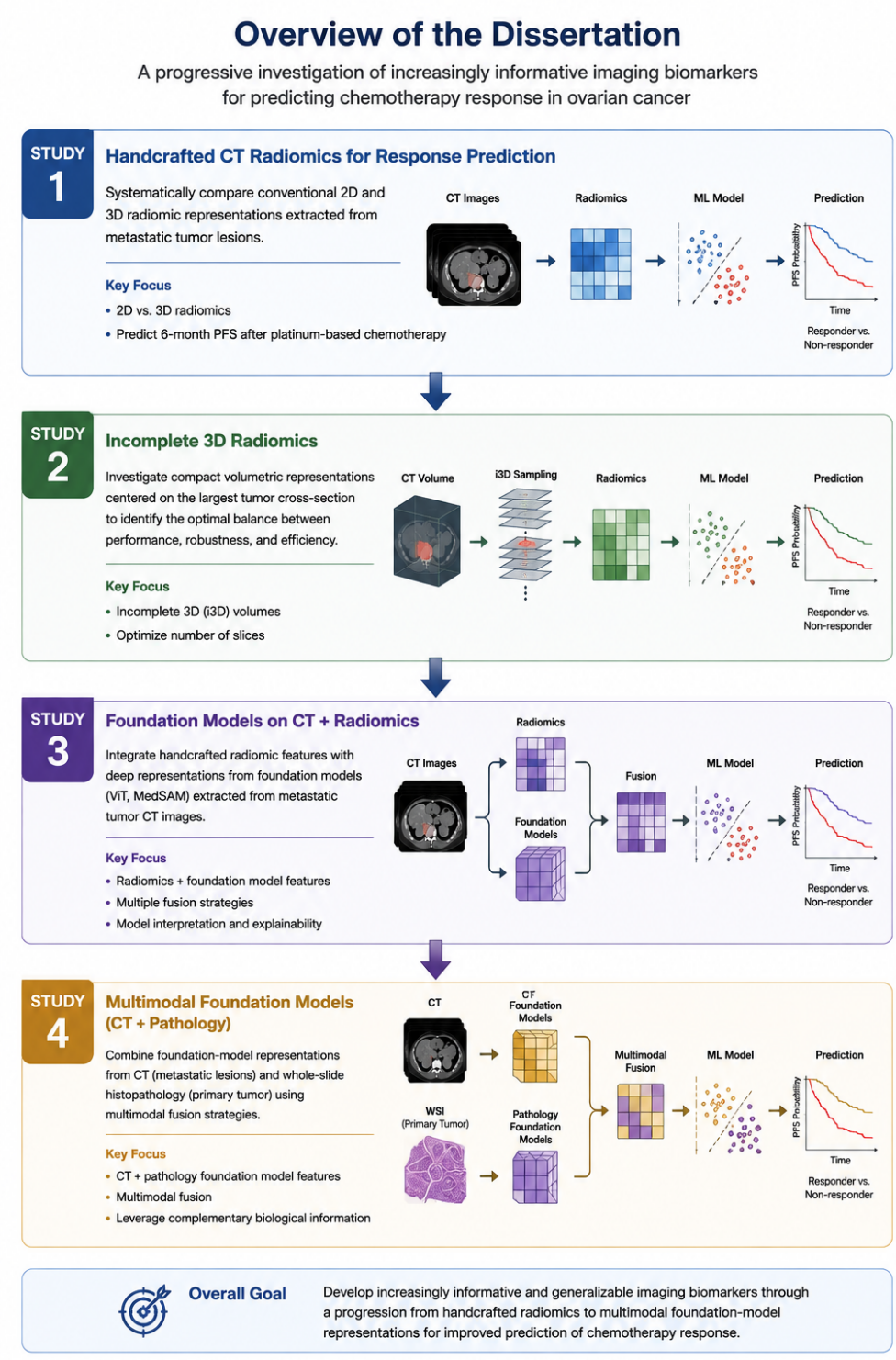


Figure 1.6: Graphical overview of the dissertation

Chapter 2

Research Objective and Hypothesis

2.1 Advances Toward Robust Imaging Biomarkers

2.1.1 Feature robustness and data standardization

Reproducibility is a critical concern in imaging-based treatment response prediction. Many studies rely on limited, single-center datasets, which restrict data diversity and introduce scanner- or protocol-specific biases that affect model generalizability [18, 23]. As a result, predictive models may inadvertently learn site-dependent artifacts rather than true biological signals relevant to therapy outcome. Therefore, to have meaningful biomarkers, it is necessary to have standardization of data, consistent preprocessing, and validation across different institutions [22].

Radiomics-based workflows are sensitive to variations in image acquisition and preprocessing parameters. For example, voxel spacing, resampling, interpolation method, and intensity normalization can significantly change numerical feature values and the stability of downstream models [24]. Tools like *PyRadiomics* [25] have standardized feature extraction workflows, but feature stability is still very reliant on image quality, reconstruction protocols, and segmentation accuracy. Manual segmentation, which is still the clinical gold standard, inherently has inter- and intra-observer variability, especially in tumors with poorly defined margins or heterogeneous texture [26]. Such inconsistencies propagate downstream analysis and degrade model reproducibility.

In our first study, we systematically investigated the impact of feature dimensionality on robustness and predictive power by comparing 2D radiomics derived from a single representative slice with full 3D radiomics computed from the entire tumor volume. We demonstrated that both 2D and 3D radiomic signatures contain biologically meaningful information capable of distinguishing responders from non-responders to chemotherapy. However, we also observed that full 3D radiomics introduces additional variability caused by unclear boundaries of the peripheral slices and the segmentation uncertainty, while 2D features may not be able to capture important volumetric heterogeneity. These results emphasize the importance of balancing biological completeness and feature stability when creating radiomics-based biomarkers.

In our second study, we addressed this challenge by introducing an "incomplete 3D" approach that sequentially incorporated varying numbers of slices around the tumor middle slice. We showed that slice selection plays a pivotal role in maintaining feature robustness. Including peripheral slices, where tumor borders blend into surrounding tissues, it introduces noise that degrades prediction accuracy and reduces signal-to-noise ratio. In contrast, features derived from the central slice and a few adjacent slices were, in fact, the most stable and accurate models. This research proves that the precise adjustment of slice depth can greatly enhance radiomic reproducibility, thus providing an efficient means of lessening segmentation uncertainty and at the same time retaining biologically relevant volumetric information.

The combined findings of these two experiments underscore the importance of preprocessing, standardized segmentation, and optimized slice selection as the key factors for reproducible and generalizable imaging biomarkers. By systematically evaluating how radiomic features behave across different spatial representations (2D, incomplete 3D, full 3D), our work provides a foundation for designing robust imaging-based models that better reflect tumor phenotype and improve treatment outcome prediction.

2.1.2 Transition from handcrafted to learned features

Although radiomics has demonstrated considerable potential as a quantitative imaging strategy, its reliance on predefined, handcrafted descriptors fundamentally limits its representational capacity. Radiomic features encode shape, intensity, and texture through mathematically engineered formulas, but because these descriptors are constrained by human intuition and low-order statistical definitions, they may be unable to capture many of the complex, nonlinear, and multiscale imaging patterns that characterize tumor phenotype and treatment sensitivity [22]. As a result, radiomics may miss hidden or subtle information embedded within CT images. This information could be critical for accurately predicting chemotherapy response in ovarian cancer. Further to these limitations, radiomic analysis are highly dependent on imaging parameters, reconstruction protocols, voxel spacing, and segmentation accuracy, making these features vulnerable to variability and often difficult to generalize across institutions or scanners.

Deep learning offers a fundamentally different paradigm by learning high-level feature representations directly from data [16]. Architectures such as convolutional neural networks, ViT, and recently developed medical foundation models like MedSAM [27] can extract rich, hierarchical embeddings that reflect global spatial context, microenvironmental cues, and complex interactions within the tumor and its surroundings. Compared to handcrafted radiomic features, learned representations are not dependent on human-designed formulas; instead, they capture imaging signatures that arise naturally during training, thereby enabling them to serve as encoders of phenotypic characteristics and heterogeneity patterns which may be totally overlooked by manual feature engineering. Transfer learning further enhances the utility of these models in medical imaging settings where datasets are often limited, allowing pretrained encoders

to contribute strong priors that improve feature quality and downstream predictive performance.

Yet, despite the rapid progress of deep learning, the role of learned features in ovarian cancer chemotherapy-response prediction remains insufficiently understood [19]. Existing studies typically evaluate radiomics or deep models in isolation, and there is limited knowledge regarding how handcrafted descriptors and learned embeddings complement or compete with one another. Particularly unclear is whether transformers pretrained on natural images, such as ViT, contribute meaningfully to medical prediction tasks, and how their learned features compare with domain-adapted models such as MedSAM. Moreover, no systematic studies have formally examined whether integrating radiomic and transformer-based representations can produce a more comprehensive and biologically informative imaging signature.

In our third study, we addressed these gaps by developing a multi-representation framework that combines handcrafted radiomics with high-level embeddings extracted from ViT and MedSAM. This effort builds directly upon the methodological foundation established in our first two studies, where we improved the robustness of radiomic features through systematic evaluation of 2D and 3D feature behavior and through optimized slice selection strategies. After we had a robust and reproducible radiomics pipeline in place, we were then motivated to break the inherent representational limits of handcrafted features by mixing them with the deep features. Deep features learned from the neural network can capture tumor biology at a higher level of abstraction. The main hypothesis of this investigation was that radiomic features alone are insufficient to fully represent the complexity of tumor phenotypes. Therefore, deep features, particularly those coming from medical foundation models like MedSAM, have extra and complementary information that is quite important for the prediction of chemotherapy response. We further hypothesized that combining these heterogeneous feature fami-

lies into a unified representation would yield synergistic improvements, provide a more complete characterization of tumor heterogeneity and translate into superior predictive performance.

Through this integrative approach, the third study advances the broader aim of this dissertation: the hand-crafted radiomics of the past are being replaced by richer, more expressive, and biologically grounded imaging biomarkers. We separately examine how radiomics, pretrained natural-image transformers, and medically trained foundation models each affect prediction performance. We also study how they interact when combined. Together, these analyses establish a new framework for building robust, generalizable, and clinically meaningful imaging models to predict ovarian cancer treatment response.

2.2 Multimodal Data Integration

Radiologic and histopathologic imaging capture complementary but biologically distinct perspectives on ovarian cancer [28]. CT imaging of metastatic lesions provides a macroscopic assessment of tumor burden, spatial heterogeneity, and treatment-relevant radiographic phenotype prior to chemotherapy [29]. In contrast, hematoxylin–eosin (H&E) histopathology obtained from the resected primary tumor reveals microscopic information such as cellular morphology, stromal architecture, and immune infiltration that cannot be captured on CT [30, 31]. Although these data originate from different anatomical sites (primary vs. metastatic tumors) and represent different stages of disease progression, both encode biological processes that influence chemotherapy sensitivity, including genomic instability, tumor–stroma interactions, and microenvironmental organization [30]. Together, they offer a multi-scale view of tumor phenotype that neither modality can provide alone [28].

In high-grade serous ovarian carcinoma, quantitative CT analysis has demonstrated predictive value beyond visual assessment [29, 32], while computational pathology of primary tumor WSIs has been shown to reflect intrinsic biology, including molecular subtype, genomic instability, and stromal activation patterns known to correlate with therapeutic response [30, 31]. Recent AI methods can extract complex morphologic features from WSIs at cellular and subcellular resolution. They reveal prognostic patterns that may miss to obtain manually. These microarchitectural signals reflect early biological processes such as proliferation, desmoplasia, or immune exclusion that may still be present in metastatic sites and influence treatment outcome [30].

Multimodal fusion of radiologic, pathologic, and genomic data has the potential to improve prediction of survival and chemotherapy response, particularly in high-grade serous ovarian cancer [28]. However, major challenges must be addressed. CT and WSI data differ drastically in spatial scale, intensity range, and feature distributions. Moreover, primary and metastatic tumors may not fully align biologically, and many cohorts lack complete multimodal data [33]. Effective harmonization strategies are therefore required to bridge the gap between macro-level radiologic phenotypes and micro-level tissue architecture [33].

Recent advances in multimodal learning offer new tools to address these barriers. Attention-based fusion models have demonstrated the ability to dynamically weigh modality-specific contributions [33], while foundation models pretrained on millions of pathology and radiology images can generate stable, transferable embeddings even when training data are limited [34]. These large pretrained encoders can extract biologically meaningful representations from both modalities. They project them into a shared latent space. This enables robust cross-modal integration, even when primary and metastatic disease sites differ [34].

In study four, we build on these developments by integrating CT-based radiomic

and deep learning features extracted from metastatic tumors with foundation-model-derived embeddings from primary tumor histopathology. This work extends the methodological progression established in the previous studies by moving beyond unimodal imaging toward a multi-scale biological representation of ovarian cancer.

The central hypothesis of this study is that combining macroscopic CT phenotypes of metastatic lesions with microscopic histopathologic embeddings from the primary tumor will improve prediction of chemotherapy response, despite the anatomical and temporal differences between these data sources. We hypothesize that biological programs governing treatment sensitivity such as proliferative capacity, stromal activation, and immune engagement are reflected both in the primary tumor microenvironment and in the radiographically visible characteristics of its metastatic progeny. Therefore, integrating these complementary modalities will yield a more complete and biologically grounded signature than either modality alone. By applying this multimodal fusion approach, the fourth study aims to develop a prediction model and cross-scale imaging biomarker that captures intrinsic tumor biology from primary-tumor histopathology and treatment-time disease burden from metastatic CT scans. This model is expected to improve the accuracy and generalizability of chemotherapy-response prediction in ovarian cancer patients.

Chapter 3

Comparing the Effectiveness of 2D and 3D Features on Predicting the Response to Chemotherapy for Ovarian Cancer Patients

3.1 Introduction

Ovarian cancer is the second leading malignancy which accounts for more deaths than other carcinomas in gynecologic oncology [35]. For this type of disease, chemotherapy is a common effective treatment method in downstaging locally advanced malignancies. However, due to instability in genetics of metastasized tumors, their behavior and response to chemotherapy varies across the individual patients [36]. To deliver an individualized cancer therapy, it is clinically imperative to identify the tumor's response at an early stage. Among all different kinds of predicting methods, the development of a CT image based clinical marker has its advantages including low cost, high availability and reliability [37]. For this kind of method, identifying effective 2D or 3D features to quantify the tumor characteristics is critically important, as it directly determines the accuracy of the final prediction model. Some previous studies utilized the radiomics

features computed from one image slice (2D features) to assist cancer diagnosis and/or prognosis prediction [9, 38], and some studies tried to use radiomics features computed from multiple CT image slices to represent 3D features of tumor volume [39]. It has not been well investigated whether using 2D or 3D radiomics features can achieve higher and more robust prediction performance. In our group, we have conducted a series of investigations on model development, but the total number of the features are limited, and the effectiveness of the 2D and 3D features are not analyzed. For this investigation we directly compare the performance of the 2D and 3D radiomics features on the same dataset. A total of 1,032 (2D) and 1,595 (3D) radiomics features were applied on the same dataset of CT images acquired from the ovarian cancer patients, which are then used as the input of the same support vector machine (SVM) models. Their performances are tabulated and compared. All the experimental details are illustrated below.

3.2 Materials and Methods

For this investigation, we have retrospectively collected a total of 99 ovarian cancer patients, who were identified by our research physicians at Stephenson Cancer Center. Each patient in the database has high-grade and recurrent ovarian carcinoma and received systemic chemotherapy to control the metastatic tumors after the primary cytoreduction. Each patient had one pre-therapy CT assessment and on 4–8-week post-therapy CT examination, and the pre-therapy CT images were used in this experiment. According to the RECIST guideline [9], a number of 1 to 5 tumors are identified by radiologists for each case. Meanwhile, we also gathered the 6-month progression-free survival data for performance evaluation. The clinical data indicated that 55 cases are responders to chemotherapy according to the 6-month PFS, and 44 individuals are

non-responders with little or no noticeable response to the therapies.

Accordingly, a computer-aided image processing scheme was developed to segment each identified tumor in the dataset. A hybrid segmenting algorithm was adopted in this study, and its effectiveness has been validated in the previous studies [38]. Given the complexity of the metastatic tumors, manual contour correction was also provided in the software to allow researchers to visually examine and improve the automatic results. Since the tumor contains more than one CT slice, the segmentation starts at the central slice marked by the radiologist and continues on its neighboring slices until the tumor disappears.

Next, radiomics features were extracted on each individual tumor to quantify their characteristics [15, 25, 39–42]. The tumor feature computation can be divided into the following steps. The original CT slices were first processed by the following filters: exponential, gradient magnitude, local binary pattern (LBP), logarithm (Log), square, square root, and wavelet transform (Coif1, 8-level decomposition). Then, three different categories of radiomics features were applied on the filtered results as well as the original image, which includes shape, density, and texture-based features. As a result, a total of 1595 and 1032 radiomics features were computed for each 3D and 2D tumor, respectively. Since one patient may have more than one tumor, the maximum feature value among all the tumors from one patient was determined as the case-based value.

After that, for the initial feature pool, the Pearson correlation coefficients were computed, and highly correlated feature pairs are discarded [43]. On the rest of the feature set, the least absolute shrinkage and selection operator (LASSO) was applied to further reduce the feature redundancy and generate an optimal feature cluster. This algorithm is a L1 regularization based optimization model, which was first developed by Tibshirani [44]. Finally, two SVM models were applied on the 2D and 3D feature clusters to generate a prediction score, indicating the likelihood that the patient is

responsive to the selected treatment. The models were optimized by a 5-fold cross-validation approach [45], while receiver operating characteristics (ROC) analysis and confusion matrix were used to assess the performance of the models.

3.3 Results

Figure 3.1 gives one example of the segmented tumor. At the central slice, the identified tumor has the largest size, and it gradually decreases until it completely disappears on the 4^{th} neighboring slice. Table 3.1 summarizes the optimal feature cluster selected by the LASSO algorithm. 34 and 38 best performed features were identified from 2D and 3D tumor image feature sets, respectively. Among the 34 features derived from 2D tumor images, 27 of them are texture-based features, while the other 7 features are from either first-order statistics or shape feature category. A similar trend was observed in 3D feature set.

Figure 3.2 a and b provide the heatmap of the 2D and 3D features, respectively. The color of each entry indicates the Pearson correlation coefficients, which was generated by all the cases. The figures reveal that some of the features extracted from wavelet and logarithm images have relatively stronger correlations than the others. Moreover, fig. 3.2 c and d demonstrate the histograms of these coefficients, which reveal that more than 80% of these values are below 0.5. These results imply that the selected features have low redundancy, which can comprehensively represent different types of tumor characteristics. Figure 3.3 demonstrates the performance of the predicting models when using the 2D and 3D features. In Figure 3.3, the ROC curves show that the 2D feature-based model achieves an AUC of 0.85 ± 0.03 with a 95% confidence interval of [0.79, 0.90]. When applying a threshold of 0.51, the 2D feature-based model yields a PPV and NPV of 91% and 66% respectively, while the overall predicting accuracy is

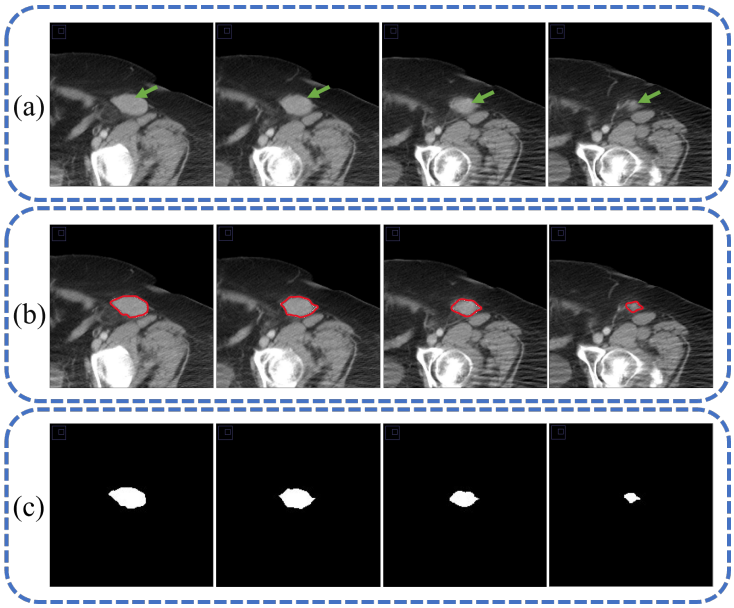


Figure 3.1: An illustration of the tumor segmentation process. (a) Identified tumors on the CT slices. (b) Tumors outlined using our developed CAD software. (c) Segmented tumor masks.

0.76 (confidence interval [0.71,0.83]). Similarly, when the 3D feature-based model was applied, the AUC value was achieved as 0.89±0.02 (confidence interval [0.84, 0.93]), and predicting accuracy increased to 0.81 (confidence interval [0.75,0.86]). These data were tabulated in table 3.2 and table 3.3 for performance comparison.

Table 3.1: Summary of the optimal feature clusters

Feature set	Texture	First-Order	Shape	Total
2D	27	7	0	34
3D	30	6	2	38

3.4 Discussion and Conclusion

In this study, we assessed the discriminative power of the 2D and 3D radiomics features in the early-stage prediction of chemotherapy treatment performance. The results

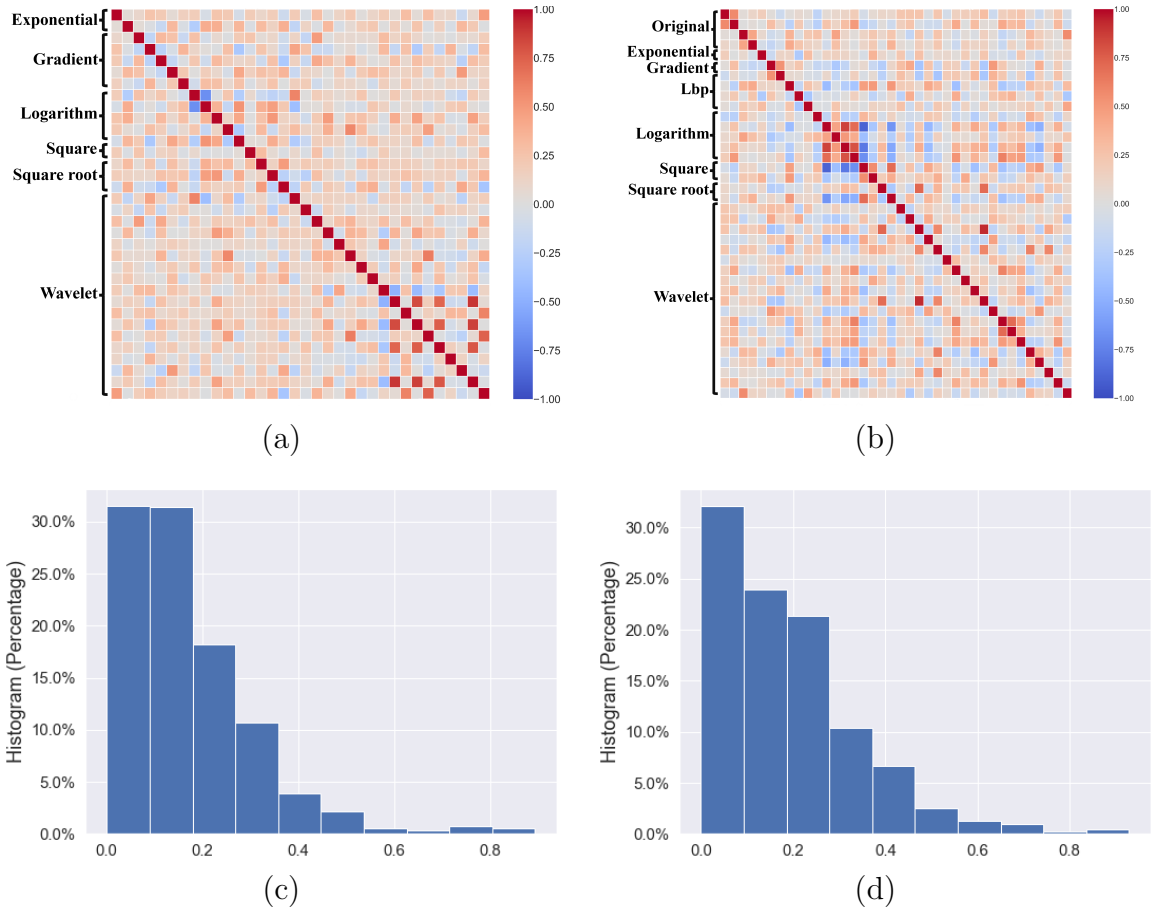


Figure 3.2: The heatmap of the Pearson correlation coefficients for (a) 2D features and (b) 3D features, and the corresponding histograms for (c) 2D features and (d) 3D features.

Table 3.2: Performance comparison of the 2D and 3D feature-based prediction models

Model	AUC (95% CI)	Accuracy (95% CI)
2D	0.85 (0.79, 0.90)	0.76 (0.71, 0.83)
3D	0.89 (0.84, 0.93)	0.81 (0.75, 0.86)

Table 3.3: Confusion matrix for 2D and 3D feature-based prediction models

6-month PFS	2D Model		3D Model	
	Yes	No	Yes	No
Yes	50	5	52	3
No	15	29	11	33

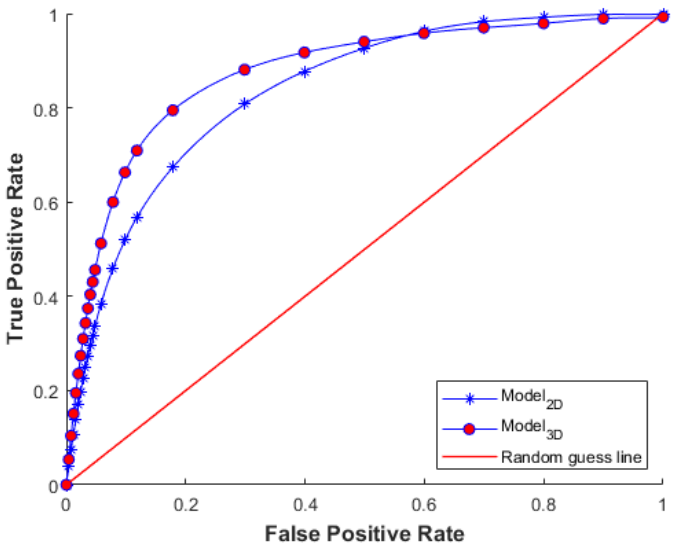


Figure 3.3: ROC curves generated by the 2D and 3D feature-based prediction models

revealed that 3D features in this task outperformed 2D features. Meanwhile, our research has the following unique findings. First, both the 2D and 3D features achieved an area under the curve (AUC) of 0.85 and 0.89, respectively, which demonstrated that these selected radiomics features can effectively characterize the tumor heterogeneities. These heterogeneities significantly contribute to the clinical outcomes [46], which is critically important for precision medicine. Second, the 3D features outperform the 2D features, which can be attributed to the fact that the 3D features are from the segmented 3D tumors on multiple slices. Thus, more useful information can be identified and extracted. Meanwhile, from the ROC curves, we also observed that the 2D features are comparable and even slightly better perform when the false positive rate is high. This implies that adding more slices may deteriorate the system accuracy in some situations, as more slices might also bring in more unrelated noises to negatively impact the prediction.

Although the encouraging results are demonstrated in this study, we still have several limitations as follows. First, only a small dataset containing 99 individuals is

used, and our results should be comprehensively evaluated on a large and diversified dataset in next step. Second, the performance of the radiomics features is highly relevant to the tumor segmentation [47], and the segmenting accuracy is not considered in this initial investigation. Third, we only compared the 2D and 3D tumors in this experiment. Our study results demonstrate that using 3D features yields a higher prediction accuracy than using the model trained by 2D features, which is different from another previously published study [48]. This previous study showed that the model using 2D features performs better than the model using 3D features. Thus, more studies are needed to further investigate and identify the optimal method. To investigate this, we can sequentially add individual slices on the central tumor slice, thus some "incomplete" 3D features will be generated, and the trend of the performance can be analyzed accordingly. This should the one meaningful direction in our future study.

Chapter 4

Evaluating the Effectiveness of 2D and 3D CT Image Features for Predicting Tumor Response to Chemotherapy

4.1 Introduction

Ovarian cancer is the most aggressive malignancy in gynecologic oncology [49]. Given the difficulty in the early stage diagnosis of ovarian carcinoma, most patients are diagnosed with ovarian cancer at advanced stages [50], which can lead to the formation of metastatic tumors on various organs. To control these metastatic tumors, chemotherapy is the only effective treatment after the primary cytoreduction [51]. However, due to the nature of metastasis heterogeneity, the patients' response to treatment varies largely. Thus, one challenge in current clinical practice is to select the appropriate chemotherapy before administrating the treatment. To address this, many studies have been conducted to identify biomarkers from either histologic or/and imaging data that might be associated with patient response to chemotherapy [52–55]. Among different technologies, imaging approaches have the advantage of evaluating the organ/tissue

noninvasively with adequate spatial resolution. Moreover, being a widely available imaging modality with relatively operating low cost, CT is one of the most popular diagnostic tools in therapy response assessment [56]. However, the current criteria requires both pre-therapy and 6–8 week follow-up examinations, and the diameter-based evaluation causes low association with the clinical outcome [57].

Meanwhile, radiomics is an emerging technology that extracts a large amount of features from tumors segmented from various medical images [15, 25, 42]. These radiomics features are able to quantify the complex textures, shapes, as well as other tumor characteristics that are highly associated with carcinogenesis [58]. Therefore, this information can be beneficial in determining treatment options and predicting the tumor's response to therapy [59]. For this technique, identifying effective radiomics features is critically important: 2D features are easier and faster to calculate but provide less information, while the 3D features use the entire 3D tumor volume and intuitively contain more comprehensive information. To this end, many studies have been conducted to explore the effectiveness of 2D and 3D features in different applications. The prognostic performance of 2D and 3D radiomics features in CT images of non-small cell lung cancer (NSCLC) was investigated in [60], which demonstrated that 2D features exhibited better performance, although both feature types had strong prognostic capability. One study compared the 2D and 3D radiomics features in characterizing gastric cancer [61], which revealed that models constructed with 2D radiomics features had comparable performance with those constructed with 3D features. In another study, Xu et al. [62] evaluated the prediction performance of 2D and 3D radiomics features in a multi-organ, multi-modality cancer study, and the authors concluded that 3D radiomics features were more effective. Lee et al. [63] compared the 2D and 3D texture features for discriminating between gastric cancer and normal gastric mucosa on CT images, and they revealed that 3D texture features were more effective than

2D features. Despite their clinical usefulness, these studies only compare features from the single-slice-based 2D tumor and the full-volume 3D tumor. Very few studies have been focused on the effectiveness of incomplete 3D tumors features, which contain only a portion of the image slices.

In this investigation, we hypothesized that the inclusion of multiple tumor slices would lead to a substantial enhancement in the predictive accuracy of our model, but it may also introduce more uncertainties that may adversely impact model accuracy. To verify this hypothesis, we extracted 2D radiomics features from a single slice of the tumors and then generated incomplete 3D features by sequentially adding the adjacent CT slices. Based on each individual feature pool, we optimized the same kind of predictive model and evaluated their performance, respectively. The details are presented in the following sections.

4.2 Materials and Methods

4.2.1 Database

In our study, we utilized a dataset of CT images involving 188 ovarian cancer patients. The CT images in the dataset were retrospectively collected from the University of Oklahoma health science center (OUHSC). Subjects for this study included the following criteria: (1) they had recurrent ovarian/peritoneal/tubal carcinoma of high-grade histology (e.g., serous, endometrioid, and undifferentiated); (2) they received systemic chemotherapy treatment after primary cytoreduction. For each patient in the dataset, there were up to 5 metastatic tumors occurred in different organs, which were marked on the pre-therapy CT images. All these images were acquired under a routing clinical protocol. CT examinations were captured utilizing either a GE Light Speed VCT

64 or a GE Discovery 600 16-row detector machine with a tube voltage of 120 kVp. The currents were adjusted from 100 to 600 mA, depending on patient body size. The clinical outcome was also collected for model assessment. In this study, 6-month progression-free survival was utilized, which a common index for assessing the efficacy of new cancer chemotherapies in clinical trials [64]. Accordingly, among all the 188 patients collected, 130 patients are responders with a 6-month PFS of 'Yes', whereas 58 patients are non-responders with little response to the treatments. The detailed demographic characteristics of our patients cohort can be found in table 4.1. The entire study protocol was approved by the institutional review board in the University (IRB13649).

Table 4.1: Patient characteristics in the study cohort

Patient Characteristics	6-Month PFS		p-value
	Responders	Non-responders	
Number of patients	130	58	–
Age (years)	63 ± 10	61 ± 10	0.24
Number of tumors	272	133	–
Average tumor diameter (mm)	31.3 ± 19.3	31.5 ± 16.4	0.94

4.2.2 Radiomics Feature Extraction

In this study, the tumor features were obtained from 2D and 3D pre-therapy CT slices. Each segmented tumor is depicted on a number of CT slices. The central slice has the largest 2D area, which was identified by radiologists. Accordingly, we first extracted the quantitative features only from the central slice to generate a 2D feature pool. We sequentially added the adjacent slices above and below the central slice until the tumor disappears. After adding one adjacent slice, we will create the corresponding incomplete 3D tumor feature pool (table 4.2). We continued the generation of the

corresponding incomplete 3D feature pool until the tumor disappears.

Table 4.2: Number of radiomics features extracted from the original and filtered images

Image Type	**2D**	**3D**
Original	104	110
Exponential	93	93
Gradient Magnitude	93	93
Local Binary Pattern (LBP)	93	279
Logarithm	93	93
Square	93	93
Square Root	93	93
Wavelet	741	741
Total	**1403**	**1595**

Before the feature computation, we developed a computer-aided image analysis scheme to segment each individual tumor from the CT images. The tumor segmentation procedure started at the central slice where the tumor was marked by a radiologist according to RECIST criteria [9]. The method then sequentially segmented the tumors on the adjacent slices until the tumor disappeared. For each slice, the tumor was segmented by a hybrid algorithm, which consists of two core algorithms: a region growth approach with adaptive thresholds [65] and a dynamic contour-seeker method [66]. The performance of this algorithm was assessed and validated by a number of our previous studies [38, 67–69]. Given that metastatic tumors occurred on a number of human organs with high heterogeneity, the automated segmentation may not be able to generate the tumor contour with satisfactory accuracy. Thus, these segmentation results were visually evaluated by experienced researchers and were manually corrected if needed. The entire scheme was developed as an ImageJ plugin [70], which was equipped with a user friendly graphical user interface (GUI) to visualize the 2D and 3D tumors.

Based on the segmented tumors, we generated the 2D and 3D feature pool using the following methods [25, 42]. The feature computation scheme is based on pyra-

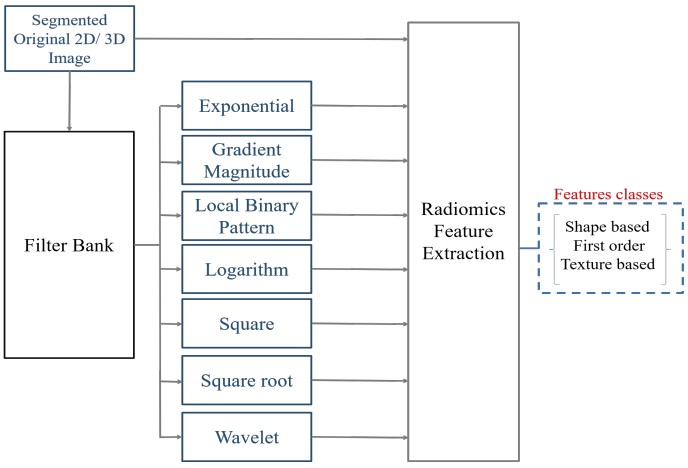


Figure 4.1: Overview of the feature extraction process.

diomics [25], an open source platform that calculates features in accordance with the definitions in the imaging biomarker standardization initiative (IBSI) [42]. The general flowchart of the calculation is described in fig. 4.1. In this module, the segmented tumor was first processed by a number of operations including exponential, gradient magnitude, local binary pattern (LBP), logarithm (Log), square, square root, and wavelet (Coif1) filters. Next, the radiomics features were then calculated on the processed images as well as the original image. The features can be divided into 3 categories, i.e., geometric, density, and texture features, which describe the tumor in various aspects [41]. Table 4.2 shows a breakdown of the number of features calculated for each image class, with the number of features per class ranging from 93 to 741. We calculated a total of 1595 and 1403 features for the 3D and 2D feature pools, respectively. Given that a total of up to five tumors were segmented by the scheme and the features were extracted on each individual tumor, the case-based value of each feature is generated by calculating the minimum $F_{\min} = \min\{\ F_1,\ F_2, F_3, \ldots, F_N\}$ all the tumor-based values. The minimum value was selected on the basis of our empirical analysis, as it consistently demonstrates its effectiveness in providing a stable and conservative feature representation across the segmented tumors for each patient.

4.2.3 Developing Machine Learning-Based Models to Predict Tumor Response to chemotherapy

Pearson correlation coefficient (PCC) and LASSO approaches were used to reduce the dimensionality of our feature space. First, the PCC was calculated, and a heatmap was generated to investigate the feature dependency. After the computation, feature pairs with a PCC exceeding the pre-set threshold were identified, and one of the correlated features was removed [71] to reduce the redundant information. Next, LASSO was applied on the rest of the feature pool to create an optimal feature cluster [44], as it has the capability to handle high-dimensional data with a limited number of observations [72]. Essentially, this method employs a variation of the least squares regression method that employs L1 regularization to produce sparse variable coefficients. Accordingly, LASSO constrains the sum of the absolute values of the regression weights to be smaller than a fixed threshold. Therefore, it improves model interpretability by removing variables irrelevant to the response variable and reduces the possibility of overfitting [73]. The LASSO objective function is defined in Equation eq. (4.1).

$$\sum_{i=1}^{n}\left(y_i - \sum_{j} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \qquad \text{subject to} \qquad \sum_{j=1}^{p}|\beta_j| < t \tag{4.1}$$

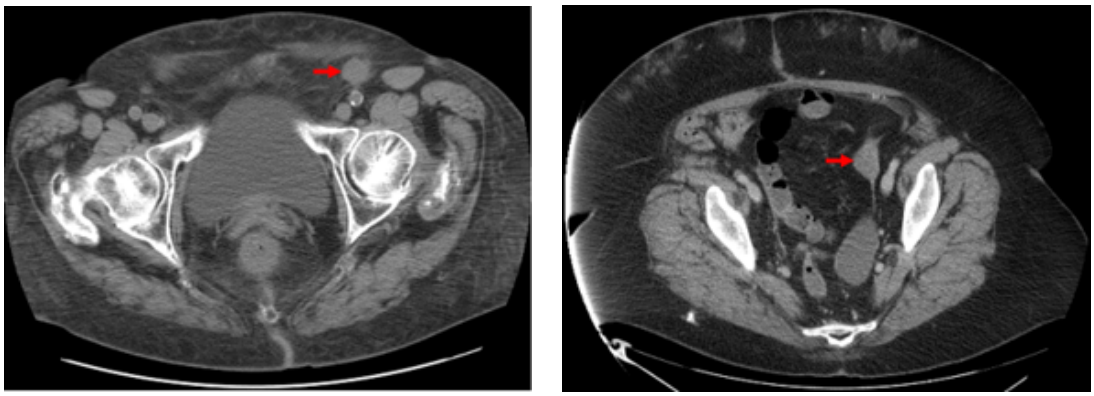
In the above error function, t is the upper-bound threshold for the sum of the absolute values of feature weights β, and λ is a non-negative tuning parameter that regulates the degree of the penalty. We employed the LASSO-cross validation (LASSO-CV) approach to tune the LASSO parameters [74]. After applying LASSO, features with non-zero weights were chosen to be included in the model. Finally, we adopted the synthetic minority over-sampling technique (SMOTE) to add more non-responder samples and balance the dataset, which generates samples by interpolating between

one selected minority sample with its nearest neighbors, which was performed using scikit-learn library [75].

Next, a SVM was used to predict the tumor responses (i.e., 6-month PFS). SVM has been widely validated as an effective classifier in various medical imaging tasks [76–79]. The method constructs one or more hyperplanes to classify the feature vectors into two classes with minimal generalization error [80]. Mathematically, the SVM model is established by solving the following optimization problem: $\min \frac{1}{2} \|w\|^2 + \frac{C}{2} \|\xi\|_2^2$ with the constraint $y_i \left(w\Phi\left(x_i\right) + b_i\right) \geq 1 - \xi_i$, $\xi_i \geq 0$, $\forall i$ or its dual form $\max W(\alpha) = e^T\alpha - \frac{1}{2}\alpha^T y_i y_j K\left(x_i, x_j\right)\alpha$, $0 \leq \ \alpha_i \leq C$, $y^T\alpha = 0$. In the above formulae, (x_i, y_j) is one pair of feature vector and output value, w and b are the normal vector and intercept of the discrimination hyperplane, ξ_i is the slack variable, C is the adjusting coefficients for optimization errors, e is the unit vector, α is the Lagrange coefficient, and K represents kernel function $K\left(x_i, x_j\right) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$ [81]. The model's parameters (i.e., C and σ) will be tuned to achieve optimal values during the training process. The SVM parameter tuning was executed using grid search [82], which varies the parameters systematically to search through a predefined range of values and find the set that produced the most desirable performance metrics (e.g., accuracy). This process was conducted to ensure the robustness and reliability of our proposed approach, resulting in the selection of the most suitable parameters for the SVM model. Finally, each classifier was then trained and tested using a 5-fold cross-validation approach [83]. To evaluate the model performance, the ROC curve was used [84], and the area under the ROC curve (AU-ROC) was estimated for each model.

4.3 Results

Figure 4.2(a), (b) illustrate tumor images from a representative non-responder and a representative responder, respectively. Although there are no visible feature differences between the responder and non-responder, the computed radiomics feature can classify them into different groups. Figure 4.3 shows an example of metastatic tumor segmentation. The tumor has the largest area in the central slice, and its area gradually decreases until the tumor eventually disappears on the fourth neighboring slices of both sides (fig. 4.3(a)). Accordingly, the segmentation procedure starts with the central slice, where the tumor was identified by a radiologist or oncologist. Then, the segmentation continues on the adjacent slices until the tumor disappears (fig. 4.3(b), (c)).



(**a**) (**b**)

Figure 4.2: Representative tumor images from (a) a responder patient and (b) a non-responder patient.

Figure 4.4(a-c) represents the heatmap of all the features extracted from 2D, incomplete 3D, and 3D tumor, respectively. The maps were produced by a total of 188 observations. Each entry in the map indicates the colorized Pearson correlation coefficient ranging from 0 (Blue) to 1 (red). The maps illustrate that most of the features extracted from 2D tumor masks have relatively low correlation. A similar pattern can

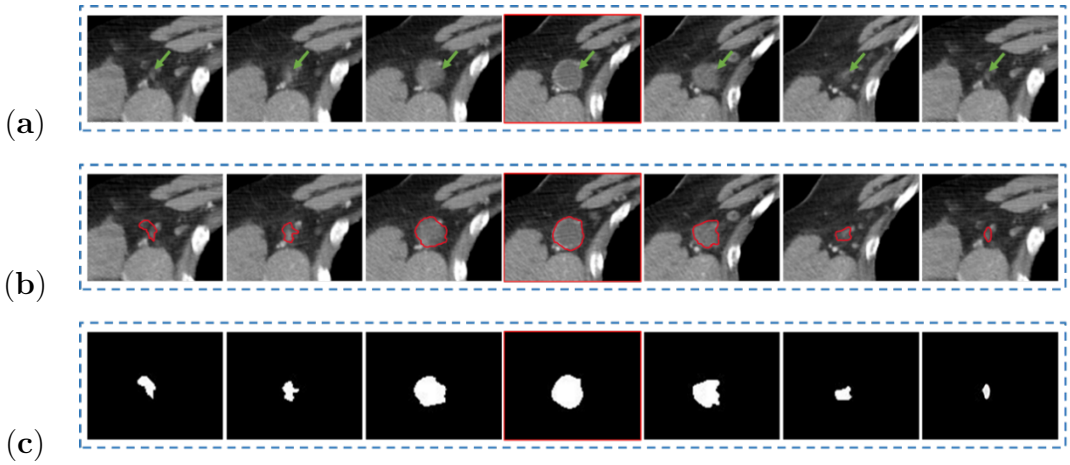


Figure 4.3: Example of the tumor segmentation process. The central slice is outlined in red. (a) Identified tumor on the CT slices. (b) Tumors outlined using the developed CAD software. (c) Segmented tumor masks.

be observed in the features obtained from incomplete 3D tumor masks. However, the correlation structure of the features extracted from whole 3D tumor masks was notably different, particularly in logarithm-based, exponential, square, and square root-based features, exhibiting higher correlation among different classes. The high correlation may be attributed to the enhancement of certain 3D structures when applying these filters, which leads to consistent value changes on some categories of the features. Furthermore, fig. 4.4(d-f) demonstrates the histograms of all feature categories for 2D, incomplete 3D, and 3D features. According to the histograms, more than 80% of correlation coefficient values for 2D and incomplete 3D features are less than 0.5, whereas a higher percentage of 3D features have correlation coefficients greater than 0.5. These results imply that 2D and incomplete-3D tumor features cover comprehensive tumor characteristics with low information redundancy.

After applying the PCC and LASSO feature selection methods, we identified 115 two-dimensional features as the most effective feature cluster for further analysis. Similarly, 56–126 features were identified for different incomplete 3D tumor features pools, and 73 features were identified from the full 3D feature pool. As illustrated in table 4.3,

a minimum of 78% of the selected features can be categorized into texture-based features, while the remainder fall into the categories of shape or density features. When dividing the features based on the filters, we observe that the largest group of features is extracted from the wavelet transform-based filtered images, with the features calculated from LBP-filtered images forming the second most prominent group. Our models used these features as inputs to generate the prediction score. Figure 4.5 illustrates the ROC curve achieved by the models trained with one to nine slices of tumor mask and whole slices, while table 4.4 summarizes the corresponding values of AUC and overall accuracy values. The results demonstrate that adding more slices to the central slice can increase the prediction performance, but only for a certain range of additional slices. The highest performance was achieved when the model was constructed with one central slice and two adjacent slices, with the model yielding an AUC of 0.91 $\pm$ 0.01 (95% confidence interval [0.88, 0.94]) and an overall predicting accuracy of 0.84 $\pm$ 0.03 (95% confidence interval [0.78, 0.88]). The 3D model with three and 5 adjacent slices were the second and third best performing models, which achieved an AUC value of 0.89 $\pm$ 0.01 and 0.86 $\pm$ 0.02, respectively. Model_3D, with an AUC of 0.86 $\pm$ 0.02 (95% confidence interval [0.82, 0.89]), performed better than Model_2D, with AUC of 0.84 $\pm$ 0.02 (95% confidence interval [0.78, 0.88]). All these trends are also illustrated in fig. 4.5(b).

4.4 Discussion

In this study, we conducted a comprehensive evaluation of various radiomic feature extraction methods for predicting early response of ovarian cancer to chemotherapy. Our investigation focused on 2D, incomplete 3D, and 3D radiomic features and their ability to represent and discriminate the underlying tumor characteristics. Our findings

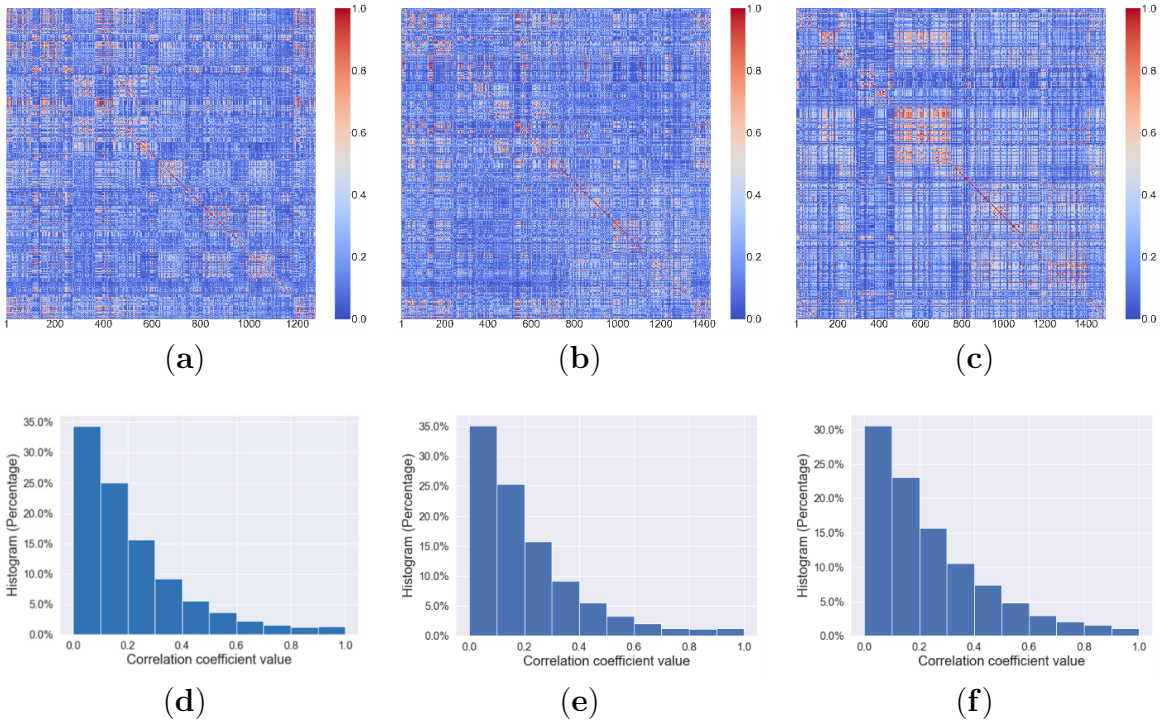


Figure 4.4: Feature heatmaps of (a) 2D tumors, (b) incomplete 3D tumors, and (c) 3D tumors. The corresponding histograms are shown in (d)–(f) for 2D, incomplete 3D, and 3D features, respectively.

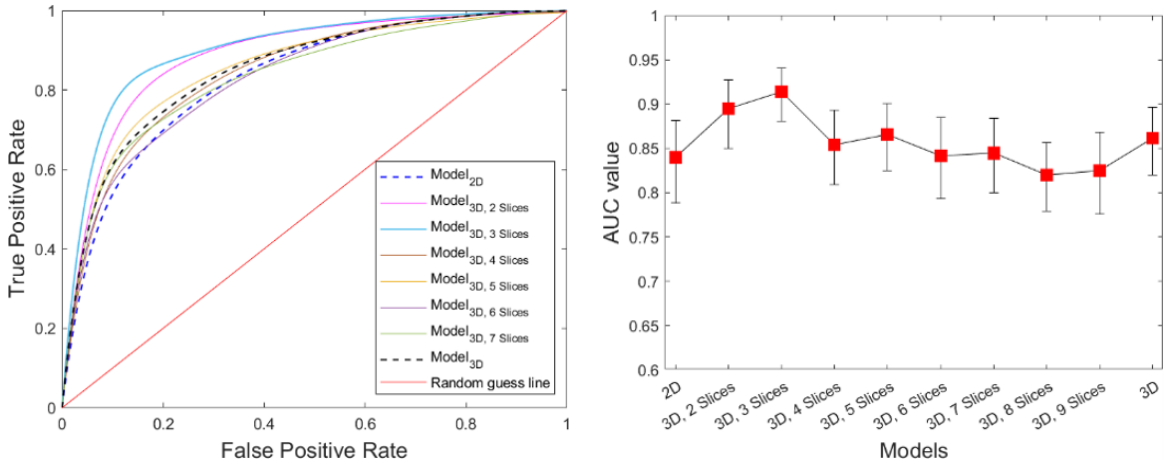


Figure 4.5: Performance comparison of the models for six-month progression-free survival prediction. (a) ROC curves of the prediction models. (b) Corresponding AUC values with 95% confidence intervals.

Table 4.3: Summary of selected features by radiomics category and filtering method

Features	Radiomics Category			Filter			Total
	Shape	Density	Texture	Wavelet	LBP	Other	
2D	0	23	92	71	21	23	115
$3D_{2\,\text{Slices}}$	4	18	104	77	19	30	126
$3D_{3\,\text{Slices}}$	0	16	75	51	9	31	91
$3D_{4\,\text{Slices}}$	1	20	75	52	13	31	96
$3D_{5\,\text{Slices}}$	1	17	64	44	12	26	82
$3D_{6\,\text{Slices}}$	1	16	72	51	14	24	89
$3D_{7\,\text{Slices}}$	1	15	61	37	16	24	77
$3D_{8\,\text{Slices}}$	0	11	51	27	7	28	62
$3D_{9\,\text{Slices}}$	1	8	47	30	4	22	56
3D	1	8	64	34	13	26	73

Table 4.4: Performance comparison of prediction models

Model	AUC ± STD (95% CI)	ACC ± STD (95% CI)
2D	0.84 ± 0.02 [0.78, 0.88]	0.75 ± 0.03 [0.69, 0.80]
$3D_{2\,\text{Slices}}$	0.89 ± 0.01 [0.85, 0.92]	0.83 ± 0.02 [0.77, 0.87]
$3D_{3\,\text{Slices}}$	0.91 ± 0.01 [0.88, 0.94]	0.84 ± 0.02 [0.78, 0.88]
$3D_{4\,\text{Slices}}$	0.85 ± 0.02 [0.80, 0.89]	0.76 ± 0.02 [0.71, 0.82]
$3D_{5\,\text{Slices}}$	0.86 ± 0.02 [0.82, 0.90]	0.78 ± 0.02 [0.72, 0.83]
$3D_{6\,\text{Slices}}$	0.84 ± 0.02 [0.79, 0.88]	0.74 ± 0.03 [0.69, 0.79]
$3D_{7\,\text{Slices}}$	0.86 ± 0.02 [0.80, 0.88]	0.76 ± 0.03 [0.70, 0.80]
$3D_{8\,\text{Slices}}$	0.84 ± 0.02 [0.78, 0.86]	0.75 ± 0.03 [0.70, 0.80]
$3D_{9\,\text{Slices}}$	0.83 ± 0.01 [0.78, 0.87]	0.73 ± 0.02 [0.69, 0.78]
3D	0.86 ± 0.02 [0.82, 0.89]	0.77 ± 0.02 [0.72, 0.83]

demonstrate that the incomplete 3D features extracted from the central tumor slice and its two adjacent slices yielded better prediction accuracy and might be recommended as the preferred approach.

As compared to the previous studies, the most unique characteristic of our investigation is that we explored the performance of the features computed from incomplete 3D tumors, which contain a portion of the tumor slices. The comparison between the 2D or 3D features has been investigated by many research groups. In general, the 2D features are calculated from a single slice containing the maximal projection within the entire tumor. Accordingly, tumor segmentation and annotation are relatively easier, but some useful information is lost. For instance, it is difficult to accurately extract tumor volume, surface area, spherical disproportion, and other features that reflect geometric irregularities from a 2D tumor. On the other end, despite the comprehensive information, the 3D feature suffers from more information uncertainties. Given that it is very difficult to clearly define the boundary between the tumor region and its surround healthy tissues, there is no segmentation algorithm which is able to perfectly extract the tumor contour [85]. Moreover, the tumor heterogeneity may only be inside several central slices. The marginal slices may only contain suspicious surrounding tissue, and they may not have any information associated with the needed treatment responses. As a result, more uncertainties will be added to the segmented 3D tumors, which may adversely affect the performance of prediction models. Therefore, there is a trade-off between adding more useful information and uncertainties when sequentially increasing the adjacent slices. In this study, we first demonstrated that using the central and its two neighboring slices may achieve better performance than other situations. To the best of our knowledge, no similar studies have been performed on this topic before.

Moreover, our study highlights the importance of texture-based features and wavelet

filtering for predicting treatment response. We observe that texture-based features were more informative, especially when they are computed from the wavelet transform-processed images. This could be attributed by the fact that texture-based features are able to capture the heterogeneity and complexity of the tumor microenvironment, which can provide insights into tumor biology and response to treatment. Wavelet transform is able to initially concentrate the tumor edge and texture information, while the other irrelevant information is discarded. Meanwhile, our study suggests that the number of slices used for CT image analysis can have an impact on the correlation structure of extracted radiomics features. The high correlation only occurs on the complete 3D tumor features from the tumor images processed by certain kinds of filters (e.g., logarithm, exponential, and square), which may be explained by the fact that such filters can enhance certain 3D structures, and these structures could lead to consistent value changes on some categories of the features.

Despite the promising results of our study, this study has the following limitations. First, the dataset used in our study only consisted of 188 patients, which were collected from a single institute. This single institute dataset may not be able to comprehensively represent the diversified population within the entire country, therefore the robustness of the results in this investigation should be further verified by a larger multi-institute patient cohort. Second, although the segmentation scheme was validated in our previous studies, the tumor segmentation accuracy was not considered. The segmentation error may also potentially introduce noise into our 3D features and consequently reduce the predictive performance [47]. Third, only radiomics features were used in the feature performance comparison. Other types of features, such as deep learning features, were not investigated. It would be worthwhile exploring the combination of deep learning features and radiomics features in future studies. In conclusion, our study may provide valuable information for the development of radiomics-based prediction models in the

future.

Chapter 5

A vision transformer-radiomics approach for enhanced chemotherapy outcome prediction in ovarian cancer

5.1 Introduction

Ovarian cancer is one of the most lethal gynecological malignancies [86]. Most ovarian carcinoma cases are diagnosed at advanced stages due to the difficulty in early detection, which increases the risk of metastatic tumor development on multiple organs [87]. Following primary cytoreduction, the standard treatment involves chemotherapy, directed at eliminating residual cancer cells and decreasing the risk of recurrence. Clinically, response to treatment is different among the patients due to the complexity of disease, particularly cellular and molecular heterogeneities of OC tumor anand2023a. Therefore, predicting treatment effectiveness in the initial stages of therapy can improve the precision of patient selection for tailored treatment approaches, notably chemotherapy [88]. A limitation in early prediction of treatment response for OC is the absence of biomarkers capable of reliably predicting chemotherapy efficacy [89, 90].

Certain biomarkers, like CA125 and human epididymis protein 4 (HE4), are associated with patient response to chemotherapy [54, 91]. However, their accuracy in predicting chemotherapy response varies among patients with different epidemiological and clinical features [92]. Medical imaging modalities offer a non-invasive tool that can capture comprehensive data characterizing tumor heterogeneity. Among different imaging modalities, CT is a widely utilized method for assessing therapy response due to its general availability and user-friendly nature ganeshan2013a. Tumor heterogeneity on CT images can be quantified using texture analysis which extracts spatial information that may not be perceptible to the naked eye [93].

Radiomics refers to the extraction and analysis of large amounts of quantitative imaging features with high throughput from two-dimensional and three-dimensional medical images [94]. These features comprise mathematical descriptions of various visual properties observed in the images, encompassing attributes such as shape, texture, and density. Radiomics provides an objective method to assess the tumor microenvironment, capturing information that is not perceptible to the human eye [95]. Several studies have been conducted to identify quantitative imaging markers related to treatment response using radiomics features [69, 96–98]. However, a major limitation of radiomics is its reliance on predefined, low-order features based on heuristic knowledge [99]. Extracting higher-level, more advanced features and integrating them into the radiomics framework could enhance predictive power.

Deep learning, including recent advancements such as vision transformers, and medical foundation models, has emerged as a prominent method for various image processing tasks, gaining increasing attention [100]. These models have been utilized to extract information of tumor phenotype and micro-environment in detection, prognosis, and treatment outcome prediction studies from medical images [101]. In the medical field, where datasets are often limited, a pretrained DL model can be applied

to the downstream tasks. This practice, known as "transfer learning," has been proven to be effective in many cases, such as predicting survival among lung adenocarcinoma patients and rectum toxicity in cervical cancer radiotherapy [102, 103]. Similarly, deep learning models trained on medical images, such as histopathology images, can be employed to predict treatment response in OC [104, 105]. However, there are limited existing studies rely on multi-method features that explore the full potential of integrating radiomics with high-level deep embeddings [106]. Furthermore, the relative contributions of models pretrained on natural images versus those fine-tuned on medical data remain insufficiently studied. There is a need for further investigation into how these different pretrained models complement radiomics features, and whether specific combinations of representations can enhance predictive performance in a synergistic manner.

Integrating advanced deep learning techniques with radiomics has the potential to overcome the limitations of traditional handcrafted features by leveraging rich, high-level embeddings. In this study, we explore the combination of radiomic features with deep representations extracted from ViT and MedSAM encoders, a medical image segmentation foundation model. We develop and evaluate multi-representation predictive models that integrate these diverse feature sets to predict chemotherapy response in OC treatment. This integrated approach aims to provide a more comprehensive representation of tumor characteristics and treatment dynamics, ultimately paving the way for more accurate and personalized cancer therapies. To the best of our knowledge, this is the first study to systematically explore the integration of radiomics and foundation model features for chemotherapy response prediction.

5.2 Materials and Methods

5.2.1 Database

In our study, we used an imaging dataset from 197 patients with OC treated at the university of oklahoma health science center. These images were retrospectively collected from patients with advanced-stage OC, presenting with recurrent peritoneal or tubal carcinoma. Patients with pre-treatment CT scans and known 6-month PFS were included and cases without imaging or follow-up data were excluded (fig. 5.1). The study population included participants with various histological subtypes, such as endometrioid, serous, clear cell, and mucinous carcinomas. For each patient, radiologists examined the pre-therapy CT images and annotated all metastatic tumors, resulting in one to five marked lesions depending on the individual's disease burden. Following diagnosis, patients underwent systemic chemotherapy treatment post primary cytoreduction. Imaging procedures followed routine clinical protocols using GE LightSpeed VCT (64-slice) or GE Discovery 600 (16-slice) machines, with tube currents ranging from 100 to 600 mA based on patient size. Clinical outcome data included 6-month progression-free survival, an endpoint frequently adopted in ovarian cancer phase II/III trials to enable early assessment of therapeutic benefit and support timely treatment adaptation. Several studies have demonstrated that improvements in 6-month PFS are associated with longer-term survival and are used to inform go/no-go decisions and regulatory evaluation in clinical trial settings [107, 108]. Of the 182 patients included in the final cohort, 124 showed positive 6-month PFS (responder), while 58 showed no or minimal response (non-responder). Table 5.1 presents the detailed demographic characteristics of the patients cohort in this study.

Table 5.1: Demographic and clinical characteristics of the study cohort.

Characteristic	Responder	Non-responder	p-value
Number of patients	124	58	–
Age (years)	63 ± 10	61 ± 10	0.24
Number of tumors	264	131	–
Average tumor diameter (mm)	31.3 ± 19.3	31.5 ± 16.4	0.94

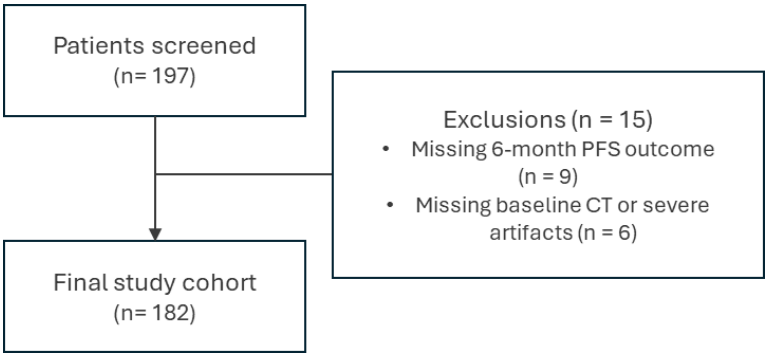


Figure 5.1: Flow diagram illustrating patient selection and cohort construction.

5.2.2 Image Feature Computation

Before feature extraction, CT volumes were first resampled to isotropic voxel spacing to ensure consistent spatial resolution across patients. A soft-tissue HU window (WL=40, WW=400) was then applied to enhance lesion visibility, and the processed slices were subsequently linearly scaled to 8-bit grayscale and exported as PNG images without further normalization for standardized handling. These windowed images were used for all subsequent segmentation and feature extraction steps. Following this, we adopted a tumor segmentation method to segment regions of interest (ROIs) from CT images [38]. This method included two key steps. First, an experienced radiologist annotated the largest tumor area on the CT images using the RECIST criteria [9]. Then, a semi-automated hybrid algorithm was applied to segment the tumor. This algorithm combines a topographic region growth approach with adaptive thresholds and dynamic edge tracking, ensuring accurate identification of the metastatic tumor slice

with the largest area. We used only the 2D slice with the largest tumor cross-section for all subsequent feature extraction to ensure consistency and minimize variability. Despite the algorithm's high performance, the inherent heterogeneity of metastatic tumors may occasionally result in suboptimal segmentation outcomes. To address this, we made manual corrections to refine the segmentation results as needed. The entire segmentation process was implemented as an ImageJ plugin [109] with a user-friendly graphical user interface (GUI), facilitating visualization and manipulation of 2D tumor contours. Based on the extracted image ROI, we computed handcrafted features and deep learning-based embeddings. In order to extract hand crafted features, we employed established radiomics methodologies to derive a comprehensive set of quantitative imaging features characterizing tumor texture, shape, and intensity [94]. In addition, we employed state-of-the-art pretrained ViTs and foundation models to extract high-level embeddings that capture intricate spatial information within the images [27, 100].

5.2.2.1 Radiomics Features

Radiomics feature computation was conducted using the pyradiomics platform [25], adhering to the definitions established by the imaging biomarker standardization initiative [22]. Initially, the 2D tumor images underwent a filtering process, including exponential, gradient magnitude, local binary pattern, logarithm, square, square root, and wavelet (Coif1) filters. Subsequently, radiomics features were computed on both the processed images and the original image, encompassing geometric, density, and texture features. Geometric features were exclusively extracted from the original image, offering insights into tumor morphology, while density features described the intensity distribution of pixels within the segmented region. Texture features are derived by matrices such as gray level co-occurrence matrix (GLCM) [110], gray level dependence

matrix (GLDM) [111], gray level run length matrix (GLRLM) [112], gray level size zone matrix (GLSZM) [113], and neighboring gray tone difference matrix (NGTDM) [111]. These features capture spatial patterns and variations in pixel intensities. In total, 1218 features were computed for the 2D radiomics feature pool. From each image type, 18 first-order features, 24 GLCM, 14 GLDM, 16 GLRLM, 16 GLSZM, and 5 NGTDM features were extracted, along with 9 geometric features extracted from the original image.

5.2.2.2 Deep Learning Features

In addition to handcrafted radiomics features, deep feature representations were extracted using pretrained transformer-based architectures. Specifically, embeddings were generated using a Vision Transformer model pretrained on natural images and MedSAM, a medical foundation model adapted from the Segment Anything Model (SAM). These embeddings were used to capture high-level image representations that may not be represented by conventional radiomics descriptors.

5.2.2.3 Vision Transformer Embeddings

In this study, we utilized the embeddings from the ViT base model [100]. The ViT architecture was originally pretrained on images from ImageNet dataset with a resolution of 224*224 for image classification purposes [114]. These embeddings encode high-level representations that reflect detailed characteristics of image intensity, texture, and edge information [115]. For this reason, each preprocessed grayscale CT image was replicated across three channels and cropped to a 224 $\times$ 224 region centered on the ROI. The resulting images were then divided into 14$\times$14 non-overlapping patches (fig. 5.2). Afterwards, the patches were flattened and linearly embedded into 1D vectors. Learnable positional encodings were added to these vectors, and a learnable [CLS] token

was prepended to the sequence. In the figure, let p_1, ..., p_{196} denote the patches of an input image. The transformer encoder comprises a stack of multiple blocks, each composed of multi-head attention and feed-forward neural networks. At its core, the transformer encoder leverages the self-attention mechanism, enabling the model to assess the importance of different patches in the sequence during data processing [100]. This self-attention mechanism effectively captures global contextual information, facilitating the learning of long-range dependencies and relationships among image patches. Finally, to refine the feature representation, we averaged the feature vectors corresponding to patches with tumor and excluded those from unaffected areas, resulting in a 1D feature vector with a size of 768 dimensions.

5.2.2.4 MedSAM Embeddings

MedSAM [27] is a foundation model built upon the SAM [116]. It is fine-tuned on more than 1.5 million medical images, including more than 500,000 CT scans, each paired with segmentation masks to capture domain-specific visual representations. The model architecture consists of a transformer-based image encoder, a prompt encoder, and a mask decoder. The image encoder is based on the ViT-Base model, which was pretrained using masked autoencoding [117]. Given the strong representational capacity of its ViT encoder, which is trained on large-scale medical datasets with extensive CT coverage, we used MedSAM as a feature extractor in this study. This allowed us to capture high-level semantic features relevant to our prediction task.

In our study, we resized each 2D CT image from 512×512 to 1024×1024 pixels to match the MedSAM input format. Embeddings were then extracted from non-overlapping 16×16 patches of the image. To ensure focus on tumor-relevant regions, we adopted a patch selection strategy similar to that used in ViT, retaining only the embedding vectors related to patches that contained at least 50% tumor area. Finally,

average pooling was applied to generate a fixed-length feature vector (of size 768) for each tumor.

When multiple lesions were present in a patient, radiomics and deep features were computed separately for each lesion. To generate a patient-level feature representation, lesion-level feature vectors were aggregated using average pooling, producing a fixed-size feature vector per patient regardless of lesion count. The resulting patient-level vectors were used as input for model training and outcome prediction.

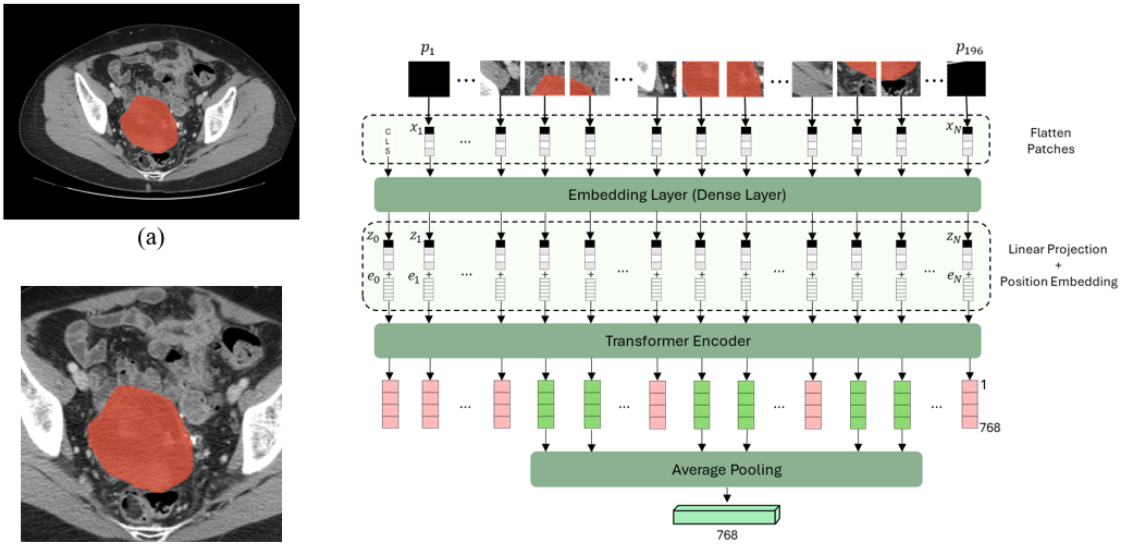


Figure 5.2: Vision Transformer-based feature extraction workflow. (a) CT image with highlighted tumor. (b) Tumor region of interest. (c) Schematic representation of ViT-based feature extraction.

5.2.3 Machine Learning Model Development

To analyze the effectiveness of features from radiomics alongside deep learning methods, we developed distinct machine learning models that utilized features from radiomics, ViT, MedSAM, and their integrations (fig. 5.3). For the integration models, we concatenated features from the different sources to form a unified feature vector. All

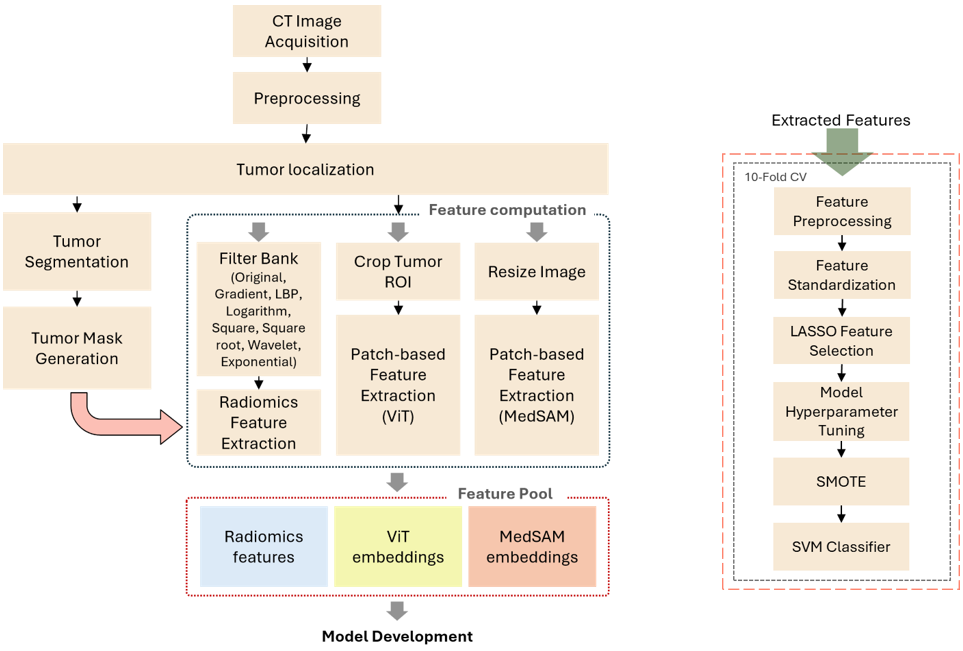


Figure 5.3: Overview of the proposed framework for chemotherapy response prediction. (a) Feature extraction pipeline illustrating the computation of handcrafted radiomics features, vision transformer embeddings, and MedSAM embeddings from segmented tumor regions. (b) Machine learning workflow including feature preprocessing, correlation filtering, SMOTE oversampling, LASSO-based feature selection, support vector machine (SVM) classification, and nested cross-validation evaluation.

models aimed to predict the 6-month PFS following chemotherapy treatment in OC patients.

Prior to model training, all preprocessing and modeling steps were performed within the inner loop of a nested, stratified cross-validation framework. All data-dependent procedures were fit exclusively on the training split and subsequently applied to the corresponding held-out data. Initially, features with zero variance were identified and removed [118]. We then standardized features using z-score normalization [119]. It transforms each feature x to have zero-mean and unit standard deviation according to $z = (x - \mu)/\sigma$ (where z is standardized feature value, x is the original feature value, μ and σ are mean and standard deviation computed on the training split). Next, we computed pairwise Pearson correlations and, for each highly correlated pair (above a

predefined threshold of 0.95), removed one feature to reduce redundancy, improving generalization and reducing overfitting risk [120].

Considering the imbalance nature of the dataset, we applied the synthetic minority oversampling technique on the training split inside the cross validation loop [75]. SMOTE addresses class imbalance by generating synthetic samples for the minority class, thereby mitigating the risk of bias towards the majority class. This algorithm works by first selecting a non-dominant sample at random and then obtaining its k nearest neighbors (k = 5). Next, a specified number of new synthetic instances are created by interpolating between the minority sample and its neighbors in the feature space. By calculating the difference in distance between the feature vector and its neighbors, SMOTE synthesizes new instances that lie within the vicinity of the existing minority instances, effectively augmenting the minority class.

Next, dimensionality reduction was performed using the least absolute shrinkage and selection operator with cross-validation method to form optimal feature ensemble. LASSO-CV is a modified version of the least squares regression technique incorporating L1 regularization to generate sparse coefficients for features [121]. In this study, feature selection was carried out using Lasso CV from scikit-learn. Although the outcome variable was binary, labels were coded as 0/1 and LASSO was used solely as a sparsity-inducing feature selector, rather than as a probabilistic classifier. All features have been z-score standardized using training data prior to LASSO, and the regularization parameter λ was selected via the default 5-fold cross-validation within the training data by minimizing mean-squared error. In LASSO regression, the coefficients are estimated by minimizing $\min_\beta \left(\frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{m} x_{ij}\beta_j \right)^2 \right) + \lambda \sum_{j=1}^{m} |\beta_j|$. Where β represents the feature coefficients, λ is the regularization parameter, n and m are number of samples and features, respectively. By setting some coefficients to zero, LASSO-CV effectively identifies and selects important features, thereby reducing the dimension-

ality of the dataset while preserving predictive accuracy [73]. Features with non-zero coefficients were retained and passed to the downstream SVM classifier, enabling dimensionality reduction while improving interpretability and generalization [73].

Subsequently, we utilized a support vector machine model [122] for predicting chemotherapy treatment responses, with hyperparameter tuning conducted via grid search cross validation [123]. The main goal of the SVM classifier is to find an optimal hyperplane in high-dimensional space to separate different classes in the feature space [80]. The hyperplane tries to maximize the margin between the closest samples of different classes. Hence, the optimization problem aims to minimize the $\frac{1}{2}w^T w + C\sum_{i=1}^{m}\zeta^{(i)}$ subject to $t^{(i)}\left(w^T x^{(i)} + b\right) \geq 1 - \zeta^{(i)}$ and $\zeta^{(i)} \geq 0$ for $i = 1, 2, \ldots, m$ [122]. w and b are normal vector and bias term, ζ denotes the slack variable, and C is a hyperparameter controlling the margin and adjusting coefficients for error optimization. Hyperparameters such as 'C', and 'kernel' were optimized through an exhaustive search. Specifically, 'C' values ranged from 0.1 to 10, and 'kernel' options included 'linear', radial basis function' (RBF), and 'polynomial'. For non-linear kernels we additionally tuned γ for RBF and the polynomial degree. The best configuration in each inner split was refit on the inner-training data and evaluated on the corresponding held-out fold.

Through model development and evaluation, we used a nested stratified cross-validation framework. The inner loop was repeated five times with stratified splits to reduce variability from random fold assignments and improve the stability of feature selection and hyperparameter tuning. Within each inner loop, feature preprocessing, SMOTE, feature selection, hyperparameter optimization, and SVM training were performed. The outer 10-fold loop was used exclusively for held-out testing to provide unbiased performance estimates. To enable probabilistic evaluation, SVM decision values were converted to class probabilities using Platt scaling (sigmoid calibration) by

enabling *probability=True* in scikit-learn's SVC. Probability estimation was learned using training data only within the nested cross-validation procedure. Calibration curves and Brier scores were computed exclusively from predicted probabilities on the held-out outer test folds. No probability threshold optimization was performed on test folds; class labels were generated using the default SVM decision rule.

We evaluated the performance of the models using the ROC curve [84], the ROC-AUC [124], and their respective confidence intervals. Additional metrics such as precision, recall, and F1-score were also assessed [125]. Confidence intervals were computed via patient-level bootstrap (N=200) on pooled out-of-fold predictions [126].

5.3 Results

Figure 5.4(a) and (b) present four sample CT images, representing both responder and non-responder cohorts. The segmented tumors in each image are highlighted in orange. Visual inspection of these images reveals no discernible features that differentiate between the responder and non-responder groups.

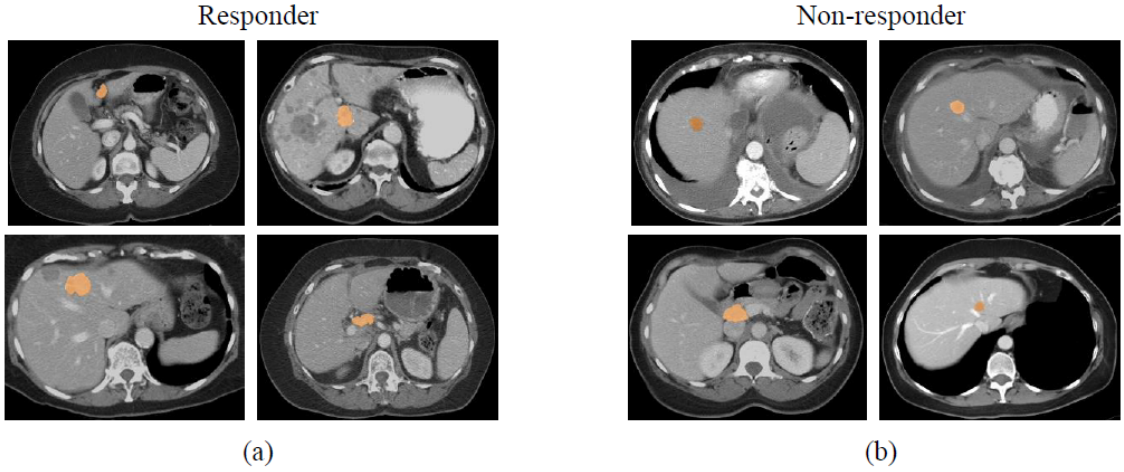


Figure 5.4: Sample CT images illustrate tumor segmentation. (a) Four images from the responder cohort, and (b) four images from the non-responder cohort.

Figure 5.5(a) displays heatmaps of the Pearson correlation coefficients among all extracted features from the radiomics, ViT-based, and MedSAM-based groups. The coefficients range from 0 to 1, with color intensity indicating the strength of correlation. As shown, while all feature groups demonstrate generally low internal dependency, the radiomics features exhibit higher intra-group correlations compared to those derived from ViT and MedSAM, as shown in fig. 5.5(b)-(d). The inter-group correlations remain relatively low, suggesting that the different feature sets provide complementary information with minimal redundancy.

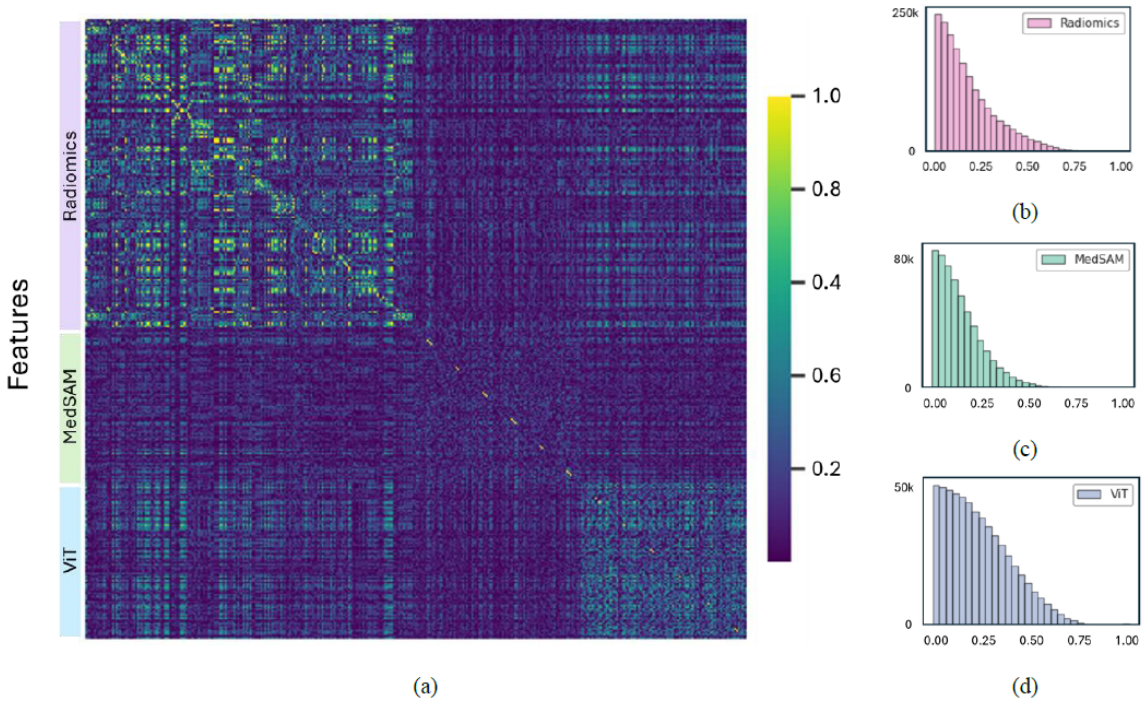


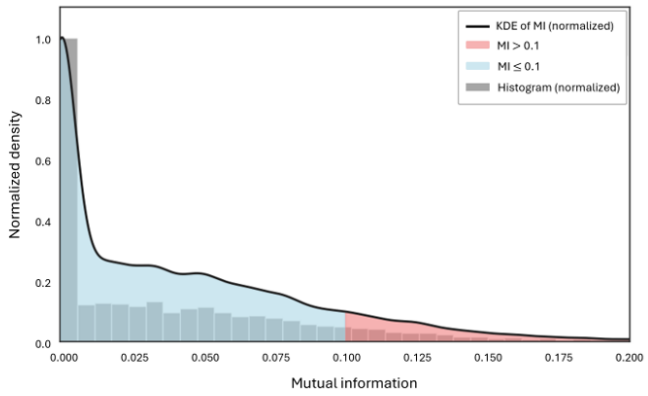
Figure 5.5: (a) Heatmap of Pearson correlation coefficients among all extracted features from radiomics, ViT-based, and MedSAM-based groups. (b–d) Individual heatmaps showing intra-group correlations.

To further examine the relationships among the selected feature groups in the context of integrated model development, fig. 5.6 presents the normalized distribution of mutual information (MI) between features employed in training the models. An

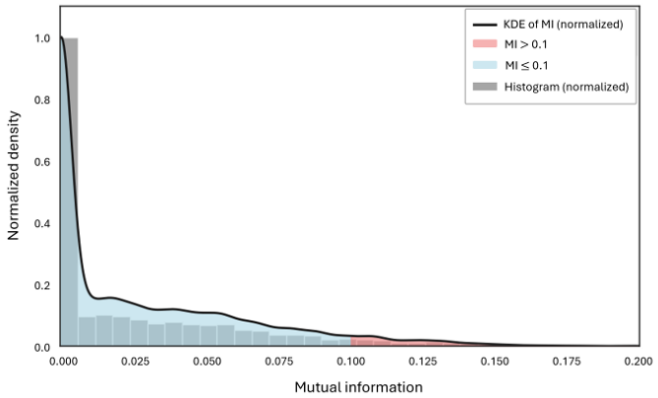
MI threshold of 0.1 is used to indicate very low statistical dependency. As shown in fig. 5.6(a), the distribution of MI values between the selected radiomic and ViT features reveals minimal redundancy, with over 90% of the values falling below the 0.1 threshold. Similar trends are observed in fig. 5.6(b) and (c), corresponding to the ViT–MedSAM and radiomics–MedSAM combinations, respectively, suggesting consistently low inter-feature dependency across all integrated feature sets. indicates minimal statistical dependency. (a) Radiomics and ViT features, (b) ViT and MedSAM features, and (c) Radiomics and MedSAM features.

Table 5.2 presents the performance of different feature sets and their combinations in predicting treatment response. Among individual modalities, MedSAM achieved the highest AUC (0.897 ± 0.034) and accuracy (0.808 ± 0.046), suggesting its effectiveness in capturing relevant imaging biomarkers. Radiomics and ViT, while lower performing individually, contributed to improved results when integrated. In the multi-representation setting, integrating all feature sets achieved the best performance (AUC 0.924 ± 0.037, accuracy 0.831 ± 0.042), indicating the complementary value of handcrafted radiomics and deep embeddings. The ViT + MedSAM model reached a similar AUC (0.924 ± 0.032) and second-highest in accuracy (0.829 ± 0.039), reinforcing the advantage of multi-representation integration. According to the radar chart in fig. 5.7, which compares models across multiple evaluation metrics, similar trends are observed in precision, recall, F1-score, and specificity. Detailed values for performance metrics of each model are reported in table B.1 (chapter B). These results demonstrate the closely matched performance of the (ViT + MedSAM) and (ViT + Radiomics + MedSAM) models, further supporting the effectiveness of combining ViT and MedSAM features.

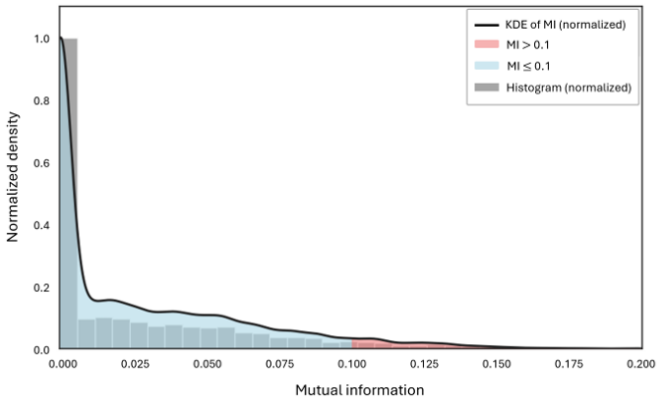
To assess the robustness of the reported performance to class imbalance handling, we additionally conducted a no-SMOTE ablation study for the best-performing integrated model (Radiomics + ViT + MedSAM). We used the same stratified nested



(a)



(b)



(c)

Figure 5.6: Normalized distributions of mutual information (MI) between feature pairs used in training multi-modal models. An MI threshold of 0.1 indicates minimal statistical dependency. (a) Radiomics and ViT features, (b) ViT and MedSAM features, and (c) Radiomics and MedSAM features.

Table 5.2: Prediction performance of the models with confidence interval.

Feature Set	AUC	Accuracy
Radiomics	0.780 ± 0.062	0.712 ± 0.044
ViT	0.679 ± 0.107	0.657 ± 0.056
MedSAM	0.897 ± 0.034	0.808 ± 0.046
Radiomics + ViT	0.865 ± 0.076	0.783 ± 0.045
Radiomics + MedSAM	0.890 ± 0.076	0.816 ± 0.050
ViT + MedSAM	$\mathbf{0.924 \pm 0.032}$	0.829 ± 0.039
Radiomics + ViT + MedSAM	$\mathbf{0.924 \pm 0.037}$	$\mathbf{0.831 \pm 0.042}$

cross-validation protocol with identical outer and inner folds, preprocessing, feature selection, and hyperparameter tuning. The full pipeline was repeated without applying SMOTE. As reported in table B.2 in chapter B, removal of SMOTE resulted in a modest reduction in performance (AUC 0.924 ± 0.037 with SMOTE vs. 0.908 ± 0.035 without SMOTE). Importantly, overall discriminative performance remained high, indicating that the observed performance is not driven by oversampling but reflects a robust predictive signal.

Calibration analysis was conducted for the best-performing model. The calibration curve (fig. 5.8) demonstrates close alignment between predicted and observed event probabilities, indicating good probability reliability. The Brier score was 0.12, reflecting good overall calibration. Figure 5.9 presents the performance of our best-performing model. Figure 5.9(a) and (b) show the ROC and precision–recall (PR) curves with ROC-AUC of 0.92 and PR-AUC of 0.94. Figure 5.9(c) depicts the confusion matrices for that model, with a decision threshold set at 0.43 on the decision scores (selected by max-F1 on the training folds). The model correctly predicts 109 out of 124 responder cases and 46 out of 58 non-responder cases, resulting in a positive predictive value (PPV) of 90% and a negative predictive value (NPV) of 75%.

After LASSO feature selection, an average of approximately 30 features were retained per run. Figure 5.10 (a)-(c) present SHAP [127] summary plots for the prognos-

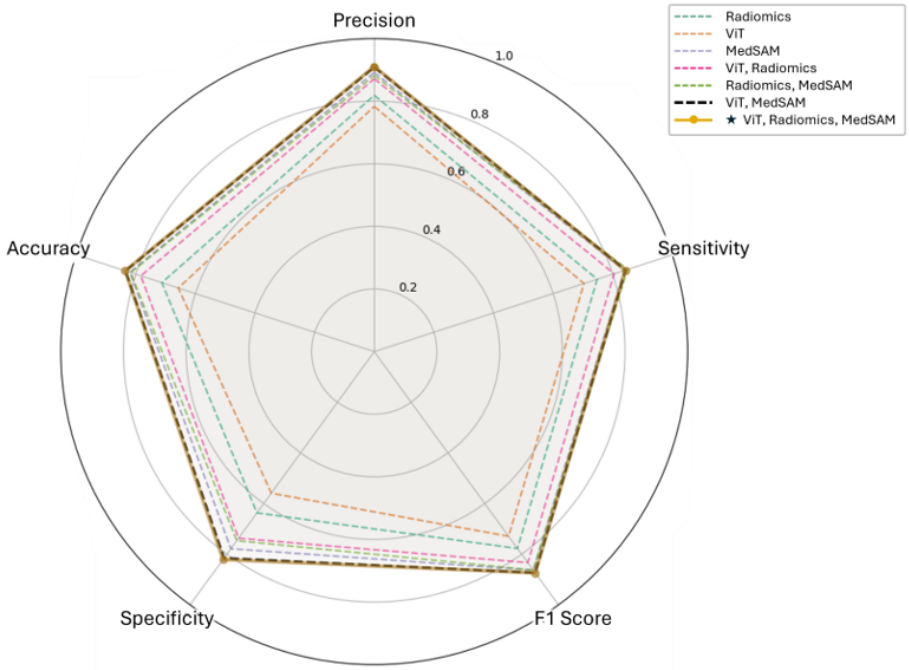


Figure 5.7: Radar chart comparing the performance of models trained on different feature sets across multiple evaluation metrics, including accuracy, precision, sensitivity, F1-score, and specificity.

tic feature subsets (MedSAM, ViT, radiomics) used by the best-performing classifier. For visual clarity, the figures illustrate the relative importance of the selected features without explicit feature name annotations. The 30 features shown were taken from a representative run whose performance closely matched the cross-validated average. Among them, 12 originated from MedSAM, 10 from ViT, and 8 from radiomics. Feature identities and detailed descriptions are provided in table B.3 in the chapter B. Across cross-validation runs, feature attributions were highly stable, with the top 20 features consistently appearing in at least 80% of the outer folds. To summarize this consistency, the mean absolute SHAP values for these features across runs are provided in fig. B.1 in the chapter B. This summary confirms that MedSAM-derived features exert the strongest influence on model predictions, while ViT-based and radiomic features also contribute meaningfully, though to a less extent.

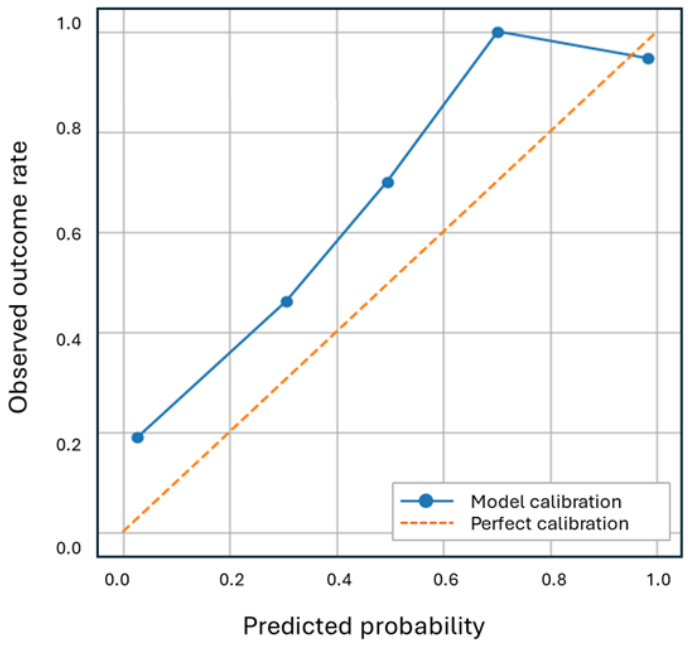
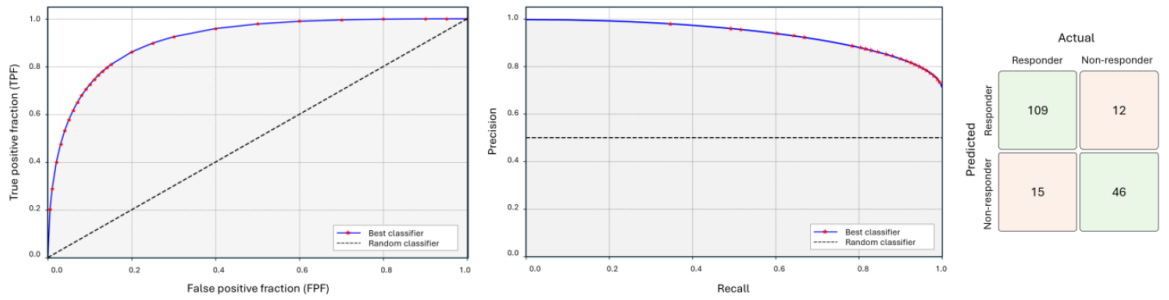


Figure 5.8: Calibration curve of the best-performing integrated model. The close agreement between predicted and observed probabilities indicates good calibration performance.

5.4 Discussion

In this study, we investigated the value of multi-method imaging biomarkers for predicting chemotherapy response in ovarian cancer patients. Compared to previous research, our study has several unique characteristics.

First, radiomics features outperformed ViT features in this task, achieving an AUC of 0.780 $\pm$ 0.062 compared to 0.679 $\pm$ 0.107. This is probably why handcrafted radiomics features have the potential to identify and extract tumor-related information clearly, and they have been proven to have potentials in characterization of tumor phenotypes and prediction of outcomes in oncologic imaging [99]. In contrast, although ViT models learn high-dimensional representations, they are pretrained primarily on non-medical datasets and therefore lack domain-specific medical knowledge. Nevertheless, their performance indicates the generalizability of transformer-based architec-



(a) (b) (c)

Figure 5.9: Performance of the six-month PFS prediction model achieved by the best performing classification model based on Radiomics, ViT and MedSAM feature integration. (a) ROC curve. (b) Precision-recall curve. (c) Confusion matrix.

tures in medical imaging applications. Prior CT radiomics studies in ovarian cancer have reported moderate-to-strong performance using handcrafted features and classical machine-learning pipelines [128–130]; in comparison, our results suggest that integrating transformer-based embeddings provides additional prognostic signal beyond traditional radiomics alone.

Second, the MedSAM based model showed better results than the radiomics and ViT based models (table 5.2), probably because it had been trained on extensive medical image datasets containing representations of clinical relevance. In particular, MedSAM by itself had a very good prediction performance, indicating that advanced foundation models already have a significant portion of the imaging signal that is typically captured by handcrafted radiomics encoded. Besides that, combining radiomics and ViT features led to an increased performance compared to each single-method that was used, which means that those features hold complementary information. Radiomics features provide interpretable, handcrafted descriptors of tumor intensity and texture, whereas ViT embeddings encode representations that capture more complex and subtle spatial patterns. The similar improvement in performance was shown through the combination of ViT and MedSAM features, which indicates that transformer models trained

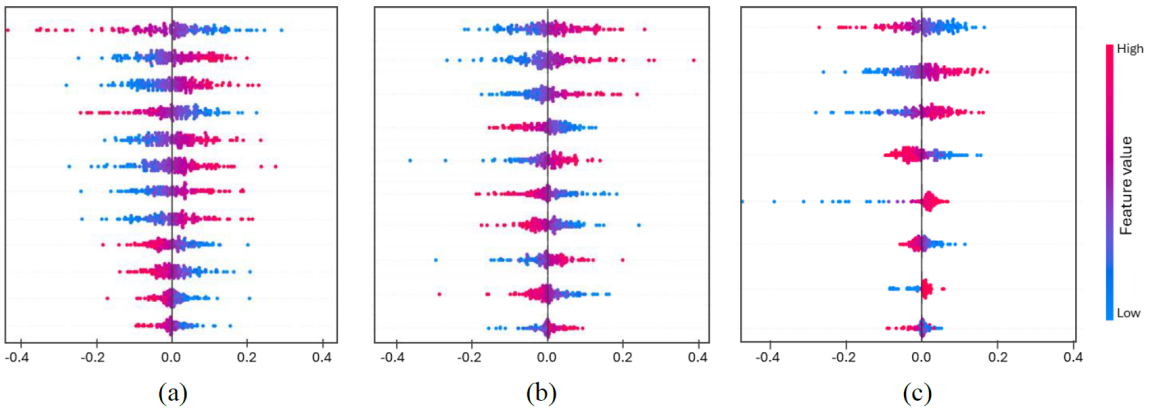


Figure 5.10: SHAP summary plots for the features selected in the best-performing model (Radiomics + ViT + MedSAM), shown separately by source: (a) MedSAM, (b) ViT, and (c) radiomics. To enhance visual clarity, feature names are not shown in the plots and are instead listed in Supplementary table B.3 in the chapter B.

in different domains (natural versus medical images) hold different, non-redundant information. Interestingly, adding radiomics features to the MedSAM-based model did not further improve performance, implying that MedSAM embeddings already encompass many structured tumor characteristics quantified by radiomics. In general, the results here suggest that multi-method fusion can capture a wider representation of tumor biology, which is reflected in the higher prediction performance of multi-method fusion models compared to single-method approaches. This conclusion is further supported by feature redundancy analyses, which revealed minimal correlation between ViT and MedSAM embeddings, consistent with their complementary nature.

Despite these promising results, several limitations should be acknowledged. First, the data used for this study was only from one institution, which may limit generalizability across scanners, acquisition protocols, and patient populations. To ensure the robustness and clinical transferability of the proposed approach, multi-center studies will be necessary in the future. Second, our feature fusion strategy relied on simple feature concatenation, which does not explicitly model interdependencies among representations; consequently, higher-dimensional or more discriminative feature sets may

dominate the learning process, potentially diminishing the contribution of more subtle or indirect predictive information such as radiomics. Additionally, imaging features were extracted from a single axial slice with the largest tumor cross-section. While 3D may capture richer spatial information, this 2D approach reduced sensitivity to slice-thickness variability and segmentation uncertainty while maintaining strong predictive performance. Finally, this study focused exclusively on CT imaging and did not incorporate additional data sources such as histopathology, clinical variables, or genomics, which could further enhance predictive performance and model robustness. Despite these limitations, our findings provide meaningful insights into imaging-based prediction of chemotherapy response and advance the development of precision oncology strategies.

Chapter 6

Multimodal Representation Fusion for Chemotherapy Response Prediction in Ovarian Cancer

6.1 Introduction

Ovarian cancer is a spatially heterogeneous disease characterized by extensive dissemination throughout the peritoneal cavity and the development of multiple metastatic lesions [131–134]. This heterogeneity extends beyond genetic and molecular differences and is reflected across distinct anatomical disease sites [132–135]. While the primary tumor represents the biological origin of the disease and preserves information regarding its intrinsic morphology, microenvironment, and evolutionary potential, metastatic lesions embody the outcome of tumor progression, dissemination, and adaptation to distant tissues [134–137]. Consequently, these disease sites may contain complementary information relevant to therapeutic response [135–138].

The standard treatment paradigm for advanced ovarian cancer further highlights the importance of considering both primary and metastatic disease [132, 133, 139]. Following primary cytoreductive surgery, treatment decisions and response assessments are primarily driven by the behavior of residual and metastatic lesions [9, 139, 140]. How-

ever, the biological mechanisms governing treatment sensitivity may originate from characteristics established within the primary tumor before dissemination [134, 135, 137]. Therefore, relying exclusively on either primary-tumor pathology or metastatic-tumor imaging may provide an incomplete representation of the disease processes associated with chemotherapy response [135, 137, 141].

Digital pathology and radiological imaging offer unique opportunities to interrogate these complementary aspects of ovarian cancer biology [15, 142–144]. Whole-slide histopathology images of the resected primary tumor provide microscopic information regarding tissue architecture, cellular morphology, stromal composition, and tumor microenvironmental interactions [142–144]. In contrast, computed tomography images of metastatic lesions capture macroscopic disease phenotypes, including lesion morphology, spatial heterogeneity, tumor burden, and patterns of dissemination [15, 58, 94]. Together, these modalities provide a multiscale characterization of the disease, spanning cellular-level tumor biology to organ-level manifestations of metastatic progression [15, 143, 145].

Despite increasing interest in multimodal artificial intelligence for oncology, most existing radiology-pathology studies analyze information derived from the same tumor site or focus on a single disease compartment [45, 146–148]. Consequently, the potential complementary value of pathology-derived representations from the primary tumor and radiology-derived representations from metastatic lesions remains largely unexplored [45, 147, 148]. Given that these disease sites reflect distinct stages of tumor evolution, integrating information from both may provide a more complete representation of the biological factors underlying chemotherapy sensitivity and resistance [135, 146, 149].

Recent advances in foundation models have created new opportunities to investigate this hypothesis [34, 150–152]. Foundation models trained on large-scale pathology

and medical imaging datasets can learn transferable and biologically meaningful visual representations that substantially outperform conventional handcrafted features and task-specific deep learning models [27, 34, 150–153]. In computational pathology, models such as CONCH have demonstrated the ability to capture rich histomorphologic patterns from whole-slide images [152, 153], while in radiology, MedSAM has shown strong generalization capabilities across diverse medical imaging tasks [27]. These models enable the extraction of high-level representations that may encode clinically relevant phenotypes across multiple imaging scales and modalities [34, 151, 152].

Therefore, rather than treating multimodal learning simply as a fusion of imaging modalities, we propose a biologically motivated framework that integrates information from two complementary disease compartments: the surgically resected primary tumor and radiographically visible metastatic lesions in CT images. Specifically, we utilize the CONCH foundation model to extract histopathological representations from primary-tumor whole-slide images and the MedSAM foundation model to derive imaging representations from metastatic lesions identified on CT. We hypothesize that the integration of microscopic biological information from the primary tumor with macroscopic phenotypic information from metastatic disease will provide complementary biomarkers of chemotherapy response and improve predictive performance beyond either modality alone.

6.2 Materials and Methods

6.2.1 Dataset

6.2.1.1 Patient Cohort and Study Design

We assembled a retrospective, single-institution dataset of 197 ovarian carcinoma patients from the university of oklahoma health sciences center who underwent primary cytoreductive (debulking) surgery, with a subset receiving neoadjuvant chemotherapy beforehand. Histological subtypes included high-grade serous, endometrioid, clear cell, and mucinous carcinoma. Eligibility required confirmed diagnosis via postoperative pathology, one of the four histological subtypes, paired pre- and post-chemotherapy CT scans with lesions annotated by a board-certified radiologist per RECIST criteria [9] (up to five target lesions per case), and matched H&E-stained whole-slide images (WSIs) from the primary debulked tumor. Cases were excluded for missing scans/WSIs or poor imaging quality (artifacts, noise, reconstruction errors, tissue impurities, uneven staining), as jointly assessed by the radiologists and pathologists. Of the 197 cases, 182 had paired pre- and post-chemotherapy CT scans; after applying all inclusion/exclusion criteria, 79 cases with both CT and WSI data formed the final analytic cohort used in all subsequent experiments. The entire study protocol was approved by the institutional review board in the University (IRB13649).

6.2.1.2 CT Imaging Acquisition and Preprocessing

CT scans were acquired at OUHSC per routine clinical protocol using a GE LightSpeed VCT (64-slice) or GE Discovery 600 (16-slice) scanner, with tube current adjusted for patient body habitus (100–600 mA). Both pre- and post-chemotherapy scans were collected for each case, and metastatic lesions were manually delineated by board-certified

radiologists per RECIST criteria [9], with up to five target lesions annotated per patient. CT volumes were resampled to isotropic voxel spacing to ensure consistent spatial resolution and improve feature reproducibility across patients, following recommendations from the image biomarker standardization initiative [42]. A soft-tissue HU window (WL = 40, WW = 400) was then applied to enhance lesion visibility. Windowed slices were linearly scaled to 8-bit grayscale and exported as PNG images without further normalization; these images were subsequently used for segmentation and feature extraction. Tumor regions of interest were segmented using a two-step method. First, an experienced radiologist identified the metastatic lesion slice containing the largest tumor area according to RECIST criteria [9]. A semi-automated hybrid algorithm, combining topographic region growing, adaptive thresholding, and dynamic edge tracking, was then used to delineate the tumor boundary.

6.2.1.3 Histopathology Whole Slide Images

Histopathological data consisted of 269 H&E-stained WSIs from the 79 eligible patients, extracted from primary debulked ovarian tumor specimens at OUHSC and digitized at 20× objective magnification using a clinical-grade slide scanner.

6.2.1.4 Clinical Data and Treatment Response Labels

The dataset also includes longitudinal clinical variables for each patient, including 6-month PFS and detailed chemotherapy treatment records. Binary treatment response labels were established from a composite assessment of pre- and post-treatment CT imaging and clinical evaluation. Within the 79-patient WSI subcohort, platinum-based chemotherapy was classified as effective in 58 cases and ineffective in 21, yielding a moderately imbalanced label distribution accounted for in our experimental design.

6.2.2 WSI Preprocessing and Feature Encoding

Whole-slide histopathology images were partitioned into non-overlapping 224 × 224 pixel tiles using the OpenSlide library [154]. Prior to analysis, a preprocessing pipeline was applied to remove background, out-of-focus regions, and artifact-containing tiles.

Background removal was performed using Otsu thresholding [155] on a downsampled grayscale representation of each slide, and only tiles containing at least 90% tissue were retained. Marker-contaminated and scanning-related artifacts were identified using intensity thresholding and morphological image processing techniques [156] and subsequently removed.

Image sharpness was quantified using a blur-based focus score computed as the mean absolute difference between each tile and its Gaussian-blurred counterpart, a commonly used focus measure for digital pathology image quality assessment [157]. Tiles with focus scores below a slide-specific threshold (one standard deviation below the mean) were classified as out of focus and excluded.

To enrich the dataset with tumor-representative regions, three pathology foundation models including CONCH [152], PLIP [158], and BioMedCLIP [159] were used to estimate the likelihood of each tile containing ovarian cancer tissue. For the CLIP-based models (PLIP and BioMedCLIP), tile embeddings were compared against textual prompts describing malignant and normal ovarian histopathology. The malignant prompts described high-grade serous ovarian carcinoma characterized by atypical epithelial cells, pleomorphic hyperchromatic nuclei, increased mitotic activity, papillary structures, stromal invasion, necrosis, and high cellularity, whereas the negative prompt described normal ovarian tissue with uniform epithelial cells and regular nuclei. Each foundation model produced a cancer probability score for every tile, and tiles were ranked according to their predicted likelihood of malignancy.

The final set of tumor-representative tiles was obtained by selecting the top 300 tiles that were consistently ranked among the most cancerous regions across all three foundation models. These selected tiles were subsequently encoded using the CONCH foundation model [152] to produce a 512-dimensional feature vector for each tile. Figure 6.1 demonstrates the overview of tile filtering workflow used in our study.

Tile-level embeddings were aggregated using mean and maximum pooling across all selected tiles to generate a slide-level representation that captured both dominant histopathologic characteristics and intra-slide heterogeneity. For patients with multiple WSIs from primary tumors, slide-level feature vectors were further aggregated using mean, standard deviation, and maximum pooling to generate a single patient-level pathology representation with 3072 elements for downstream multimodal analysis and treatment outcome prediction.

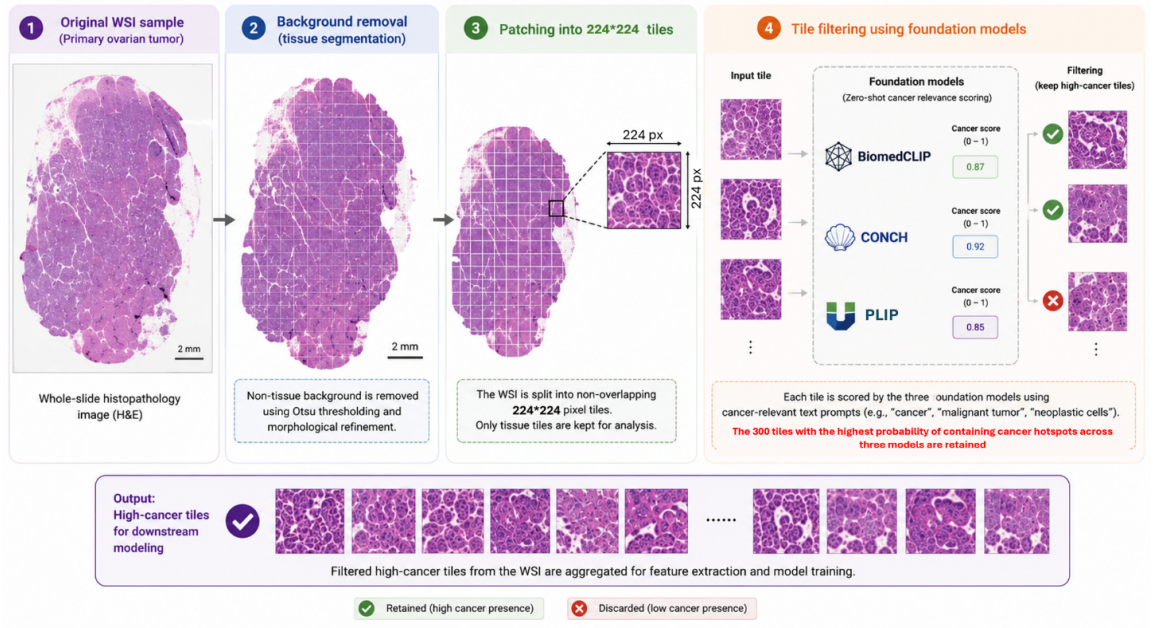


Figure 6.1: Overview of the tile filtration workflow for cancerous hotspot selection.

6.2.3 CT Feature Encoding

Deep imaging features were extracted from segmented metastatic lesions using the frozen image encoder of MedSAM [27], a medical-domain adaptation of the segment anything model [160] pretrained on more than 1.5 million medical image-mask pairs.

For each lesion, the axial slice containing the largest tumor area was selected for feature extraction. The grayscale image was converted to a three-channel image by replicating the intensity values across RGB channels and resized to 1024×1024 pixels to match the MedSAM input requirements.

The MedSAM image encoder, based on the vision transformer (ViT-B) architecture [161], partitioned each image into 4096 non-overlapping patches and generated a 768-dimensional embedding for each patch.

To focus feature extraction on tumor tissue, the lesion segmentation mask was mapped onto the corresponding patch grid, and only patches containing at least 50% tumor tissue were retained. The retained embeddings were averaged to produce a lesion-level representation. Patient-level CT representations were subsequently generated by averaging lesion-level embeddings across all metastatic lesions.

6.2.4 Feature Selection

Feature selection was incorporated within a nested cross-validation framework to prevent information leakage and ensure unbiased performance estimation [162]. For each outer cross-validation fold, recursive feature elimination (RFE) [163] was applied exclusively to the training portion of the data. Prior to feature selection, features were standardized using z-score normalization, with normalization parameters estimated from the training data and subsequently applied to the corresponding validation and test sets.

RFE was implemented using a logistic regression estimator with L2 regularization and class-balanced weighting [164]. Features were iteratively removed according to their contribution to the model until a predefined number of features remained. The optimal number of retained features (k) was treated as a hyperparameter and determined within the inner cross-validation loop. Candidate values of k ranged from 2 to 16 features. For each inner training fold, RFE was performed independently, and the feature subset yielding the highest mean validation AUC was selected.

Following hyperparameter optimization, RFE was repeated on the entire outer training set using the optimal k value to obtain the final feature subset for that fold. The selected features were then applied to the corresponding held-out outer test set, which remained completely unseen during feature selection and model optimization. This procedure was repeated across all outer folds, and feature selection stability was assessed by tracking the frequency with which individual features were selected across folds.

6.2.5 Multimodal Fusion Strategies and Model Development

To evaluate the contribution of imaging and histology modalities to treatment-response prediction, we compared two unimodal baselines with three multimodal fusion strategies representing early fusion, feature attention, and cross-modal attention, which are commonly used paradigms in multimodal machine learning [165] (fig. 6.2). All strategies share a common backbone consisting of modality-specific feature extraction (MedSAM for CT and CONCH for WSI), followed by dimensionality reduction using recursive feature elimination and a logistic regression classifier for the final prediction of the response probability, $P(\text{responder})$.

6.2.5.1 Unimodal baselines

As reference points, we trained independent logistic regression classifiers on each modality in isolation. CT features were extracted as 768-dimensional embeddings using MedSAM, whereas WSI features were extracted as 3072-dimensional embeddings using CONCH. Each feature set was independently reduced via RFE to its top-K most predictive features, yielding

$z_{CT} \in \mathbb{R}^{K_{CT}}$ or $z_{WSI} \in \mathbb{R}^{K_{WSI}}$. The resulting feature vector z was then used as input to the logistic regression classifier, which estimates the probability of treatment response according to

$$P(\text{responder}) = \sigma\left(\beta^\top z + \beta_0\right). \tag{6.1}$$

These unimodal models establish the baseline predictive performance achievable from imaging or histology alone, against which multimodal gains can be measured.

6.2.5.2 Early fusion with joint feature selection

In the early-fusion strategy, CT and WSI embeddings are first concatenated at the raw feature level before dimensionality reduction. The concatenated feature representation is given by

$$z_{\text{concat}} = [z_{CT}^{\text{raw}}; z_{WSI}^{\text{raw}}] \in \mathbb{R}^{768+512}. \tag{6.2}$$

RFE is then applied jointly across the combined feature space to select the top-K features regardless of their modality of origin, producing a single reduced feature vector that is subsequently passed to the logistic regression classifier. Unlike the modality-specific RFE used in later strategies, this joint selection allows features from either modality to compete directly for inclusion but does not explicitly model interactions

or relative weighting between modalities. Instead, the fusion occurs solely through feature concatenation before feature selection.

6.2.5.3 Feature attention fusion

The feature-attention strategy first performs RFE independently within each modality, producing reduced feature vectors $z_{CT} \in \mathbb{R}^{K_{CT}}$ and $z_{WSI} \in \mathbb{R}^{K_{WSI}}$. The two vectors are then concatenated to form the multimodal representation

$$z = [z_{CT}; z_{WSI}] \in \mathbb{R}^{K_{CT}+K_{WSI}}. \tag{6.3}$$

A logistic regression model is then trained using the concatenated feature vector in eq. (6.3). The learned regression coefficients β_i are subsequently used to derive feature-level attention weights according to

$$w_i = \text{softmax}(|\beta_i|) = \frac{e^{|\beta_i|}}{\sum_{j=1}^{d} e^{|\beta_j|}}, \qquad \widetilde{z}_i = w_i z_i. \tag{6.4}$$

As shown in eq. (6.4), the attention weights are computed directly from the magnitude of the learned logistic regression coefficients, resulting in a post-hoc, model-derived attention mechanism. This weighting highlights the relative importance of individual features from either modality but does not explicitly model cross-modal interactions, since each feature is weighted independently.

6.2.5.4 Cross-modal attention with modality gating

To explicitly model interactions between CT and WSI representations, we implemented a learned cross-modal attention and gating mechanism. The approach follows the transformer attention principle introduced by Vaswani et al. [166], while adapting it to scalar modality-level gating for multimodal medical imaging.

Given modality-specific, RFE-reduced feature vectors projected to a common dimensionality, $z_{CT}, z_{WSI} \in \mathbb{R}^k$, query and key representations are first computed using learned projection matrices $W_q, W_k \in \mathbb{R}^{k \times k}$:

$$q = W_q z_{CT}, \qquad \kappa = W_k z_{WSI}. \tag{6.5}$$

Using the query and key vectors defined in eq. (6.5), a scalar attention coefficient is computed using scaled dot-product attention followed by a sigmoid activation,

$$a = \sigma\left(\frac{q^\top \kappa}{\sqrt{d}}\right). \tag{6.6}$$

The attention coefficient in eq. (6.6) determines how much of the WSI representation is retained relative to the CT representation, yielding the attended WSI embedding

$$z^*_{WSI} = a\, z_{WSI} + (1 - a)\, z_{CT}. \tag{6.7}$$

A second modality-level gate, parameterized by a learned vector v, is then used to determine the final contribution of each modality. The gating function and fused representation are computed as

$$g = \sigma\left(v^\top [z_{CT}; z^*_{WSI}]\right), \qquad z = g\, z_{CT} + (1 - g)\, z^*_{WSI}. \tag{6.8}$$

Finally, the fused feature vector defined in eq. (6.8) is passed to the logistic regression classifier for treatment-response prediction. This formulation allows the model to adaptively weigh each modality on a per-sample basis by conditioning fusion on the degree of agreement between CT and WSI representations, in contrast to the static, post-hoc weighting used in the feature-attention strategy.

For all fusion strategies, predictive models were implemented using logistic regres-

sion with L2 regularization and class-balanced weighting to account for class imbalance. Model hyperparameters, including the regularization strength (C) and the number of selected features (k), were jointly optimized within the inner cross-validation loop using a grid-search strategy. Candidate values of C included 0.01, 0.1, and 1.0, whereas k was selected from the even values between 2 and 16. Following model training, an optimal classification threshold was determined on the training data by maximizing the F1-score and subsequently applied to the corresponding held-out test fold.

6.2.6 Model Evaluation and Statistical Analysis

Model development and evaluation were performed using repeated nested stratified cross-validation [162]. The outer cross-validation consisted of five stratified folds and was repeated 100 times using different random seeds to assess the stability and robustness of model performance across different data partitions. Within each outer training set, a five-fold stratified inner cross-validation was performed exclusively for hyperparameter optimization and model selection, ensuring that the outer test sets remained completely independent throughout model development.

Model performance was evaluated using the area under the receiver operating characteristic curve, area under the precision–recall curve, sensitivity, specificity, positive predictive value, and F1-score. During each repetition, every patient received one out-of-fold (OOF) prediction [167]. Performance metrics were first computed from the pooled OOF predictions of each complete repetition, and the mean ± standard deviation across the ten repetitions was reported to assess the stability of model performance. The OOF predicted probabilities for each patient were then averaged across all repetitions to obtain a single patient-level prediction.

Statistical comparisons between competing models were performed using the DeLong test for correlated ROC curves [168]. All reported performance estimates and

statistical analyses were derived exclusively from held-out test predictions generated within the outer cross-validation framework, with no information from the outer test sets used during model selection or hyperparameter optimization.

6.3 Results

6.3.1 Correlation Analysis of Imaging Features

Figure 6.3 presents the intra- and inter-modality correlations among the extracted feature groups. fig. 6.3a shows the distribution of the absolute Pearson correlation coefficients within each feature group. Overall, most feature groups exhibited relatively weak internal correlations, indicating limited redundancy among the extracted descriptors. In contrast, the WSI feature group demonstrated moderately stronger intra-group correlations, which may be attributed to the inclusion of statistical summary features (e.g., standard deviation) computed across multiple slides from the same patient.

Figure 6.3b presents the Pearson correlation matrix between the patient-level CT and WSI feature sets. Correlation coefficients range from -1 to 1, where warmer colors represent stronger positive correlations and cooler colors indicate stronger negative correlations. As shown, correlations between CT- and WSI-derived features were generally weak, suggesting that the two imaging modalities capture complementary information with minimal redundancy. These findings support the rationale for multimodal feature fusion by indicating that each modality provides distinct yet potentially complementary representations of tumor characteristics.

6.3.2 Model Complexity

The computational complexity of the evaluated logistic regression models is summarized in table 6.1. Model complexity primarily depended on the number of features retained after recursive feature elimination. Among the single-modality models, CT-only and WSI-only classifiers exhibited comparable complexity. Within the multimodal approaches, Early Fusion and Feature Attention produced the most compact models, followed by Late Fusion, whereas the Cross-Modal Attention architecture resulted in the highest model complexity due to its additional attention mechanisms.

Table 6.1: Model complexity comparison for single-modality and fusion models.

Model	**Modalities**	**Fusion Type**	**LogReg Parameters**
CT-only	CT-only	None	$k_{\mathrm{CT}} + 1$
WSI-only	WSI-only	None	$k_{\mathrm{WSI}} + 1$
Early fusion	CT + WSI	Joint RFE	$(k_{\mathrm{CT}} + 1) + (k_{\mathrm{WSI}} + 1) + 2$
Late fusion	CT + WSI	Learned weights	$k_{\mathrm{CT}} + k_{\mathrm{WSI}} + 1$
Feature attn.	CT + WSI	Feature softmax	$k_{\mathrm{CT}} + k_{\mathrm{WSI}} + 1$
Cross-modal attn.	CT + WSI	CT queries WSI	$k_{\mathrm{CT}} + 2k^2 + 2k + 1$

6.3.3 Predictive Performance

Table 6.2 summarizes the predictive performance of all evaluated models using repeated nested stratified cross-validation (100 repeats $\times$ 5 outer folds). Among the single-modality models, the CT-only model outperformed the WSI-only model, achieving a mean cross-validation AUC of 0.791 $\pm$ 0.027 compared with 0.755 $\pm$ 0.027. The CT-only model also demonstrated higher specificity (0.449 $\pm$ 0.055 vs. 0.319 $\pm$ 0.090), whereas the WSI-only model achieved higher sensitivity (0.943 $\pm$ 0.035 vs. 0.893 $\pm$ 0.040) and a slightly higher F1-score (0.861 $\pm$ 0.025 vs. 0.850 $\pm$ 0.018).

Overall, multimodal fusion strategies consistently improved predictive performance

over the single-modality models. The Feature Attention model achieved the highest mean cross-validation AUC (0.853 ± 0.056) and the highest specificity (0.671 ± 0.164), while also achieving the highest mean AUPRC (0.936 ± 0.038), comparable to that of the Early Fusion model (0.936 ± 0.020). The Early Fusion model demonstrated similar discriminative performance, achieving a mean AUC of 0.850 ± 0.030, while yielding the highest F1-score (0.875 ± 0.023) among all evaluated models. The Late Fusion model also improved upon both single-modality approaches, achieving a mean AUC of 0.835 ± 0.042. In contrast, the cross-modal attention model exhibited the weakest performance among the multimodal approaches, with a mean AUC of 0.747 ± 0.034 and specificity of 0.305 ± 0.072, despite maintaining a relatively high sensitivity (0.931 ± 0.037).

Table 6.2: Prediction performance of the models with confidence intervals.

Model	**AUC**	**AUPRC**	**Sensitivity**	**Specificity**	**F1**	***p*-value**
CT-only	0.791 ± 0.027	0.893 ± 0.018	0.893 ± 0.040	0.449 ± 0.055	0.850 ± 0.018	n.s.
WSI-only	0.755 ± 0.027	0.902 ± 0.021	0.943 ± 0.035	0.319 ± 0.090	0.861 ± 0.025	n.s.
Early fusion	0.850 ± 0.030	0.936 ± 0.020	0.881 ± 0.042	0.638 ± 0.082	0.875 ± 0.023	$p < 0.05$
Late fusion	0.835 ± 0.042	0.933 ± 0.019	0.878 ± 0.051	0.529 ± 0.163	0.857 ± 0.025	n.s.
Feature attn.	0.853 ± 0.056	0.936 ± 0.038	0.853 ± 0.034	0.671 ± 0.164	0.866 ± 0.016	$p < 0.05$
Cross-modal attn.	0.747 ± 0.034	0.881 ± 0.023	0.931 ± 0.037	0.305 ± 0.072	0.853 ± 0.017	n.s.

$^{*}p < 0.05$ compared with the CT-only model.

To facilitate visual comparison across models, fig. 6.4 presents a heatmap of the principal performance metrics. Higher values are represented by warmer colors, highlighting the superior overall performance of the Feature Attention and Early Fusion models across multiple evaluation criteria.

6.3.4 Receiver Operating Characteristic and Precision–Recall Analysis

Receiver operating characteristic (ROC) and precision–recall (PR) curves for all evaluated models are presented in fig. 6.5a, b. Consistent with the quantitative results, the Feature Attention model achieved the largest ROC-AUC and PR-AUC, followed by the Early Fusion model. The remaining multimodal architectures showed only modest improvements over the single-modality baselines.

6.3.5 Confusion Matrix Analysis

Confusion matrices for all evaluated models are summarized in table 6.3. Classification thresholds were determined independently for each model using the maximum F1-score obtained from the training folds.

The Feature Attention model achieved the highest overall classification accuracy (83.5%) using a decision threshold of 0.499. The model correctly identified 50 of 58 responders and 16 of 21 non-responders, corresponding to a sensitivity of 86.2% and a specificity of 76.1%, representing the best balance between correctly identifying responders and minimizing false-positive predictions. In comparison, although the Cross-Modal Attention model achieved the highest sensitivity (91.3%), this was accompanied by substantially reduced specificity (38.0%), indicating an increased false-positive rate.

6.3.6 Feature Stability Analysis

Stable features identified during repeated nested cross-validation are summarized in fig. 6.6. Only features selected in more than 60% of the outer folds and repetitions were included in the analysis. The figure reports the eight most frequently selected features for the best-performing multimodal model (Feature Attention). Several highly ranked

Table 6.3: Confusion matrices for single-modality and fusion models.

Model	Threshold	TN	FP	FN	TP	Accuracy	Sensitivity	Specificity	F1
CT-only	0.384	8	13	3	55	0.797	0.948	0.381	0.873
WSI-only	0.382	7	14	5	53	0.759	0.913	0.333	0.848
Early fusion	0.484	14	7	9	49	0.810	0.845	0.666	0.859
Late fusion	0.696	9	12	7	51	0.759	0.845	0.429	0.843
Feature attn.	0.500	15	6	6	52	0.848	0.897	0.714	0.896
Cross-modal attn.	0.355	6	15	3	55	0.772	0.948	0.285	0.859

features appeared consistently in both the single-modality and multimodal models, suggesting that these imaging biomarkers contribute robustly to treatment response prediction regardless of the fusion strategy.

6.3.7 Relative Contribution of CT and WSI Features

The relative contributions of the CT and WSI modalities within the fusion models are illustrated in fig. 6.7. In the best-performing feature-attention model, the learned fusion weights assigned greater importance to CT-derived features (0.556 ± 0.075) than to WSI-derived features (0.444 ± 0.075). This pattern was consistent across most outer folds, suggesting that CT features had a stronger influence on treatment-response prediction, while WSI features provided complementary information that further improved model performance. A similar trend was observed for the late-fusion attention model. In contrast, the cross-modal attention model assigned greater importance to WSI features.

6.4 Discussion

In this study, we investigated the value of combining CT-derived and histopathology-derived imaging biomarkers for predicting chemotherapy response in ovarian cancer patients. Compared with the single-modality models, multimodal fusion consistently

improved predictive performance, demonstrating that the two imaging modalities provide complementary information for treatment outcome prediction [146].

First, among the single-modality models, the CT-based classifier outperformed the WSI-based classifier, achieving a higher cross-validation AUC (0.791 ± 0.027 vs. 0.755 ± 0.027). A possible explanation is that CT images were acquired from metastatic tumors, whereas the histopathology slides originated from the primary tumor. Since chemotherapy response is strongly influenced by the biological behavior, tumor burden, and aggressiveness of metastatic disease [94, 132], imaging characteristics extracted directly from metastatic lesions may provide more prognostic information regarding treatment resistance [58, 94]. In contrast, although primary tumor histopathology contains valuable information regarding tumor morphology, tissue architecture, and the tumor microenvironment [142, 143], it may not fully represent the biological characteristics acquired during metastatic dissemination and evolution [134–137] that ultimately determine therapeutic response.

Second, all multimodal fusion strategies, except the cross-modal attention model, improved predictive performance over both single-modality models. These findings are further supported by the weak inter-modality correlations observed between CT and WSI features, suggesting that the two modalities capture distinct aspects of tumor biology with minimal redundancy. Consequently, integrating CT and WSI features allows the prediction model to exploit complementary information that is unavailable from either modality alone. Similar observations have been reported in previous multimodal imaging studies, where combining radiological and histopathological information improved prediction performance by capturing complementary tumor characteristics [146]. Among all evaluated approaches, the feature attention model achieved the highest overall performance, obtaining the best cross-validation AUC (0.853 ± 0.056), the highest specificity (0.671 ± 0.164), and one of the highest AUPRC values.

The modality contribution analysis further supports these findings. In the feature attention model, the learned attention weights consistently assigned greater importance to CT-derived features (0.556 ± 0.075) than to WSI-derived features (0.444 ± 0.075). Similar trends were observed in the late fusion model. This result is consistent with the superior performance of the CT-only classifier and suggests that CT provides the dominant predictive signal for chemotherapy response. This is likely because CT directly characterizes metastatic disease burden, lesion morphology, intratumoral heterogeneity, and imaging manifestations associated with treatment resistance [15, 58, 94]. Nevertheless, the fact that all successful fusion strategies outperformed the CT-only model indicates that WSI features provide complementary biological information that is not captured by radiological imaging alone. Histopathology contains microscopic information regarding tumor architecture, cellular morphology, stromal composition, immune infiltration, and the tumor microenvironment [142, 144, 149], complementing the macroscopic anatomical information available from CT and ultimately improving prediction performance.

Third, the cross-modal attention model achieved the lowest performance among the multimodal approaches despite its theoretical ability to explicitly learn relationships between CT and WSI representations. Cross-attention mechanisms have demonstrated remarkable performance in large-scale vision-language and multimodal learning tasks by modeling interactions between heterogeneous data sources [166]. However, these architectures generally contain substantially more trainable parameters than simpler fusion strategies and typically require large training datasets to achieve robust generalization [161, 166]. Given the relatively limited number of patients available in this study, the increased model complexity likely exceeded the amount of information that could be reliably learned from the training data, increasing the risk of overfitting. Similar observations have been reported in medical imaging studies, where simpler fu-

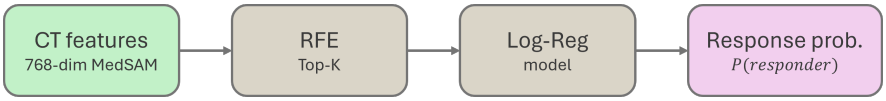
sion strategies frequently outperform more complex attention-based architectures on relatively small clinical datasets [146]. In contrast, the Feature Attention and Early Fusion approaches employ simpler fusion mechanisms that appear to be better suited for moderate-sized clinical datasets.

Finally, CONCH was selected as the foundation model for extracting WSI embeddings because previous studies have demonstrated its superior performance over several pathology foundation models in pathology representation learning, survival prediction, and clinical outcome prediction tasks [34, 45, 152]. These studies suggest that CONCH captures clinically relevant tissue representations beyond those learned by alternative pretrained models, making it an appropriate choice for chemotherapy response prediction in ovarian cancer. However, like other pathology foundation models, CONCH generates high-dimensional latent embeddings with limited biological interpretability [34, 151, 152]. Unlike handcrafted imaging biomarkers, these learned representations cannot be directly associated with specific histopathological characteristics, making clinical interpretation more challenging [169, 170].

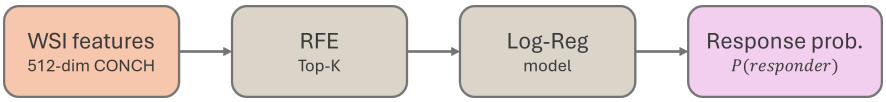
This study has several limitations. First, the relatively small cohort size limits the ability of complex deep learning architectures, particularly attention-based fusion methods with a large number of trainable parameters, to fully realize their potential [15, 58]. Second, the study was conducted using data collected from a single institution without independent external validation, which may limit the generalizability of the proposed models. Third, only a single pathology foundation model (CONCH) was used for WSI feature extraction, and only imaging information was incorporated into the prediction framework. Finally, although foundation model embeddings improved predictive performance, their limited interpretability makes it difficult to understand the biological mechanisms underlying the model predictions [169, 170].

Future work should focus on validating the proposed framework using larger multi-

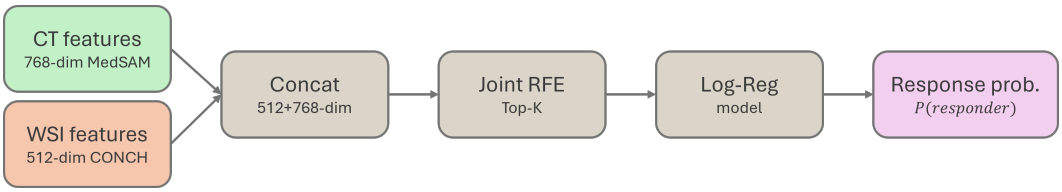
center cohorts to improve robustness and generalizability. With substantially larger datasets, more sophisticated cross-modal attention architectures may better exploit complex interactions between radiology and histopathology features [15, 58, 94]. In addition, integrating clinical variables, molecular biomarkers, genomic information, and recently developed pathology foundation models such as Virchow, UNI, and related vision-language pathology models [34] may further improve prediction accuracy and model robustness. Domain adaptation techniques [171, 172] could also be incorporated to reduce inter-institutional variability in both CT and WSI data, thereby improving generalization across imaging centers. Finally, developing more interpretable multimodal models through attention visualization, explainable AI techniques, and biologically meaningful embedding analysis [169, 170, 173] will be essential for increasing clinician confidence and facilitating translation into precision oncology applications.



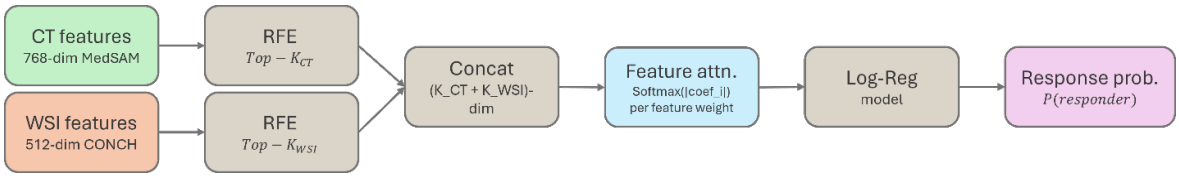
(a) CT-only baseline.



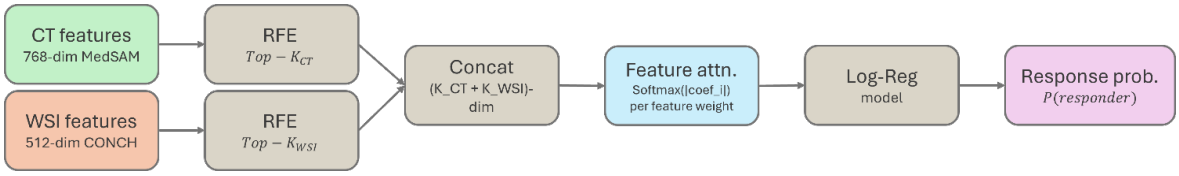
(b) WSI-only baseline.



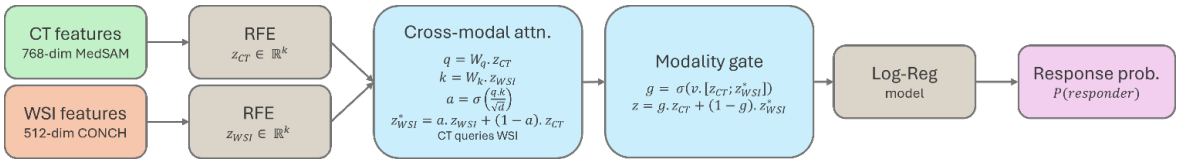
(c) Early fusion.



(d) Late fusion.

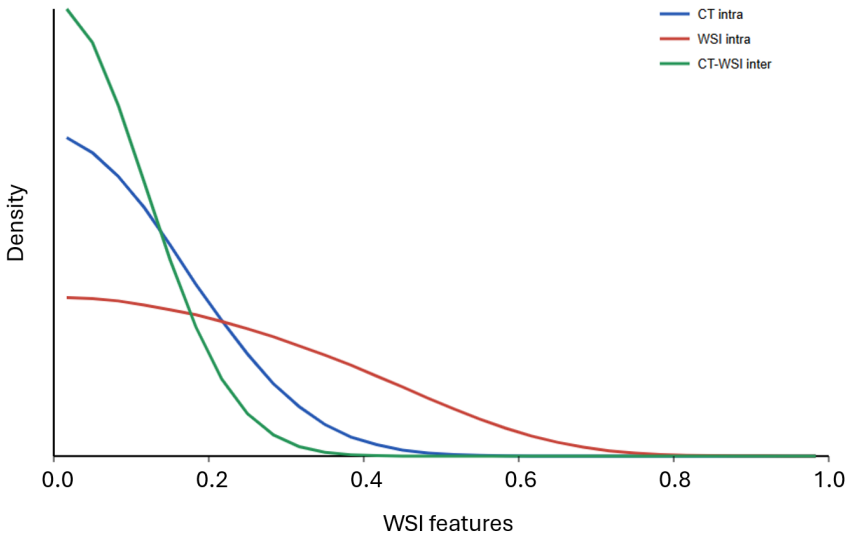


(e) Feature attention fusion.

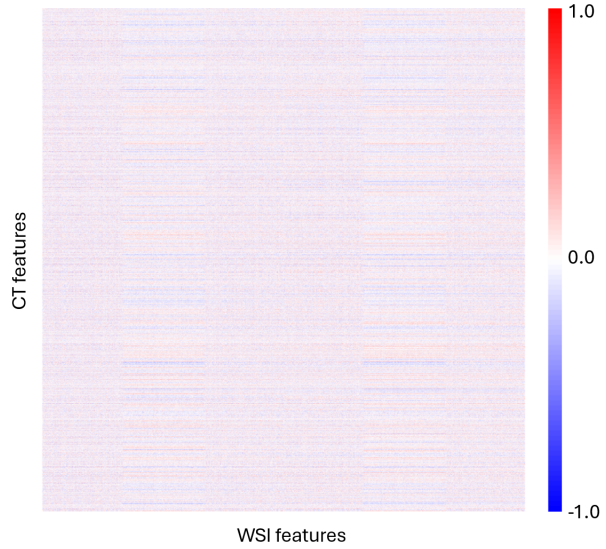


(f) Cross-modal attention fusion.

Figure 6.2: Structural block diagrams of the evaluated models: (a) CT-only baseline, (b) WSI-only baseline, (c) Early fusion, (d) Late fusion, (e) Feature attention fusion, and (f) Cross-modal attention fusion.



(a) Intra-group Pearson correlation distributions.



(b) Cross-modality CT–WSI correlation matrix.

Figure 6.3: Intra- and inter-modality feature correlations across the CT and WSI feature sets.

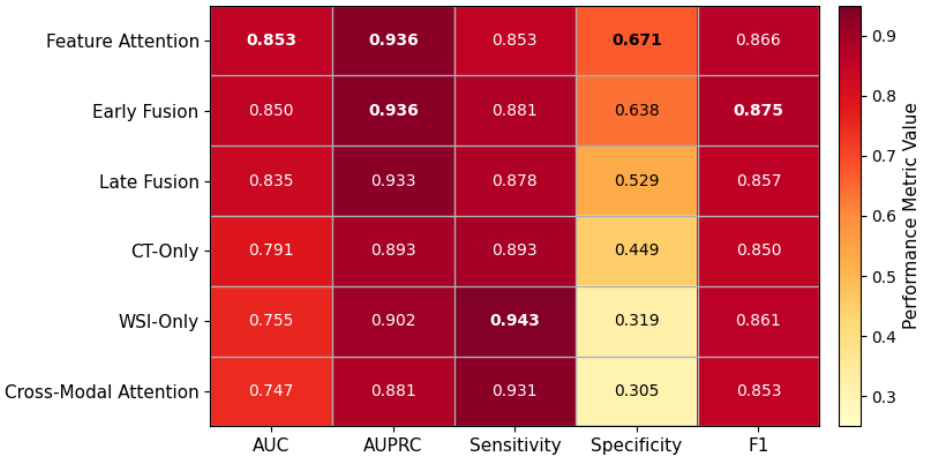
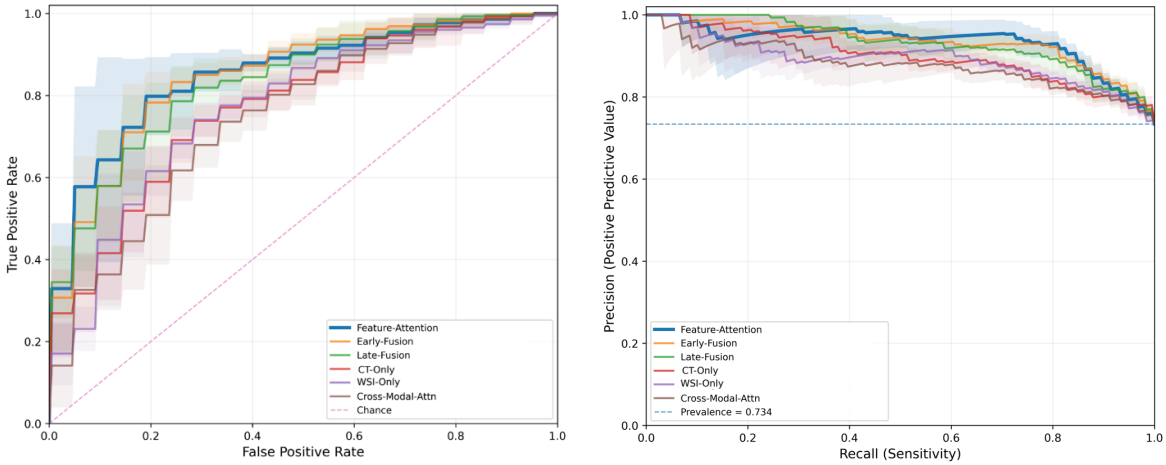


Figure 6.4: Heatmap of model performance metrics.



(a) ROC curves.

(b) Precision–recall curves.

Figure 6.5: Model performance curves. (a) Receiver operating characteristic (ROC) curves and (b) precision–recall (PR) curves for the single-modality and multimodal fusion models.

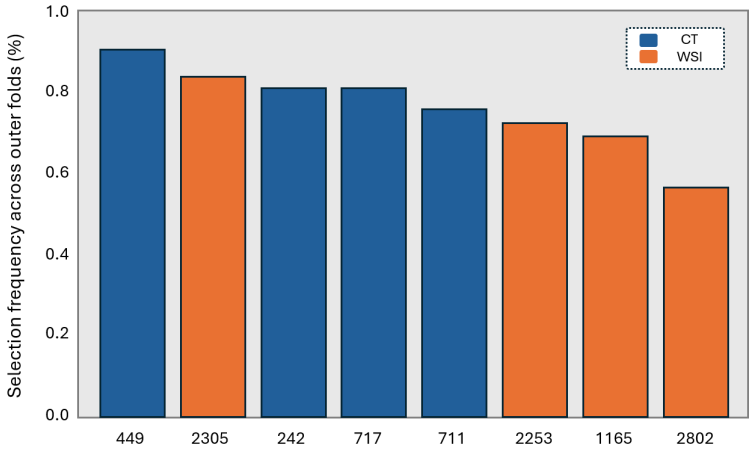


Figure 6.6: Stable predictive features for treatment response. Key features selected across cross-validation folds in the Feature Attention model, highlighting robust biomarkers across fusion strategies.

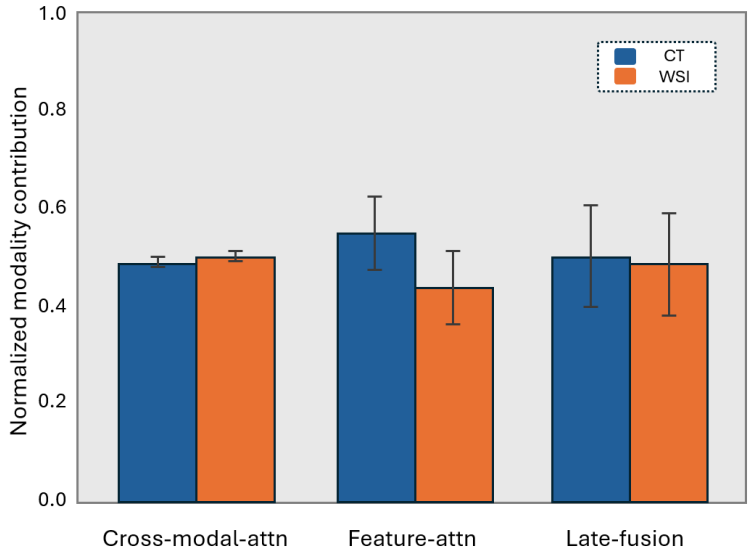


Figure 6.7: Relative modality contributions across fusion models. Distribution of learned fusion weights assigned to CT and WSI features.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

The main focus of this dissertation was the formulation of robust, generalizable, and biologically significant imaging biomarkers that would allow the early prediction of chemotherapy response in ovarian cancer. The accurate and timely treatment stratification in clinical practice was the main motivation of this work which, through a systematic manner, explored the capacity of machine learning and deep learning to unveil the tumor phenotypes hidden in radiologic and pathologic imaging that ultimately determine the therapeutic outcomes. This dissertation followed a coherent methodological trajectory through these four major studies: first, it judged the value of handcrafted radiomic features, then, it advanced their reliability through slice-selection strategies, thirdly, it enriched the representations with deep and transformer-based architectures, and finally, it employed foundation models to merge multi-modal radiologic and histopathologic data.

The primary breakthrough was the demonstration that 2D and 3D radiomics together hold biologically relevant information to differentiate responders from non-responders. Although the full 3D radiomics demonstrated higher accuracy, the results also revealed that it was more vulnerable to noise and segmentation uncertainty. As a consequence, the researchers noticed such a limitation that they decided to work on

the next study where radiomic feature robustness through slice-selection was the focus. Radiomic features from the central slice, and its two neighboring slices were found to be consistently more robust than features from single slices or fully 3D volumes, i.e., 'partial 3D' allowed for a better approximation of the complete spatial context than just one slice. These results gave convincing proof that having detailed control of the spatial context not only makes the radiomic features more reproducible but also stabilizes the prediction thus serving as a handy tool to lessen the variability that comes from the differently shaped tumor boundaries.

Building on this optimized radiomics foundation, the third study explored how handcrafted features compare and interact with the high-level deep representations that are extracted from vision transformers and MedSAM, a medical foundation model adapted for CT. The results revealed that deep embeddings, especially from MedSAM, contain a lot of phenotypic information that extends beyond the capabilities of handcrafted features. Multi-representation models that combine radiomics, ViT, and MedSAM embeddings have reached the highest predictive performance, thus confirming the complementary strengths of low-level texture descriptors and high-level semantic features. SHAP analyses further showed that deep features contributed to the majority of predictive signals, while radiomics provided interpretable descriptors linked to tumor heterogeneity.

The fourth study is going to expand the scope by fusing radiologic features from metastatic tumors with histopathologic embeddings derived from primary-tumor whole-slide images. This multimodal integration utilizes foundation models to comprehend the cross-scale biology of ovarian cancer. Macroscopic tumor burden on CT and microscopic tissue architecture on H&E slides. Despite anatomical and temporal differences between primary and metastatic tumor sites, the multimodal models are used to demonstrate the influence of intrinsic tumor biology from both micro– and macro–scale

phenotypes in improving chemotherapy-response prediction.

Collectively, the studies in this dissertation make the following key contributions:

1. Clarified the strengths and limitations of 2D and 3D radiomics. We recognized the necessity of balanced volumetric representation for the purpose of noise and segmentation uncertainty reduction.

2. Presented an incomplete 3D slice-selection framework that significantly improves feature robustness and predictive power across radiomic pipelines.

3. Proved that deep embeddings of foundation models can capture complementary biological information, surpassing handcrafted radiomics and enhancing prediction when fused.

4. Established a multimodal CT–histopathology fusion strategy, showing that integrating macro– and micro–scale tumor features creates more comprehensive imaging biomarkers.

These findings collectively advance the overarching objective of the dissertation: to build reliable, interpretable, and generalizable imaging biomarkers that support early treatment-response prediction in ovarian cancer. Although the results are promising, several limitations remain. Datasets from single institutions might limit the extent to which the findings can be generalized; segmentation variability continues to be a source of noise; and the availability of multimodal data is often scarce. The next steps in research should include validation that is large-scale and multi-center; adjusting acquisition protocols; using genomic and clinical data; as well as creating attention-based fusion models that interact more effectively with cross-representation. Besides that, a prospective clinical evaluation will be imperative to confirm the utility in the real world of decision support and patient stratification. In conclusion, the dissertation

provides evidence that combining radiomics, deep learning, and multimodal imaging has the potential to greatly enhance the prediction of the chemotherapy response in ovarian cancer. The study, by systematically addressing feature robustness, representation limits, and cross-modal integration issues, offers a fundamental framework for imaging biomarker development that is scalable, interpretable, and clinically meaningful. Such progress moves the discipline closer to the realization of precision oncology whereby treatment decisions are guided by quantitative, biologically grounded, and patient-specific imaging signatures.

7.1.1 Journal Articles

1. Chen, X., Zhang, K., **Abdoli, N.**, Gilley, P. W., Wang, X., Liu, H., Zheng, B., and Qiu, Y. (2022). Transformers improve breast cancer diagnosis from unregistered multi-view mammograms. *Diagnostics*, 12(7), 1549.

2. Zhang, K., **Abdoli, N.**, Gilley, P., Sadri, Y., Chen, X., Thai, T. C., Dockery, L., Moore, K., Mannel, R. S., and Qiu, Y. (2024). Developing a novel image marker to predict the clinical outcome of neoadjuvant chemotherapy (NACT) for ovarian cancer patients. *Computers in Biology and Medicine*, 172, 108240.

3. Zhang, K., Gilley, P., **Abdoli, N.**, Chen, X., Fung, K.-M., and Qiu, Y. (2023). Using symmetric illumination and color camera to achieve high throughput Fourier ptychographic microscopy. *Journal of Biophotonics*, 16(5), e202200303.

4. **Abdoli, N.**, Zhang, K., Gilley, P., Chen, X., Sadri, Y., Thai, T., Dockery, L., Moore, K., Mannel, R., and Qiu, Y. (2023). Evaluating the effectiveness of 2D and 3D CT image features for predicting tumor response to chemotherapy. *Bioengineering*, 10(11), 1334.

5. **Abdoli, N.**, Gilley, P., Zhang, K., Sadri, Y., Thai, T., Chen, Y., Dockery, L., Moore, K., Mannel, R., and Qiu, Y. (2026). A vision transformer-radiomics approach for enhanced chemotherapy outcome prediction in ovarian cancer. *Frontiers in Radiology*, 6, 1702977.

6. Gilley, P., Zhang, K., **Abdoli, N.**, Sadri, Y., Adhikari, L., Fung, K.-M., and Qiu, Y. (2024). Utilizing a pathomics biomarker to predict the effectiveness of bevacizumab in ovarian cancer treatment. *Bioengineering*, 11(7), 678.

7. Islam, W., **Abdoli, N.**, Alam, T. E., Jones, M., Mutembei, B. M., Yan, F., and Tang, Q. (2024). A neoteric feature extraction technique to predict the survival of gastric cancer patients. *Diagnostics*, 14(9), 954.

8. Zhang, K., Gilley, P., **Abdoli, N.**, Thai, T. C., Chen, Y., Dockery, L., Moore, K., Mannel, R. S., Tang, Q., and Qiu, Y. (2025). Parameter-efficient fine-tuning of foundation models to facilitate tumor response prediction for ovarian cancer patients. *BMC Medical Imaging*, 25(1), 504.

9. **Abdoli, N.**, Sadri, Y., Gilley, P., Zhang, K., Chen, Y., and Qiu, Y. (2025). From CNNs to SAM: A survey of deep learning techniques for liver tumor segmentation in CT images. *IEEE Access*, 13, 199074–199104.

10. Gilley, P., Zhang, K., **Abdoli, N.**, Sadri, Y., Adhikari, L., Fung, K.-M., and Qiu, Y. (2024). Development and assessment of multiple illumination color Fourier ptychographic microscopy for high throughput sample digitization. *Sensors*, 24(14), 4505.

7.1.2 Conference Proceedings

1. **Abdoli, N.**, Gilley, P., Zhang, K., Chen, X., Thai, T. C., Moore, K., Mannel, R. S., and Qiu, Y. (2023). Comparing the effectiveness of 2D and 3D features on predicting the response to chemotherapy for ovarian cancer patients. In *Biophotonics and Immune Responses XVIII* (Vol. 12380, pp. 13–18). SPIE.

2. Chen, X., Hu, M., Zhang, K., **Abdoli, N.**, Sadri, Y., Gilley, P., Chekuri, O. S. V., Omoumi, F. H., Qiu, Y., Zheng, B., et al. (2024). David vs. Goliath: Large foundation models are not outperforming small models in multi-view mammogram breast cancer prediction. In *Medical Imaging 2024: Computer-Aided Diagnosis* (Vol. 12927, pp. 217–224). SPIE.

3. Gilley, P. W., Zhang, K., **Abdoli, N.**, Sadri, Y., Chen, X., Fung, K.-M., Liu, H., and Qiu, Y. (2024). Evaluating the imaging quality of multiplexed illumination Fourier ptychography microscopy. In *Biophotonics and Immune Responses XIX* (Vol. 12843, pp. 30–35). SPIE.

4. Zhang, K., Gilley, P., **Abdoli, N.**, Fung, K.-M., and Qiu, Y. (2023). Evaluating the depth of field for Fourier ptychography microscopy. In *Biophotonics and Immune Responses XVIII* (Vol. 12380, pp. 7–12). SPIE.

5. Zhang, K., Gilley, P., Lu, X., Chen, X., **Abdoli, N.**, Zhang, R., Fung, K.-M., Liu, H., Zheng, B., Li, S., et al. (2022). Improving the resolution of chromosome imaging by high numerical aperture Fourier ptychography microscopy. In *Biophotonics and Immune Responses XVII* (Vol. 11961, Paper 1196102). SPIE.

6. **Abdoli, N.**, Zhang, K., Gilley, P., Thai, T. C., Chen, Y., Moore, K., Mannel, R. S., and Qiu, Y. (2025). Segment Anything Model (SAM) for metastatic tumor

segmentation: Evaluating its impact on chemotherapy response prediction. In *Biophotonics and Immune Responses XX* (Vol. 13296, pp. 22–28). SPIE.

7. Gai, T., Wang, J., Zhang, K., **Abdoli, N.**, Gilley, P., Henson, C., Chen, Y., and Qiu, Y. (2026). Using deep reinforcement learning to facilitate automated adaptive radiation therapy planning: An initial study. In *Biophotonics and Immune Responses XXI* (Vol. 13827, Paper 1382703). SPIE.

8. Zhang, K., Gilley, P., **Abdoli, N.**, Adhikari, L., Fung, K.-M., Tang, Q., and Qiu, Y. (2026). Fourier ptychographic microscopy using a doublet as objective lens: An initial evaluation. In *Biophotonics and Immune Responses XXI* (Vol. 13827, pp. 7–13). SPIE.

9. Zhang, K., Gai, T., Gilley, P., **Abdoli, N.**, Sadri, Y., Thai, T., Fung, K.-M., Chen, Y., Moore, K., Mannel, R., et al. (2025). A preliminary study on fusing CT and pathology information to improve the performance of tumor response prediction. In *Biophotonics and Immune Responses XX* (Vol. 13296, pp. 16–21). SPIE.

10. Sadri, Y., **Abdoli, N.**, Zhang, K., Gilley, P., Zhao, X., Prather, K., Shakir, H., and Qiu, Y. (2024). Developing a CT-image-based clinical marker to predict vasospasm for aneurysmal subarachnoid hemorrhage patients. In *Biophotonics and Immune Responses XIX* (Vol. 12843, pp. 13–18). SPIE.

7.2 Future Studies

The studies presented in this dissertation demonstrate that integrating radiomics, deep learning, foundation models, and multimodal imaging can improve the prediction of chemotherapy response in ovarian cancer. Building on these findings, future studies

should focus on four complementary directions: expanding multimodal biomarkers, improving multi-center robustness, developing reasoning-based clinical decision support, and advancing prospective clinical translation.

7.2.1 Expanding Multimodal Biomarkers

The framework developed in this dissertation integrates radiologic imaging and digital pathology to capture complementary macroscopic and microscopic characteristics of ovarian cancer. However, treatment response is influenced by biological and clinical factors that extend beyond imaging alone. Future studies should therefore incorporate additional patient-specific information, including genomic profiles, molecular biomarkers, and clinical variables. Multimodal cancer-learning studies have shown that integrating pathology, genomic, radiologic, and clinical information can provide a more comprehensive representation of disease and support prediction tasks such as survival, molecular subtype, and treatment outcome [146, 174, 175]. Clinical information presents a particular methodological challenge because it often consists of low-dimensional, heterogeneous, structured, and unstructured variables. Recent work has shown that clinical variables can be transformed into higher-dimensional representations through text prompting and foundation-model encoding and then integrated with pathology and genomic features through cross-attention mechanisms [175]. A similar strategy could be investigated for predicting chemotherapy response and 6-month progression-free survival in ovarian cancer. Future work should also incorporate imaging-derived biomarkers that characterize the overall physiological status of the patient rather than focusing exclusively on the tumor. Potential factors include visceral and subcutaneous adipose tissue, skeletal muscle volume and attenuation, bone mineral density, and indicators of cardiac and pulmonary function. CT-derived body-composition measurements have been associated with chemotoxicity and prognosis in patients with ovarian cancer, while

systematic reviews have identified relationships between muscle quantity or quality and ovarian cancer survival, although the reported evidence remains heterogeneous [176–178]. The number, size, and anatomical distribution of metastatic lesions could provide additional information regarding disease burden and treatment resistance. Foundation models may facilitate this expansion by providing transferable representations for multiple data types. UNI has demonstrated general-purpose representation learning across diverse computational pathology tasks, while CONCH combines histopathology images with biomedical language supervision [152, 153]. Future studies could evaluate whether embeddings derived from imaging, pathology, genomic data, and clinical narratives can be mapped into a shared representation and fused to produce more accurate and generalizable biomarkers than any individual modality alone.

7.2.2 Improving Generalizability and Clinical Robustness

A major limitation of the present work is its reliance on relatively small, single-institution cohorts. Future studies should validate the proposed biomarkers using larger multi-center datasets representing diverse patient populations, scanners, reconstruction settings, tissue-processing procedures, and clinical practices. External validation is particularly important because differences between institutions can produce domain shifts that reduce the performance of models developed at a single center. Domain adaptation provides one potential strategy for reducing these differences. In computational pathology, Adversarial Fourier-based Domain Adaptation (AIDA) was developed to learn representations that are less sensitive to variations in staining, tissue preparation, and scanner characteristics. The method improved multi-center histopathology classification across several cancer types, including ovarian cancer [172]. For radiologic imaging, domain-adaptation methods can similarly reduce discrepancies associated with scanner manufacturers, acquisition parameters, reconstruction kernels, and image appearance.

Existing approaches include domain-adversarial representation learning and image-to-image translation methods such as CycleGAN-based adaptation [179]. Segmentation variability represents another source of uncertainty. Manual delineation remains an important clinical reference standard, but it is labor-intensive and sensitive to interobserver and interdisciplinary variation. Previous studies have shown that segmentation differences can substantially affect CT radiomic features and their downstream reproducibility [180]. Future studies should therefore systematically assess foundation segmentation models for automated delineation of primary and metastatic tumors across multiple organs. Both 2D and 3D segmentation should be evaluated, and performance should be assessed not only using overlap measures but also according to the stability of extracted biomarkers and predictive models. Multi-center robustness should ultimately be assessed through independent external validation, standardized preprocessing, harmonized feature extraction, and prospective testing. Public resources such as TCGA and TCIA may provide useful imaging, pathology, genomic, and clinical data for preliminary external evaluation, although differences in treatment regimens and endpoint definitions must be considered when interpreting such validation results.

7.2.3 Reasoning-Based Artificial Intelligence for Clinical Decision Support

In the current dissertation, foundation models were primarily used as feature extractors. A more exploratory future direction is to move beyond static embeddings toward reasoning-based systems that integrate pathology, radiology, molecular findings, treatment history, and clinical information. One possible approach is agentic reasoning, in which specialized language-model agents simulate the contributions of different members of a multidisciplinary tumor board. Preliminary research in gynecologic

oncology has shown that teams of role-specialized language-model agents can outperform single-model prompting approaches on complex staging, treatment-selection, complication-management, and follow-up questions [181]. However, this work remains a preprint and requires independent clinical validation. More broadly, recent studies indicate that case-grounded AI agents can synthesize longitudinal clinical records and molecular evidence to generate auditable treatment recommendations. A case-grounded agent for hematological malignancies demonstrated substantial concordance with multidisciplinary tumor-board decisions across internal, external, and prospective evaluations [182]. Reviews of AI-supported multidisciplinary tumor boards also report potential benefits in information synthesis, guideline alignment, documentation, and workflow efficiency, while emphasizing the continued need for expert oversight and prospective validation [183]. Future work could investigate whether structured outputs from specialist-agent discussions provide useful predictive variables for chemotherapy response or progression-free survival. Such outputs might include the inferred diagnosis, relevant biomarkers, treatment hypotheses, unresolved questions, and explicitly identified uncertainty. Two questions would be particularly important: whether multidisciplinary reasoning provides predictive information beyond a single-pass extraction of the same patient record, and whether explicitly represented uncertainty improves the calibration of outcome predictions. Because these systems may generate incorrect or unsupported recommendations, they should be treated as decision-support tools rather than autonomous decision makers. Their evaluation should include factual grounding, reproducibility, uncertainty calibration, clinician agreement, and the clinical consequences of errors.

7.2.4 Translation Toward Precision Oncology

The ultimate objective of this research is not only to improve predictive performance but also to support clinically meaningful treatment stratification. Future studies should therefore evaluate the proposed framework prospectively within routine clinical workflows. Prospective evaluation would determine whether model predictions remain reliable when generated before treatment decisions and whether they provide information that influences therapy selection, monitoring, or patient counseling. Clinical translation will require robust external validation, automated and reproducible preprocessing, transparent model outputs, and efficient integration with existing radiology, pathology, and electronic health record systems. The framework should also be evaluated for calibration and clinical utility rather than discrimination alone. Decision-curve analysis and prospective impact studies could help determine whether the predictions produce a meaningful net benefit compared with current clinical assessment. Interpretability will remain particularly important when using foundation-model embeddings, because individual latent features may not correspond directly to recognizable biological characteristics. Future studies should therefore combine attention visualization, feature attribution, representative image retrieval, and association with known histopathologic, radiologic, molecular, and clinical variables. These analyses may help determine whether the models rely on biologically plausible patterns and may increase clinician confidence in their use. Collectively, these future directions would extend the work from tumor-centered imaging biomarkers toward a multi-scale, multi-institutional, and clinically integrated decision-support framework. Such a framework could combine tumor phenotype, tissue architecture, molecular characteristics, systemic patient factors, and uncertainty-aware reasoning to support more individualized treatment of ovarian cancer.

Appendix A

Radiomics Features

This appendix summarizes the radiomics features extracted using the PyRadiomics package. Features were computed from both 2D and 3D tumor segmentations and grouped into three major categories: (i) shape-based features, (ii) first-order (density) features, and (iii) texture features. Texture features were further divided into Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighboring Gray Tone Difference Matrix (NGTDM), and Gray Level Dependence Matrix (GLDM).

More detailed descriptions of these features can be found in the official PyRadiomics documentation at `https://pyradiomics.readthedocs.io`.

Table A.1: Summary of extracted radiomics feature classes.

Feature Category	**Feature Subclass**	**2D**	**3D**
Shape	–	10	16
First-order (Density)	–	19	19
Texture	Gray Level Co-occurrence Matrix (GLCM)	24	24
	Gray Level Run Length Matrix (GLRLM)	16	16
	Gray Level Size Zone Matrix (GLSZM)	16	16
	Neighboring Gray Tone Difference Matrix (NGTDM)	5	5
	Gray Level Dependence Matrix (GLDM)	14	14
Total		**104**	**110**

A.1 First-order Features (Density)

- Energy
- Total Energy
- Entropy
- Minimum
- 10th Percentile
- 90th Percentile
- Maximum
- Mean
- Median
- Interquartile Range
- Range
- Mean Absolute Deviation (MAD)
- Robust Mean Absolute Deviation (rMAD)
- Root Mean Squared (RMS)
- Standard Deviation
- Variance
- Skewness
- Kurtosis
- Uniformity

A.2 Shape Features (2D)

- Mesh Surface
- Pixel Surface
- Perimeter
- Perimeter-to-Surface Ratio
- Sphericity
- Spherical Disproportion
- Maximum 2D Diameter

- Major Axis Length
- Minor Axis Length
- Elongation

A.3 Shape Features (3D)

- Mesh Volume
- Voxel Volume
- Surface Area
- Surface Area-to-Volume Ratio
- Sphericity
- Compactness 1
- Spherical Disproportion
- Maximum 3D Diameter
- Maximum 2D Diameter (Slice)
- Maximum 2D Diameter (Column)
- Maximum 2D Diameter (Row)
- Major Axis Length
- Minor Axis Length
- Least Axis Length
- Elongation
- Flatness

A.4 Gray Level Co-occurrence Matrix Features (GLCM)

- Autocorrelation
- Joint Average
- Cluster Prominence
- Cluster Shade

- Cluster Tendency
- Contrast
- Correlation
- Difference Average
- Difference Entropy
- Difference Variance
- Joint Energy
- Joint Entropy
- Informational Measure of Correlation (IMC1)
- Informational Measure of Correlation (IMC2)
- Inverse Difference Moment (IDM)
- Maximal Correlation Coefficient (MCC)
- Inverse Difference Moment Normalized (IDMN)
- Inverse Difference (ID)
- Inverse Difference Normalized (IDN)
- Inverse Variance
- Maximum Probability
- Sum Average
- Sum Entropy
- Sum of Squares

A.5 Gray Level Run Length Matrix Features (GLRLM)

- Short Run Emphasis (SRE)
- Long Run Emphasis (LRE)
- Gray Level Non-Uniformity (GLN)
- Gray Level Non-Uniformity Normalized (GLNN)
- Run Length Non-Uniformity (RLN)
- Run Length Non-Uniformity Normalized (RLNN)

- Run Percentage (RP)
- Gray Level Variance (GLV)
- Run Variance (RV)
- Run Entropy (RE)
- Low Gray Level Run Emphasis (LGLRE)
- High Gray Level Run Emphasis (HGLRE)
- Short Run Low Gray Level Emphasis (SRLGLE)
- Short Run High Gray Level Emphasis (SRHGLE)
- Long Run Low Gray Level Emphasis (LRLGLE)
- Long Run High Gray Level Emphasis (LRHGLE)

A.6 Gray Level Size Zone Matrix Features (GLSZM)

- Small Area Emphasis (SAE)
- Large Area Emphasis (LAE)
- Gray Level Non-Uniformity (GLN)
- Gray Level Non-Uniformity Normalized (GLNN)
- Size Zone Non-Uniformity (SZN)
- Size Zone Non-Uniformity Normalized (SZNN)
- Zone Percentage (ZP)
- Gray Level Variance (GLV)
- Zone Variance (ZV)
- Zone Entropy (ZE)
- Low Gray Level Zone Emphasis (LGLZE)
- High Gray Level Zone Emphasis (HGLZE)
- Small Area Low Gray Level Emphasis (SALGLE)
- Small Area High Gray Level Emphasis (SAHGLE)
- Large Area Low Gray Level Emphasis (LALGLE)
- Large Area High Gray Level Emphasis (LAHGLE)

A.7 Neighboring Gray Tone Difference Matrix Features (NGTDM)

- Coarseness
- Contrast
- Busyness
- Complexity
- Strength

A.8 Gray Level Dependence Matrix Features (GLDM)

- Small Dependence Emphasis (SDE)
- Large Dependence Emphasis (LDE)
- Gray Level Non-Uniformity (GLN)
- Dependence Non-Uniformity (DN)
- Dependence Non-Uniformity Normalized (DNN)
- Gray Level Variance (GLV)
- Dependence Variance (DV)
- Dependence Entropy (DE)
- Low Gray Level Emphasis (LGLE)
- High Gray Level Emphasis (HGLE)
- Small Dependence Low Gray Level Emphasis (SDLGLE)
- Small Dependence High Gray Level Emphasis (SDHGLE)
- Large Dependence Low Gray Level Emphasis (LDLGLE)
- Large Dependence High Gray Level Emphasis (LDHGLE)

Appendix B

Integrated Radiomics

Table B.1: Performance metrics of the models trained with different feature sets.

Features	**Accuracy**	**Precision**	**Sensitivity**	**F1 Score**	**Specificity**
Radiomics	0.71 ± 0.04	0.81 ± 0.03	0.74 ± 0.09	0.77 ± 0.04	0.63 ± 0.12
ViT	0.65 ± 0.05	0.78 ± 0.05	0.70 ± 0.14	0.72 ± 0.07	0.56 ± 0.17
MedSAM	0.80 ± 0.04	0.88 ± 0.03	0.83 ± 0.06	0.86 ± 0.03	0.77 ± 0.08
Radiomics, ViT	0.78 ± 0.04	0.87 ± 0.04	0.80 ± 0.07	0.83 ± 0.04	0.73 ± 0.13
Radiomics, MedSAM	0.81 ± 0.05	0.88 ± 0.04	0.84 ± 0.07	0.86 ± 0.04	0.74 ± 0.14
ViT, MedSAM	0.82 ± 0.03	0.90 ± 0.03	0.84 ± 0.06	0.87 ± 0.03	0.81 ± 0.06
Radiomics, ViT, MedSAM	0.83 ± 0.04	0.91 ± 0.03	0.84 ± 0.05	0.87 ± 0.03	0.82 ± 0.06

Table B.2: No-SMOTE ablation results for the best-performing integrated model (Radiomics + ViT + MedSAM) using the same stratified nested cross-validation protocol.

Features	**AUC**	**Accuracy**	**Precision**	**Sensitivity**	**F1 Score**	**Specificity**
Radiomics, ViT, MedSAM	0.90 ± 0.03	0.81 ± 0.04	0.89 ± 0.03	0.84 ± 0.06	0.86 ± 0.03	0.79 ± 0.06

Table B.3: Features selected by the best-performing integrated model (Radiomics + ViT + MedSAM). Features within each source are ordered according to decreasing mean absolute SHAP value.

MedSAM	**ViT**	**Radiomics**
449	236	Wavelet-HLH_GLSZM_SizeZoneNonUniformityNormalized
705	154	Wavelet-LHL_GLDM_DependenceEntropy
247	22	Wavelet-HHH_GLDM_LargeDependenceEmphasis
499	503	Wavelet-LHH_GLSZM_GrayLevelVariance
649	466	Wavelet-LHL_firstorder_Skewness
260	329	Logarithm_GLSZM_SmallAreaEmphasis
463	704	LBP_firstorder_10Percentile
71	591	Squareroot_GLDM_DependenceVariance
455	385	–
491	674	–
145	–	–
549	–	–

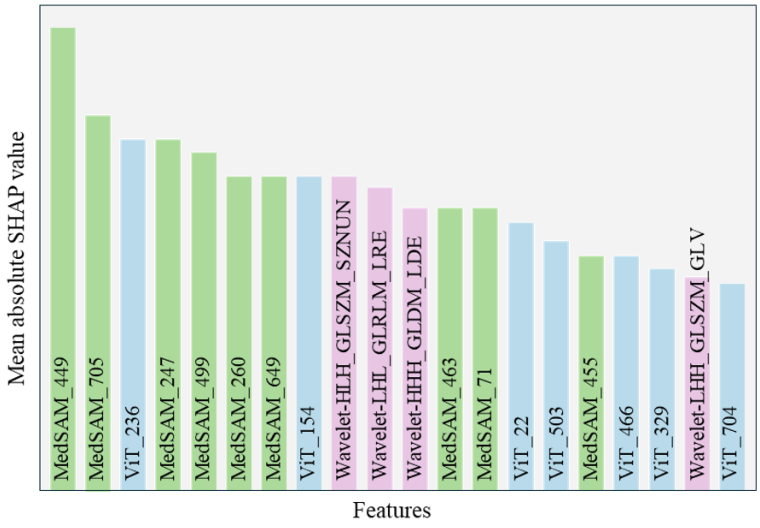


Figure B.1: Contribution of the top 20 optimal features to the best-performing classifier, as quantified by SHAP values.

Reference List

[1] Gordon C Jayson, Elise C Kohn, Henry C Kitchener, and Jonathan A Ledermann. Ovarian cancer. *The Lancet*, 384(9951):1376–1388, 2014.

[2] Rebecca L Siegel, Tyler B Kratzer, Angela N Giaquinto, Hyuna Sung, and Ahmedin Jemal. Cancer statistics, 2025. *CA: A Cancer Journal for Clinicians*, 75(1):10, 2025.

[3] Danijela Jelovac and Deborah K Armstrong. Recent progress in the diagnosis and treatment of ovarian cancer. *CA: A Cancer Journal for Clinicians*, 61(3): 183–203, 2011.

[4] Penelope M Webb and Susan J Jordan. Global epidemiology of epithelial ovarian cancer. *Nature Reviews Clinical Oncology*, 21(5):389–400, 2024.

[5] Wolfgang A Weber. Assessing tumor response to therapy. *Journal of Nuclear Medicine*, 50(Suppl 1):1S–10S, 2009.

[6] National Cancer Institute (US). Surveillance Program, SEER Program (National Cancer Institute (US)), and National Center for Health Statistics (US). *Cancer statistics review*. US Department of Health and Human Services, Public Health Service, National . . . , 1973.

[7] Thomas J Herzog, Deborah K Armstrong, Mark F Brady, Robert L Coleman, Mark H Einstein, Bradley J Monk, Robert S Mannel, J Tate Thigpen, Sharee A Umpierre, Jeannine A Villella, et al. Ovarian cancer clinical trial endpoints: Society of Gynecologic Oncology white paper. *Gynecologic Oncology*, 132(1): 8–17, 2014.

[8] Sevinj Yolchuyeva, Elena Giacomazzi, Marion Tonneau, Fabien Lamaze, Michele Orain, François Coulombe, Julie Malo, Wiam Belkaid, Bertrand Routy, Philippe Joubert, et al. Radiomics approaches to predict PD-L1 and PFS in advanced non-small cell lung patients treated with immunotherapy: a multi-institutional study. *Scientific Reports*, 13(1):11065, 2023.

[9] Elizabeth A Eisenhauer, Patrick Therasse, Jan Bogaerts, Lawrence H Schwartz, Danielle Sargent, Robert Ford, Janet Dancey, Stephen Arbuck, Steve Gwyther, Margaret Mooney, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2): 228–247, 2009.

[10] Leonard Fass. Imaging and cancer: a review. *Molecular Oncology*, 2(2):115–152, 2008.

[11] Marco Gerlinger, Andrew J Rowan, Stuart Horswell, James Larkin, David Endesfelder, Eva Gronroos, Pierre Martinez, Nicholas Matthews, Aengus Stewart, Patrick Tarpey, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine*, 366(10):883–892, 2012.

[12] Roger Sun, Elaine Johanna Limkin, Maria Vakalopoulou, Laurent Dercle, Stéphane Champiat, Shan Rong Han, Loïc Verlingue, David Brandao, Andrea Lancia, Samy Ammari, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *The Lancet Oncology*, 19(9):1180–1191, 2018.

[13] Roger Sun, Nora Sundahl, Markus Hecht, Florian Putz, Andrea Lancia, Angela Rouyar, Marina Milic, Alexandre Carré, Enzo Battistella, Emilie Alvarez Andres, et al. Radiomics to predict outcomes and abscopal response of patients with cancer treated with immunotherapy combined with radiotherapy using a validated signature of CD8 cells. *Journal for ImmunoTherapy of Cancer*, 8(2): e001429, 2020.

[14] Evis Sala, E Mema, Y Himoto, H Veeraraghavan, JD Brenton, A Snyder, B Weigelt, and HA Vargas. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clinical Radiology*, 72(1):3–10, 2017.

[15] Robert J Gillies, Paul E Kinahan, and Hedvig Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016.

[16] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature Medicine*, 25(1): 24–29, 2019.

[17] Dow-Mu Koh, Nickolas Papanikolaou, Ulrich Bick, Rowland Illing, Charles E Kahn Jr, Jayshree Kalpathi-Cramer, Celso Matos, Luis Martí-Bonmatí, Anne Miles, Seong Ki Mun, et al. Artificial intelligence and machine learning in cancer imaging. *Communications Medicine*, 2(1):133, 2022.

[18] Alberto Traverso, Leonard Wee, Andre Dekker, and Robert Gillies. Repeatability and reproducibility of radiomic features: a systematic review. *International Journal of Radiation Oncology, Biology, Physics*, 102(4):1143–1158, 2018.

[19] He-Li Xu, Ting-Ting Gong, Fang-Hua Liu, Hong-Yu Chen, Qian Xiao, Yang Hou, Ying Huang, Hong-Zan Sun, Yu Shi, Song Gao, et al. Artificial intelligence performance in image-based ovarian cancer identification: a systematic review and meta-analysis. *eClinicalMedicine*, 53, 2022.

[20] Pratik Adusumilli, Nishant Ravikumar, Geoff Hall, Sarah Swift, Nicolas Orsi, and Andrew Scarsbrook. Radiomics in the evaluation of ovarian masses—a systematic review. *Insights into Imaging*, 14(1):165, 2023.

[21] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud GPM Van Stiphout, Patrick Granton, Catharina ML Zegers, Robert Gillies, Ronald Boellard, André Dekker, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4):441–446, 2012.

[22] Alex Zwanenburg, Stefan Leger, Martin Vallières, and Steffen Löck. Image biomarker standardisation initiative. *arXiv preprint arXiv:1612.07003*, 2016.

[23] Roberto Berenguer, María Del Rosario Pastor-Juan, Jesús Canales-Vázquez, Miguel Castro-García, María Victoria Villas, Francisco Mansilla Legorburo, and Sebastià Sabater. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. *Radiology*, 288(2):407–415, 2018.

[24] Muhammad Shafiq-ul Hassan, Geoffrey G Zhang, Kujtim Latifi, Ghanim Ullah, Dylan C Hunt, Yoganand Balagurunathan, Mahmoud Abrahem Abdalah, Matthew B Schabath, Dmitry G Goldgof, Dennis Mackin, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Medical Physics*, 44(3):1050–1062, 2017.

[25] Joost JM Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JWL Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer Research*, 77(21):e104–e107, 2017.

[26] Chintan Parmar, Emmanuel Rios Velazquez, Ralph Leijenaar, Mohammed Jermoumi, Sara Carvalho, Raymond H Mak, Sushmita Mitra, B Uma Shankar, Ron Kikinis, Benjamin Haibe-Kains, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS ONE*, 9(7):e102107, 2014.

[27] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.

[28] Mireia Crispin-Ortuzar, Ramona Woitek, Marika AV Reinius, Elizabeth Moore, Lucian Beer, Vlad Bura, Leonardo Rundo, Cathal McCague, Stephan Ursprung, Lorena Escudero Sanchez, et al. Integrated radiogenomics models predict response to neoadjuvant chemotherapy in high grade serous ovarian cancer. *Nature Communications*, 14(1):6756, 2023.

[29] L. Rundo. Clinically interpretable radiomics-based prediction of histopathologic response to neoadjuvant chemotherapy in high-grade serous ovarian carcinoma. *Frontiers in Oncology*, 12:868265, 2022.

[30] Jack Breen, Katie Allen, Kieran Zucker, Pratik Adusumilli, Andrew Scarsbrook, Geoff Hall, Nicolas M Orsi, and Nishant Ravikumar. Artificial intelligence in ovarian cancer histopathology: a systematic review. *npj Precision Oncology*, 7 (1):83, 2023.

[31] Sandra Orsulic, Joshi John, Ann E Walts, and Arkadiusz Gertych. Computational pathology in ovarian cancer. *Frontiers in Oncology*, 12:924945, 2022.

[32] Hui-zhu Chen, Xin-rong Wang, Fu-min Zhao, Xi-jian Chen, Xue-sheng Li, Gang Ning, and Ying-kun Guo. A CT-based radiomics nomogram for predicting early recurrence in patients with high-grade serous ovarian cancer. *European Journal of Radiology*, 145:110018, 2021.

[33] Yihao Li, Mostafa El Habib Daho, Pierre-Henri Conze, Rachid Zeghlache, Hugo Le Boité, Ramin Tadayoni, Béatrice Cochener, Mathieu Lamard, and Gwenolé Quellec. A review of deep learning-based information fusion techniques for multimodal medical image classification. *Computers in Biology and Medicine*, 177: 108635, 2024.

[34] Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, Nicolo Fusi, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature Medicine*, 30(10):2924–2935, 2024.

[35] Rebecca L Siegel, Kimberly D Miller, Hannah E Fuchs, and Ahmedin Jemal. Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians*, 71(1):7–33, 2021.

[36] Elaine Johanna Limkin, Sylvain Reuzé, Alexandre Carré, Roger Sun, Antoine Schernberg, Anthony Alexis, Eric Deutsch, Charles Ferté, and Charlotte Robert. The complexity of tumor shape, spiculatedness, correlates with tumor radiomic shape features. *Scientific Reports*, 9(1):4329, 2019.

[37] Stavroula Kyriazi, Stan B Kaye, and Nandita M Desouza. Imaging ovarian cancer and peritoneal metastases—current and emerging techniques. *Nature Reviews Clinical Oncology*, 7(7):381–393, 2010.

[38] Gopichandh Danala, Yunzhi Wang, Theresa Thai, Camille C Gunderson, Katherine M Moxley, Kathleen Moore, Robert S Mannel, Samuel Cheng, Hong Liu, Bin Zheng, et al. Improving efficacy of metastatic tumor segmentation to facilitate early prediction of ovarian cancer patients' response to chemotherapy. In *Biophotonics and Immune Responses XII*, volume 10065, pages 47–52. SPIE, 2017.

[39] Xuxin Chen, Wei Liu, Theresa C Thai, Tara Castellano, Camille C Gunderson, Kathleen Moore, Robert S Mannel, Hong Liu, Bin Zheng, and Yuchen Qiu. Developing a new radiomics-based CT image marker to detect lymph node

metastasis among cervical cancer patients. *Computer Methods and Programs in Biomedicine*, 197:105759, 2020.

[40] Stefan P Haider, Barbara Burtness, Wendell G Yarbrough, and Seyedmehdi Payabvash. Applications of radiomics in precision diagnosis, prognostication and treatment planning of head and neck squamous cell carcinomas. *Cancers of the Head & Neck*, 5(1):6, 2020.

[41] Zhenyu Liu, Shuo Wang, Di Dong, Jingwei Wei, Cheng Fang, Xuezhi Zhou, Kai Sun, Longfei Li, Bo Li, Meiyun Wang, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics*, 9(5):1303, 2019.

[42] Alex Zwanenburg, Martin Vallières, Mahmoud A Abdalah, Hugo JWL Aerts, Vincent Andrearczyk, Aditya Apte, Saeed Ashrafinia, Spyridon Bakas, Roelof J Beukinga, Ronald Boellaard, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2):328–338, 2020.

[43] Inzamam Mashood Nasir, Muhammad Attique Khan, Mussarat Yasmin, Jamal Hussain Shah, Marcin Gabryel, Rafał Scherer, and Robertas Damaševičius. Pearson correlation-based feature selection for document classification using balanced training. *Sensors*, 20(23):6793, 2020.

[44] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1):267–288, 1996.

[45] Xuxin Chen, Ximin Wang, Ke Zhang, Kar-Ming Fung, Theresa C Thai, Kathleen Moore, Robert S Mannel, Hong Liu, Bin Zheng, and Yuchen Qiu. Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, 79:102444, 2022.

[46] Ibiayi Dagogo-Jack and Alice T Shaw. Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology*, 15(2):81–94, 2018.

[47] Stefania Rizzo, Francesca Botta, Sara Raimondi, Daniela Origgi, Cristiana Fanciullo, Alessio Giuseppe Morganti, and Massimo Bellomi. Radiomics: the facts and the challenges of image analysis. *European Radiology Experimental*, 2(1):36, 2018.

[48] Lingwei Meng, Di Dong, Xin Chen, Mengjie Fang, Rongpin Wang, Jing Li, Zaiyi Liu, and Jie Tian. 2D and 3D CT radiomic features performance comparison in characterization of gastric cancer: a multi-center study. *IEEE Journal of Biomedical and Health Informatics*, 25(3):755–763, 2020.

[49] Rebecca L Siegel, Kimberly D Miller, Hannah E Fuchs, and Ahmedin Jemal. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, 72(1):7–33, 2022.

[50] Rasha Kamal, Soha Hamed, Sahar Mansour, Yasmine Mounir, and Sahar Abdel Sallam. Ovarian cancer screening—ultrasound; impact on ovarian cancer mortality. *The British Journal of Radiology*, 91(1090):20170571, 2018.

[51] Steven Narod. Can advanced-stage ovarian cancer be cured? *Nature Reviews Clinical Oncology*, 13(4):255–261, 2016.

[52] Brandon JD Rein, Sajal Gupta, Rima Dada, Joelle Safi, Chad Michener, and Ashok Agarwal. Potential markers for detection and monitoring of ovarian cancer. *Journal of Oncology*, 2011(1):475983, 2011.

[53] Asha Mahadevappa, Shruthi Mysore Krishna, and Manjunath Gubbanna Vimala. Diagnostic and prognostic significance of Ki-67 immunohistochemical expression in surface epithelial ovarian carcinoma. *Journal of Clinical and Diagnostic Research: JCDR*, 11(2):EC08, 2017.

[54] Kristin K Zorn, Chunqiao Tian, William P McGuire, William J Hoskins, Maurie Markman, Franco M Muggia, Peter G Rose, Robert F Ozols, David Spriggs, and Deborah K Armstrong. The prognostic value of pretreatment CA 125 in patients with advanced ovarian carcinoma: a Gynecologic Oncology Group study. *Cancer*, 115(5):1028–1035, 2009.

[55] György Sölétormos, Michael J Duffy, Suher Othman Abu Hassan, René HM Verheijen, Bengt Tholander, Robert C Bast Jr, Katja N Gaarenstroom, Catharine M Sturgeon, Johannes M Bonfrer, Per Hyltoft Petersen, et al. Clinical use of cancer biomarkers in epithelial ovarian cancer: updated guidelines from the European Group on Tumor Markers. *International Journal of Gynecological Cancer*, 26(1): 43–51, 2016.

[56] Bum Woo Park, Jeong Kon Kim, Changhoe Heo, and Kye Jin Park. Reliability of CT radiomic features reflecting tumour heterogeneity according to image quality and image processing parameters. *Scientific Reports*, 10(1):3852, 2020.

[57] Manish R Sharma, Michael L Maitland, and Mark J Ratain. RECIST: no longer the sharpest tool in the oncology clinical trials toolbox—point. *Cancer Research*, 72(20):5145–5149, 2012.

[58] Hugo JWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5(1):4006, 2014.

[59] Adam C Palmer and Peter K Sorger. Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell*, 171(7):1678–1691, 2017.

[60] Chen Shen, Zhenyu Liu, Min Guan, Jiangdian Song, Yucheng Lian, Shuo Wang, Zhenchao Tang, Di Dong, Lingfei Kong, Meiyun Wang, et al. 2D and 3D CT radiomics features prognostic performance comparison in non-small cell lung cancer. *Translational Oncology*, 10(6):886–894, 2017.

[61] Lingwei Meng, Di Dong, Xin Chen, Mengjie Fang, Rongpin Wang, Jing Li, Zaiyi Liu, and Jie Tian. 2D and 3D CT radiomic features performance comparison in characterization of gastric cancer: a multi-center study. *IEEE Journal of Biomedical and Health Informatics*, 25(3):755–763, 2020.

[62] Lei Xu, Pengfei Yang, Eric Alexander Yen, Yidong Wan, Yangkang Jiang, Zuozhen Cao, Xiaoyong Shen, Yan Wu, Jing Wang, Chen Luo, et al. A multi-organ cancer study of the classification performance using 2D and 3D image features in radiomics analysis. *Physics in Medicine & Biology*, 64(21):215009, 2019.

[63] Geewon Lee, Ho Yun Lee, Hyunjin Park, Mark L Schiebler, Edwin JR van Beek, Yoshiharu Ohno, Joon Beom Seo, and Ann Leung. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: state of the art. *European Journal of Radiology*, 86:297–307, 2017.

[64] Lesley J Fallowfield and Anne Fleissig. The value of progression-free survival to patients with advanced-stage cancer. *Nature Reviews Clinical Oncology*, 9(1): 41–47, 2012.

[65] Payel Roy, Saurab Dutta, Nilanjan Dey, Goutami Dey, Sayan Chakraborty, and Ruben Ray. Adaptive thresholding: a comparative study. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pages 1182–1186. IEEE, 2014.

[66] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[67] Yuchen Qiu, Maxine Tan, Scott McMeekin, Theresa Thai, Kai Ding, Kathleen Moore, Hong Liu, and Bin Zheng. Early prediction of clinical benefit of treating ovarian cancer using quantitative CT image feature analysis. *Acta Radiologica*, 57(9):1149–1155, 2016.

[68] Gopichandh Danala, Theresa Thai, Camille C Gunderson, Katherine M Moxley, Kathleen Moore, Robert S Mannel, Hong Liu, Bin Zheng, and Yuchen Qiu. Applying quantitative CT image feature analysis to predict response of ovarian cancer patients to chemotherapy. *Academic Radiology*, 24(10):1233–1239, 2017.

[69] Neman Abdoli, Patrik Gilley, Ke Zhang, Xuxin Chen, Theresa C Thai, Kathleen Moore, Robert S Mannel, and Yuchen Qiu. Comparing the effectiveness of 2D and 3D features on predicting the response to chemotherapy for ovarian cancer patients. In *Biophotonics and Immune Responses XVIII*, volume 12380, pages 13–18. SPIE, 2023.

[70] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675, 2012.

[71] Inzamam Mashood Nasir, Muhammad Attique Khan, Mussarat Yasmin, Jamal Hussain Shah, Marcin Gabryel, Rafał Scherer, and Robertas Damaševičius. Pearson correlation-based feature selection for document classification using balanced training. *Sensors*, 20(23):6793, 2020.

[72] Miriam Piles, Rob Bergsma, Daniel Gianola, Hélène Gilbert, and Llibertat Tusell. Feature selection stability and accuracy of prediction models for genomic prediction of residual feed intake in pigs using machine learning. *Frontiers in Genetics*, 12:611506, 2021.

[73] Huaqing Zhang, Jian Wang, Zhanquan Sun, Jacek M Zurada, and Nikhil R Pal. Feature selection for neural networks using group lasso regularization. *IEEE Transactions on Knowledge and Data Engineering*, 32(4):659–673, 2019.

[74] M Graziano Usai, Mike E Goddard, and Ben J Hayes. LASSO with cross-validation for genomic selection. *Genetics Research*, 91(6):427–436, 2009.

[75] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[76] Graziella Orru, William Pettersson-Yeo, Andre F Marquand, Giuseppe Sartori, and Andrea Mechelli. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews*, 36(4):1140–1152, 2012.

[77] Luca Steardo Jr, Elvira Anna Carbone, Renato De Filippis, Claudia Pisanu, Cristina Segura-Garcia, Alessio Squassina, Pasquale De Fazio, and Luca Steardo. Application of support vector machine on fMRI data as biomarkers in schizophrenia diagnosis: a systematic review. *Frontiers in Psychiatry*, 11:588, 2020.

[78] A Sheryl Oliver, M Anuradha, J Jean Justus, Kiranmai Bellam, and T Jayasankar. An efficient coding network based feature extraction with support vector machine based classification model for CT lung images. *Journal of Medical Imaging and Health Informatics*, 10(11):2628–2633, 2020.

[79] Haifeng Wang, Bichen Zheng, Sang Won Yoon, and Hoo Sang Ko. A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, 267(2):687–699, 2018.

[80] Shujun Huang, Nianguang Cai, Pedro Penzuti Pacheco, Shavira Narrandes, Yang Wang, and Wayne Xu. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics & Proteomics*, 15(1):41–51, 2018.

[81] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2 (3):1–27, 2011.

[82] Ruiming Liu, Erqi Liu, Jie Yang, Ming Li, and Fanglin Wang. Optimizing the hyper-parameters for SVM by combining evolution strategies with a grid search. In *Intelligent Control and Automation: International Conference on Intelligent Computing, ICIC 2006, Kunming, China, August 16–19, 2006*, pages 712–721. Springer, 2006.

[83] Qiang Li and Kunio Doi. Reduction of bias and variance for evaluation of computer-aided diagnostic schemes. *Medical Physics*, 33(4):868–875, 2006.

[84] Jayawant N Mandrekar. Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, 5(9):1315–1316, 2010.

[85] Arnau Oliver, Jordi Freixenet, Joan Marti, Elsa Perez, Josep Pont, Erika RE Denton, and Reyer Zwiggelaar. A review of automatic mass detection and segmentation in mammographic images. *Medical Image Analysis*, 14(2):87–110, 2010.

[86] Rebecca L Siegel, Angela N Giaquinto, and Ahmedin Jemal. Cancer statistics, 2024. *CA: A Cancer Journal for Clinicians*, 74(1):12–49, 2024.

[87] Ernst Lengyel. Ovarian cancer development and metastasis. *The American Journal of Pathology*, 177(3):1053–1064, 2010.

[88] Mina Sharbatoghli, Somayeh Vafaei, Hamidreza Aboulkheyr Es, Mohsen Asadi-Lari, Mehdi Totonchi, and Zahra Madjd. Prediction of the treatment response in ovarian cancer: a ctDNA approach. *Journal of Ovarian Research*, 13(1):124, 2020.

[89] Pierre-Emmanuel Colombo, Michel Fabbro, Charles Theillet, Frédéric Bibeau, Philippe Rouanet, and Isabelle Ray-Coquard. Sensitivity and resistance to treatment in the primary management of epithelial ovarian cancer. *Critical Reviews in Oncology/Hematology*, 89(2):207–216, 2014.

[90] Donavon Hiss. Optimizing molecular-targeted therapies in ovarian cancer: the renewed surge of interest in ovarian cancer biomarkers and cell signaling pathways. *Journal of Oncology*, 2012(1):737981, 2012.

[91] Xiaohong Chang, Xue Ye, Li Dong, Hongyan Cheng, Yexia Cheng, Lirong Zhu, Qinping Liao, Yang Zhao, Li Tian, As Tianyun Fu, et al. Human epididymis protein 4 (HE4) as a serum tumor biomarker in patients with ovarian carcinoma. *International Journal of Gynecological Cancer*, 21(5):852–858, 2011.

[92] Abolfazl Zargari, Yue Du, Morteza Heidari, Theresa C Thai, Camille C Gunderson, Kathleen Moore, Robert S Mannel, Hong Liu, Bin Zheng, and Yuchen Qiu. Prediction of chemotherapy response in ovarian cancer patients using a new clustered quantitative image marker. *Physics in Medicine & Biology*, 63(15):155020, 2018.

[93] Balaji Ganeshan and Kenneth A Miles. Quantifying tumour heterogeneity with CT. *Cancer Imaging*, 13(1):140, 2013.

[94] Philippe Lambin, Ralph TH Leijenaar, Timo M Deist, Jurgen Peerlings, Evelyn EC De Jong, Janita Van Timmeren, Sebastian Sanduleanu, Ruben THM Larue, Aniek JG Even, Arthur Jochems, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology*, 14(12):749–762, 2017.

[95] Mitesh Patel, J Zhan, K Natarajan, Robert Flintham, Nigel Davies, Paul Sanghera, J Grist, V Duddalwar, A Peet, and V Sawlani. Machine learning-based radiomic evaluation of treatment response prediction in glioblastoma. *Clinical Radiology*, 76(8):628–e17, 2021.

[96] Neman Abdoli, Ke Zhang, Patrik Gilley, Xuxin Chen, Youkabed Sadri, Theresa Thai, Lauren Dockery, Kathleen Moore, Robert Mannel, and Yuchen Qiu. Evaluating the effectiveness of 2D and 3D CT image features for predicting tumor response to chemotherapy. *Bioengineering*, 10(11):1334, 2023.

[97] Ke Zhang, Neman Abdoli, Patrik Gilley, Youkabed Sadri, Xuxin Chen, Theresa C Thai, Lauren Dockery, Kathleen Moore, Robert S Mannel, and Yuchen Qiu. Developing a novel image marker to predict the clinical outcome of neoadjuvant chemotherapy (NACT) for ovarian cancer patients. *Computers in Biology and Medicine*, 172:108240, 2024.

[98] Yuan-Peng Zhang, Xin-Yun Zhang, Yu-Ting Cheng, Bing Li, Xin-Zhi Teng, Jiang Zhang, Saikit Lam, Ta Zhou, Zong-Rui Ma, Jia-Bao Sheng, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. *Military Medical Research*, 10(1):22, 2023.

[99] W Han, L Qin, C Bay, X Chen, K-H Yu, N Miskin, A Li, X Xu, and G Young. Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas. *American Journal of Neuroradiology*, 41(1):40–48, 2020.

[100] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[101] Yuming Jiang, Zhicheng Zhang, Wei Wang, Weicai Huang, Chuanli Chen, Sujuan Xi, M Usman Ahmad, Yulan Ren, Shengtian Sang, Jingjing Xie, et al. Biology-guided deep learning predicts prognosis and cancer immunotherapy response. *Nature Communications*, 14(1):5135, 2023.

[102] Rahul Paul, Samuel H Hawkins, Yoganand Balagurunathan, Matthew B Schabath, Robert J Gillies, Lawrence O Hall, and Dmitry B Goldgof. Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. *Tomography*, 2(4):388, 2016.

[103] Xin Zhen, Jiawei Chen, Zichun Zhong, Brian Hrycushko, Linghong Zhou, Steve Jiang, Kevin Albuquerque, and Xuejun Gu. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study. *Physics in Medicine & Biology*, 62(21):8246–8263, 2017.

[104] Blessed Ziyambe, Abid Yahya, Tawanda Mushiri, Muhammad Usman Tariq, Qaisar Abbas, Muhammad Babar, Mubarak Albathan, Muhammad Asim, Ayyaz Hussain, and Sohail Jabbar. A deep learning framework for the prediction and diagnosis of ovarian cancer in pre- and post-menopausal women. *Diagnostics*, 13 (10):1703, 2023.

[105] Ching-Wei Wang, Yu-Ching Lee, Cheng-Chang Chang, Yi-Jia Lin, Yi-An Liou, Po-Chao Hsu, Chun-Chieh Chang, Aung-Kyaw-Oo Sai, Chih-Hung Wang, and Tai-Kuang Chao. A weakly supervised deep learning method for guiding ovarian cancer treatment and identifying an effective biomarker. *Cancers*, 14(7):1651, 2022.

[106] Zongsheng Hu, Zhenyu Yang, Haozhao Zhang, Eugene Vaios, Kyle Lafata, Fang-Fang Yin, and Chunhao Wang. A deep learning model with radiomics analysis integration for glioblastoma post-resection survival prediction. *arXiv preprint arXiv:2203.05891*, 2022.

[107] Ursula A Matulonis, Amit M Oza, Tony W Ho, and Jonathan A Ledermann. Intermediate clinical endpoints: a bridge between progression-free survival and overall survival in ovarian cancer trials. *Cancer*, 121(11):1737–1746, 2015.

[108] Katrin M Sjoquist, Sarah J Lord, Michael L Friedlander, Robert John Simes, Ian C Marschner, and Chee Khoon Lee. Progression-free survival as a surrogate endpoint for overall survival in modern ovarian cancer trials: a meta-analysis. *Therapeutic Advances in Medical Oncology*, 10:1758835918788500, 2018.

[109] Johannes Schindelin, Curtis T Rueden, Mark C Hiner, and Kevin W Eliceiri. The ImageJ ecosystem: an open platform for biomedical image analysis. *Molecular Reproduction and Development*, 82(7-8):518–529, 2015.

[110] Bino Sebastian V, A Unnikrishnan, and Kannan Balakrishnan. Gray level co-occurrence matrices: generalisation and some new features. *arXiv preprint arXiv:1205.4831*, 2012.

[111] Chengjun Sun and William G Wee. Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing*, 23 (3):341–352, 1983.

[112] Belur V Dasarathy and Edwin B Holder. Image characterizations based on joint gray level—run length distributions. *Pattern Recognition Letters*, 12(8):497–502, 1991.

[113] Andrea Baraldi and Flavio Panniggiani. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2):293–304, 1995.

[114] Jinyu Yang, Jingjing Liu, Ning Xu, and Junzhou Huang. TVT: transferable vision transformer for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 520–530, 2023.

[115] Mostafa Dehghani, Josip Djolonga, Basil Mustafa, Piotr Padlewski, Jonathan Heek, Justin Gilmer, Andreas Peter Steiner, Mathilde Caron, Robert Geirhos, Ibrahim Alabdulmohsin, et al. Scaling vision transformers to 22 billion parameters. In *International Conference on Machine Learning*, pages 7480–7512. PMLR, 2023.

[116] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

[117] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16000–16009, 2022.

[118] Huan Liu and Hiroshi Motoda. *Feature selection for knowledge discovery and data mining*, volume 454. Springer Science & Business Media, 2012.

[119] Huan Liu and Hiroshi Motoda. *Feature selection for knowledge discovery and data mining*, volume 454. Springer Science & Business Media, 2012.

[120] Verónica Bolón-Canedo, Noelia Sánchez-Maroño, and Amparo Alonso-Betanzos. Feature selection for high-dimensional data. *Progress in Artificial Intelligence*, 5 (2):65–75, 2016.

[121] Trevor Hastie. The elements of statistical learning: data mining, inference, and prediction, 2009.

[122] Mariette Awad and Rahul Khanna. Support vector machines for classification. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, pages 39–66. Springer, 2015.

[123] Daniel Mesafint Belete and Manjaiah D Huchaiah. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *International Journal of Computers and Applications*, 44(9):875–886, 2022.

[124] Andrew P Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

[125] Kendrick Boyd, Kevin H Eng, and C David Page. Area under the precision-recall curve: point estimates and confidence intervals. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 451–466. Springer, 2013.

[126] Siti Fairus Mokhtar, Zahayu Md Yusof, and Hasimah Sapiri. Confidence intervals by bootstrapping approach: a significance review. *Malaysian Journal of Fundamental and Applied Sciences*, 19(1):30–42, 2023.

[127] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 2017.

[128] Wang Xiang, Hui Guo, Yan Sun, Xia Jin, Hucheng Li, and Shuang Xia. Tumor stroma ratio-based radiomics model for predicting platinum resistance and prognosis in epithelial ovarian cancer. *Journal of Ovarian Research*, 19(1):106, 2026.

[129] Xiaoyu Huang, Yong Huang, Kexin Liu, Fenglin Zhang, Zhou Zhu, Kai Xu, and Ping Li. Intratumoral and peritumoral radiomics for predicting the prognosis of high-grade serous ovarian cancer patients receiving platinum-based chemotherapy. *Academic Radiology*, 32(2):877–887, 2025.

[130] Mengge He, Rahul Singh, Mandi Wang, Grace Ho, Esther MF Wong, Keith WH Chiu, Anthony KT Leung, Ka Yu Tse, Philip PC Ip, Andy Hwang, et al. CT-based radiomics model to predict platinum sensitivity in epithelial ovarian carcinoma: a multicentre study. *Cancer Imaging*, 25(1):85, 2025.

[131] David D Bowtell, Steffen Böhm, Ahmed A Ahmed, Paul-Joseph Aspuria, Robert C Bast Jr, Valerie Beral, Jonathan S Berek, Michael J Birrer, Sarah Blagden, Michael A Bookman, et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nature Reviews Cancer*, 15 (11):668–679, 2015.

[132] Stephanie Lheureux, Charlie Gourley, Ignace Vergote, and Amit M Oza. Epithelial ovarian cancer. *The Lancet*, 393(10177):1240–1253, 2019.

[133] Lindsay Kuroki and Saketh R Guntupalli. Treatment of epithelial ovarian cancer. *BMJ*, 371, 2020.

[134] Roland F Schwarz, Charlotte KY Ng, Susanna L Cooke, Scott Newman, Jillian Temple, Anna M Piskorz, Davina Gale, Karen Sayal, Muhammed Murtaza, Peter J Baldwin, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS Medicine*, 12(2):e1001789, 2015.

[135] Andrew McPherson, Andrew Roth, Emma Laks, Tehmina Masud, Ali Bashashati, Allen W Zhang, Gavin Ha, Justina Biele, Damian Yap, Adrian Wan, et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nature Genetics*, 48(7):758–767, 2016.

[136] Ahmed Ashour Ahmed, Dariush Etemadmoghadam, Jillian Temple, Andy G Lynch, Mohamed Riad, Raghwa Sharma, Colin Stewart, Sian Fereday, Carlos Caldas, Anna DeFazio, et al. Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *The Journal of Pathology*, 221(1):49–56, 2010.

[137] Ann-Marie Patch, Elizabeth L Christie, Dariush Etemadmoghadam, Dale W Garsed, Joshy George, Sian Fereday, Katia Nones, Prue Cowin, Kathryn Alsop, Peter J Bailey, et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature*, 521(7553):489–494, 2015.

[138] Douglas Hanahan et al. Hallmarks of cancer: new dimensions. *Cancer Discovery*, 12(1):31–46, 2022.

[139] National Comprehensive Cancer Network et al. NCCN clinical practice guidelines in oncology. *http://www. nccn. org/professionals/physician_gls/PDF/occult. pdf*, 2008.

[140] Andreas Du Bois, Alexander Reuss, Eric Pujade-Lauraine, Philipp Harter, Isabelle Ray-Coquard, and Jacobus Pfisterer. Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials: by the Arbeitsgemeinschaft Gynaekologische Onkologie Studiengruppe Ovarialkarzinom (AGO-OVAR) and the Groupe d'Investigateurs Nationaux Pour les Etudes des Cancers de l'Ovaire (GINECO). *Cancer*, 115(6):1234–1244, 2009.

[141] Ignace Vergote, Claes G Tropé, Frédéric Amant, Gunnar B Kristensen, Tom Ehlen, Nick Johnson, René HM Verheijen, Maria EL Van Der Burg, Angel J Lacave, Pierluigi Benedetti Panici, et al. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *New England Journal of Medicine*, 363(10):943–953, 2010.

[142] Kaustav Bera, Kurt A Schalper, David L Rimm, Vamsidhar Velcheti, and Anant Madabhushi. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology*, 16(11):703–715, 2019.

[143] Balázs Acs, Mattias Rantalainen, and Johan Hartman. Artificial intelligence as the next step towards precision pathology. *Journal of Internal Medicine*, 288(1): 62–81, 2020.

[144] Andrew H Song, Guillaume Jaume, Drew FK Williamson, Ming Y Lu, Anurag Vaidya, Tiffany R Miller, and Faisal Mahmood. Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering*, 1(12):930–949, 2023.

[145] Wenya Linda Bi, Ahmed Hosny, Matthew B Schabath, Maryellen L Giger, Nicolai J Birkbak, Alireza Mehrtash, Tavis Allison, Omar Arnaout, Christopher Abbosh, Ian F Dunn, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA: A Cancer Journal for Clinicians*, 69(2):127–157, 2019.

[146] Jana Lipkova, Richard J Chen, Bowen Chen, Ming Y Lu, Matteo Barbieri, Daniel Shao, Anurag J Vaidya, Chengkuan Chen, Luoting Zhuang, Drew FK Williamson, et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell*, 40(10):1095–1110, 2022.

[147] Aya M Al-Zoghby, Ahmed Ismail Ebada, Aya S Saleh, Mohammed Abdelhay, and Wael A Awad. A comprehensive review of multimodal deep learning for enhanced medical diagnostics. *Computers, Materials & Continua*, 84(3):4155–4193, 2025.

[148] Shih-Cheng Huang, Anuj Pareek, Saeed Seyyedi, Imon Banerjee, and Matthew P Lungren. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *npj Digital Medicine*, 3(1):136, 2020.

[149] Artem Shmatko, Narmin Ghaffari Laleh, Moritz Gerstung, and Jakob Nikolas Kather. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nature Cancer*, 3(9):1026–1038, 2022.

[150] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma

Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

[151] Michael Moor, Oishi Banerjee, Zahra Shakeri Hossein Abad, Harlan M Krumholz, Jure Leskovec, Eric J Topol, and Pranav Rajpurkar. Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956):259–265, 2023.

[152] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, 2024.

[153] Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862, 2024.

[154] Adam Goode, Benjamin Gilbert, Jan Harkes, Drazen Jukic, and Mahadev Satyanarayanan. OpenSlide: a vendor-neutral software foundation for digital pathology. *Journal of Pathology Informatics*, 4(1):27, 2013.

[155] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[156] Pierre Soille et al. *Morphological image analysis: principles and applications*, volume 2. Springer, 1999.

[157] Said Pertuz, Domenec Puig, and Miguel Angel Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5):1415–1432, 2013.

[158] Zhi Huang, Federico Bianchi, Mert Yuksekgonul, Thomas J Montine, and James Zou. A visual–language foundation model for pathology image analysis using medical Twitter. *Nature Medicine*, 29(9):2307–2316, 2023.

[159] Sheng Zhang, Yanbo Xu, Naoto Usuyama, Hanwen Xu, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, Mu Wei, Naveen Valluri, et al. BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. *arXiv preprint arXiv:2303.00915*, 2023.

[160] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3992–4003. IEEE, 2023.

[161] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg

Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[162] Sudhir Varma and Richard Simon. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7(1):91, 2006.

[163] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1):389–422, 2002.

[164] David W Hosmer, Stanley Lemeshow, Rodney X Sturdivant, et al. *Applied logistic regression*, volume 2. Wiley, 2000.

[165] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: a survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443, 2018.

[166] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[167] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27 (8):861–874, 2006.

[168] Elizabeth R DeLong, David M DeLong, and Daniel L Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845, 1988.

[169] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

[170] Wojciech Samek and Klaus-Robert Müller. Towards explainable artificial intelligence. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, pages 5–22. Springer, 2019.

[171] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.

[172] Maryam Asadi-Aghbolaghi, Amirali Darbandsari, Allen Zhang, Alberto Contreras-Sanz, Jeffrey Boschman, Pouya Ahmadvand, Martin Köbel, David Farnell, David G Huntsman, Andrew Churg, et al. Learning generalizable AI models for multi-center histopathology image classification. *npj Precision Oncology*, 8(1):151, 2024.

[173] Frederick Klauschen, Jonas Dippel, Philipp Keyl, Philipp Jurmeister, Michael Bockmayr, Andreas Mock, Oliver Buchstab, Maximilian Alber, Lukas Ruff, Grégoire Montavon, et al. Toward explainable artificial intelligence for precision pathology. *Annual Review of Pathology: Mechanisms of Disease*, 19(1):541–570, 2024.

[174] Y. Li, L. Pan, and Y. Peng. Application of deep learning-based multimodal fusion technology in cancer diagnosis: a survey. *Engineering Applications of Artificial Intelligence*, 143(109972), 2025. doi: 10.1016/j.engappai.2024.109972.

[175] J. Hou, R. Zhang, Y. Xie, C. Li, and W. Qin. Multimodal deep learning for cancer prognosis prediction with clinical information prompts integration. *npj Digital Medicine*, 9(76), 2026. doi: 10.1038/s41746-025-02257-y. The article was published online in December 2025 and assigned to the 2026 volume.

[176] M. Del Grande, S. Rizzo, and G. M. Nicolino. Computed tomography-based body composition in patients with ovarian cancer: association with chemotoxicity and prognosis. *Frontiers in Oncology*, 11(718815), 2021. doi: 10.3389/fonc.2021.718815.

[177] J. Ubachs, J. Ziemons, and I.J.G. Minis-Rutten. Sarcopenia and ovarian cancer survival: a systematic review and meta-analysis. *Journal of Cachexia, Sarcopenia and Muscle*, 10(6):1165–1174, 2019. doi: 10.1002/jcsm.12468.

[178] D.D.B. Bates and P.J. Pickhardt. CT-derived body composition assessment as a prognostic tool in oncologic patients: from opportunistic research to artificial intelligence-based clinical implementation. *American Journal of Roentgenology*, 219(4):671–680, 2022. doi: 10.2214/AJR.22.27749.

[179] H. Guan and M. Liu. Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2022. doi: 10.1109/TBME.2021.3117407.

[180] J. Wong, M. Baine, and S. Wisnoskie. Effects of interobserver and interdisciplinary segmentation variabilities on CT-based radiomics for pancreatic cancer. *Scientific Reports*, 11(16328), 2021. doi: 10.1038/s41598-021-95152-x.

[181] W. Kuerbanjiang, X. Wang, Y. Jiamaliding, M. Maimaitiaili, and Y. Yi. Multidisciplinary large language model agent teams for precision oncology enhance complex gynecologic oncology decision support. *medRxiv*, 2025. doi: 10.1101/2025.10.30.25339199. This is a preprint and has not yet undergone peer review.

[182] J. Zoller, M. Kalz, and X. Wu. Clinical decision support in hematological malignancies using a case-grounded AI agent. *Nature Medicine*, 30, 2026. doi: 10.1038/s41591-026-04494-4. Published online June.

[183] X. Wang, Q. Wang, and G. Ding. Artificial intelligence in multidisciplinary tumor boards enhancing decision making and clinical outcomes in oncology. *iScience*, 28(12), 2025. doi: 10.1016/j.isci.2025.114082.