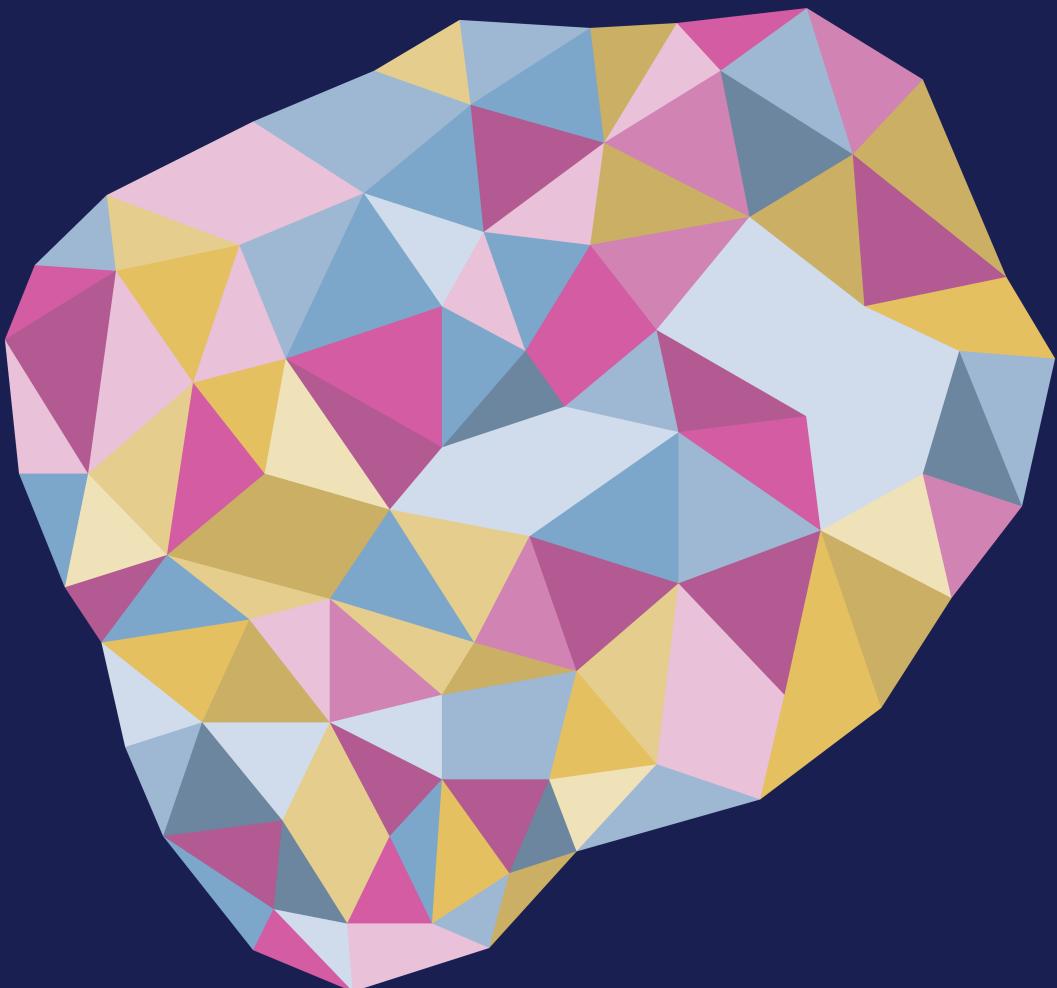


# Quantifying Tumor Heterogeneity in Colorectal Liver Metastases with CT-Derived Habitats



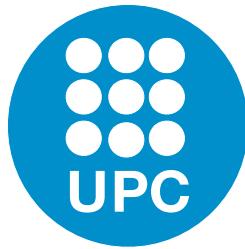
Olivia Prior



Cover design by scientist and scientific illustrator Maria Balaguer-Montero, Barcelona, January 2026.

Commissioned for the price of one extraordinarily large hug.





PhD Program in Biomedical Engineering

Universitat Politècnica de Catalunya

# **Quantifying Tumor Heterogeneity in Colorectal Liver Metastases with CT-Derived Habitats**

*By*

**Olivia Prior**

**Thesis advisor:** Raquel Perez-Lopez, MD, PhD

**Academic tutor:** Eduardo Soudah, PhD

A thesis submitted in partial fulfillment of the requirements  
for the degree of *Doctor of Philosophy*

Barcelona, February 2026



## **Olivia Prior**

*Email:* theanega@gmail.com

*Website:* theanega.github.io

*ORCID ID:* 0009-0000-4913-6716

© Olivia Prior 2026

## **FUNDING**

The research presented in this thesis dissertation was financially supported by a CRIS Foundation Talent Award (TALENT19-05), the FERO Foundation, the Instituto de Salud Carlos III-Investigación en Salud (PI18/01395 and PI21/01019), the Prostate Cancer Foundation (18YOUN19) and the Asociación Española Contra el Cáncer (AECC) (PRYCO211023SERR). All of these grants were awarded to Dr. Raquel Perez-Lopez as principal investigator. Olivia Prior was financially supported by a Fundació “LaCaixa” IN-PhINIT Fellowship (ID 100010434, code LCF/BQ/DR21/11880027). The international stay at the Early Cancer Institute (Cambridge, UK) was covered by a European Association for Cancer Research Travel Fellowship (ID 942).



*Esta tesis está dedicada a todas las personas que conviven con una enfermedad, y también a sus familias. Cuando la salud nos abandona, toda la vida gira en torno a recuperarla. A través de la ciencia intentamos allanar ese camino y ayudar a encontrarla.*



*Aquesta tesi està dedicada a totes les persones que conviuen amb una malaltia, i també a les seves famílies. Quan la salut ens abandona, tota la vida gira entorn de recuperar-la. A través de la ciència intentem aplanar aquest camí i ajudar a trobar-la.*



*This thesis is dedicated to all people who live with an illness, and also to their families. When health abandons us, life revolves around recovering it. Through science we try to smooth that path and help find it.*



# Abstract

In colorectal cancer, mortality is primarily driven by metastatic disease, most commonly to the liver. Colorectal liver metastases are biologically heterogeneous, composed of varying proportions of viable tumor cells, fibrosis, and necrosis. Both the presence and spatial organization of these tissue components influence treatment response and patient outcome. In clinical practice, however, such information is available only through biopsy, which samples a limited tumor region, or after surgical resection, which is feasible in only a minority of patients. As a result, most patients receive systemic therapy without direct knowledge of whole-tumor tissue composition.

Computed tomography (CT) is the standard imaging modality for managing colorectal liver metastases and is acquired repeatedly throughout the disease course. Despite providing non-invasive, spatially resolved information at the whole-tumor level, clinical interpretation largely focuses on lesion size, number, and location. Consequently, the ability of CT to characterize intratumor heterogeneity has not yet been fully leveraged.

Habitat imaging has been proposed as a framework for capturing tumor heterogeneity by partitioning tumors into spatial subregions with similar imaging properties (habitats). Most habitat imaging studies rely on multiparametric MRI (mpMRI), whose quantitative maps are biologically interpretable, while applications to CT remain limited despite its widespread clinical use. In addition, existing studies rarely assess the robustness of CT-derived features, often define habitats through purely data-driven optimization without biological grounding, and generally do not report clinical relevance beyond tumor volume.

This thesis addresses these gaps by investigating whether routine CT can capture biologically and clinically meaningful intratumor heterogeneity in colorectal liver metastases. First, we identified 26 radiomic features suitable for robust CT-based habitat computation based on repeatability and reproducibility criteria. Second, we developed a biologically anchored CT habitat model by incorporating co-registered mpMRI as a reference during habitat definition, rather than relying solely on statistical optimization. Within this framework, we compared multiple CT representations, including handcrafted radiomic features and deep learning embeddings, and found that handcrafted features produced more biologically coherent habitats. The resulting three habitats reflected vascular architecture: an avascular core, a cellular, perfused intermediate zone, and a highly vascularized outer rim.

Third, we evaluated clinical relevance by assessing associations between habitat-derived metrics and patient outcomes in two independent cohorts. Habitat metrics provided prognostic information beyond tumor volume, but only in specific treatment contexts. In particular, habitat entropy at the tumor–liver interface predicted survival in settings in which treatment may alter tissue composition without inducing

measurable size changes. Across all treatment contexts, prognostic information consistently localized to the invasive tumor rim rather than being uniformly distributed throughout the lesion.

Overall, this thesis contributes both methodological and clinical advances: an open-source CT habitat imaging pipeline, a comprehensive assessment of handcrafted radiomic feature robustness, the first comparison of CT representations for habitat computation, biological grounding of CT-derived habitats using mpMRI, and demonstration of their context-dependent clinical relevance. Together, these findings establish that routine CT scans contain clinically meaningful information about tumor heterogeneity that current assessment strategies do not capture, and provide a framework for extracting it.

**Keywords:** liver metastases, colorectal cancer, habitat imaging, intratumor heterogeneity, radiomics, computed tomography, multiparametric magnetic resonance imaging, precision oncology

# Resumen

En el cáncer colorrectal, la mortalidad está impulsada principalmente por la enfermedad metastásica, más comúnmente al hígado. Las metástasis hepáticas de cáncer colorrectal son biológicamente heterogéneas y están compuestas por proporciones variables de células tumorales viables, fibrosis y necrosis. Tanto la presencia como la organización espacial de estos componentes tisulares influyen en la respuesta al tratamiento y en el pronóstico del paciente. Sin embargo, en la práctica clínica, esta información solo está disponible mediante biopsia, que evalúa una región tumoral limitada, o tras la resección quirúrgica, que solo es factible en una minoría de pacientes. Como resultado, la mayoría de los pacientes recibe tratamiento sistémico sin conocimiento directo de la composición tisular del tumor completo.

La tomografía computarizada (TC) es la modalidad de imagen estándar para el manejo de las metástasis hepáticas de cáncer colorrectal y se adquiere de forma repetida a lo largo de la evolución de la enfermedad. A pesar de proporcionar información no invasiva y espacialmente resuelta a nivel de todo el tumor, la interpretación clínica continúa centrándose principalmente en el tamaño, número y localización de las lesiones. En consecuencia, la capacidad de la TC para caracterizar la heterogeneidad intratumoral aún no se ha aprovechado plenamente.

El análisis de hábitats de imagen se ha propuesto como una herramienta para capturar la heterogeneidad tumoral mediante la partición de los tumores en subregiones espaciales con propiedades de imagen similares (hábitats). La mayoría de los estudios de imagen de hábitats se basan en resonancia magnética multiparamétrica (RMmp), cuyos mapas cuantitativos son biológicamente interpretables, mientras que las aplicaciones en TC siguen siendo limitadas a pesar de su amplio uso clínico. Además, los estudios existentes rara vez evalúan la robustez de las características radiómicas derivadas de TC, a menudo definen los hábitats mediante optimización puramente basada en datos sin una base biológica, y generalmente no informan sobre su relevancia clínica más allá del volumen tumoral.

Esta tesis aborda estas limitaciones investigando si la TC de rutina puede capturar heterogeneidad intratumoral biológica y clínicamente relevante en las metástasis hepáticas de cáncer colorrectal. En primer lugar, identificamos 26 características radiómicas adecuadas para el cálculo robusto de hábitats basados en TC, según criterios de repetibilidad y reproducibilidad. En segundo lugar, desarrollamos un modelo de hábitats en TC con anclaje biológico incorporando RMmp coregistrada como referencia durante la definición de los hábitats, en lugar de basarnos exclusivamente en la optimización estadística. Dentro de este marco, comparamos múltiples representaciones de TC, incluidas características radiómicas artesanales y representaciones obtenidas mediante aprendizaje profundo, y observamos que las características artesanales producían hábitats biológicamente más coherentes. Los tres hábitats resultantes reflejaron la

arquitectura vascular: un núcleo avascular, una zona intermedia celular y perfundida, y un borde externo altamente vascularizado.

En tercer lugar, evaluamos la relevancia clínica analizando la asociación entre métricas derivadas de los hábitats y los resultados clínicos en dos cohortes independientes. Las métricas de hábitat proporcionaron información pronóstica más allá del volumen tumoral, pero solo en contextos terapéuticos específicos. En particular, la entropía del hábitat en la interfaz tumor–hígado predijo la supervivencia en situaciones en las que el tratamiento puede alterar la composición tisular sin inducir cambios medibles en el tamaño. En todos los contextos terapéuticos, la información pronóstica se localizó de forma consistente en el borde tumoral invasivo, en lugar de distribuirse uniformemente en toda la lesión.

En conjunto, esta tesis aporta avances tanto metodológicos como clínicos: un pipeline de imagen de hábitats en TC de código abierto, una evaluación exhaustiva de la robustez de las características radiómicas artesanales, la primera comparación de representaciones de TC para el cálculo de hábitats, el anclaje biológico de los hábitats derivados de TC utilizando RMmp y la demostración de su relevancia clínica dependiente del contexto. En conjunto, estos hallazgos establecen que las TC de rutina contienen información clínicamente relevante sobre la heterogeneidad tumoral que las estrategias actuales de evaluación no capturan, y proporcionan un marco para su extracción.

**Palabras clave:** metástasis hepáticas, cáncer colorrectal, análisis de hábitats de imagen, heterogeneidad intratumoral, radiómica, tomografía computarizada, resonancia magnética multiparamétrica, oncología de precisión

# Resum

En el càncer colorectal, la mortalitat està impulsada principalment per la malaltia metastàtica, més habitualment al fetge. Les metàstasis hepàtiques del càncer colorectal són biològicament heterogènies i estan compostes per proporcions variables de cèl·lules tumorals viables, fibrosi i necrosi. Tant la presència com l'organització espacial d'aquests components tissulars influeixen en la resposta al tractament i en el pronòstic del pacient. Tanmateix, en la pràctica clínica, aquesta informació només està disponible mitjançant biòpsia, que evalua una regió tumoral limitada, o després de la resecció quirúrgica, que només és factible en una minoria de pacients. Com a resultat, la majoria dels pacients rep tractament sistèmic sense coneixement directe de la composició tissular del tumor complet.

La tomografia computada (TC) és la modalitat d'imatge estàndard per al maneig de les metàstasis hepàtiques del càncer colorectal i s'adquireix de manera repetida al llarg de l'evolució de la malaltia. Tot i proporcionar informació no invasiva i amb resolució espacial a nivell de tot el tumor, la interpretació clínica continua centrant-se principalment en la mida, el nombre i la localització de les lesions. En conseqüència, la capacitat de la TC per caracteritzar la heterogeneïtat intratumoral encara no s'ha aprofitat plenament.

L'anàlisi d'hàbitats d'imatge s'ha proposat com una eina per capturar la heterogeneïtat tumoral mitjançant la partició dels tumors en subregions espacials amb propietats d'imatge similars (hàbitats). La majoria dels estudis d'hàbitats d'imatge es basen en la ressonància magnètica multiparamètrica (RMmp), els mapes quantitatius de la qual són biològicament interpretables, mentre que les aplicacions en TC continuen sent limitades malgrat el seu ampli ús clínic. A més, els estudis existents rarament avaluen la robustesa de les característiques radiomíiques derivades de la TC, sovint defineixen els hàbitats mitjançant una optimització purament basada en dades sense una base biològica, i generalment no informen sobre la seva rellevància clínica més enllà del volum tumoral.

Aquesta tesi aborda aquestes limitacions investigant si la TC de rutina pot capturar una heterogeneïtat intratumoral biològica i clínicament rellevant en les metàstasis hepàtiques del càncer colorectal. En primer lloc, vam identificar 26 característiques radiomíiques adequades per al càlcul robust d'hàbitats basats en TC, segons criteris de repetibilitat i reproductibilitat. En segon lloc, vam desenvolupar un model d'hàbitats en TC amb ancoratge biològic incorporant la RMmp coregistrada com a referència durant la definició dels hàbitats, en lloc de basar-nos exclusivament en l'optimització estadística. Dins d'aquest marc, vam comparar múltiples representacions de la TC, incloent-hi característiques radiomíiques artesanals i representacions obtingudes mitjançant aprenentatge profund, i vam observar que les característiques artesanals produïen hàbitats biològicament més coherents. Els tres hàbitats resultants reflectien

l’arquitectura vascular: un nucli avascular, una zona intermèdia cel·lular i perfosa, i un marge extern altament vascularitzat.

En tercer lloc, vam avaluar la rellevància clínica analitzant l’associació entre mètriques derivades dels hàbitats i els resultats clínics en dues cohorts independents. Les mètriques d’hàbitat van proporcionar informació pronòstica més enllà del volum tumoral, però només en contextos terapèutics específics. En particular, l’entropia de l’hàbitat a la interfície tumor–fetge va predir la supervivència en situacions en què el tractament pot alterar la composició tissular sense induir canvis mesurables en la mida. En tots els contextos terapèutics, la informació pronòstica es va localitzar de manera consistent al marge tumoral invasiu, en lloc de distribuir-se uniformement a tota la lesió.

En conjunt, aquesta tesi aporta avenços tant metodològics com clínics: un *pipeline* d’imatge d’hàbitats en TC de codi obert, una evaluació exhaustiva de la robustesa de les característiques radiomiques artesanals, la primera comparació de representacions de TC per al càlcul d’hàbitats, l’ancoratge biològic dels hàbitats derivats de la TC utilitzant RMmp i la demostració de la seva rellevància clínica dependent del context. En conjunt, aquests resultats estableixen que les TC de rutina contenen informació clínicament rellevant sobre la heterogeneïtat tumoral que les estratègies actuals d’avaluació no capten, i proporcionen un marc per a la seva extracció.

**Paraules clau:** metàstasis hepàtiques, càncer colorectal, anàlisi d’hàbitats d’imatge, heterogeneïtat intra-tumoral, radiòmica, tomografia computada, ressonància magnètica multiparamètrica, oncologia de precisió

# Acknowledgements

Cuando pienso en todas las personas que han ayudado a que este documento cobre vida durante los últimos cuatro años, me doy cuenta de que las palabras no bastan para expresar plenamente la gratitud que siento. Lo que sigue es, por tanto, un agradecimiento necesariamente incompleto.

Gracias a mi directora de tesis, **Raquel Pérez-López**, por “quererme con o sin beca”, por confiar en mí en un momento pospandemia en el que muchas puertas estaban cerradas y por darme la oportunidad de crecer con libertad. Admiro profundamente tu capacidad de trabajo y no puedo esperar a ver todo lo que el grupo logrará en las próximas décadas.

Thank you to my former co-director, **Kinga Bernatowicz**. Beginnings are hard, and you were the first person to patiently teach me what habitats were. Your early enthusiasm laid the foundation for everything that followed.

I am eternally grateful to **Francesco Grussu**. You treated me as a scientific peer from day one, encouraged me to submit habitats research to conferences, and spent countless hours patiently discussing experiments with me and teaching me how to write convincingly about them. You pushed me to think deeper, first by sitting next to you in the early days (one learns a lot by watching how someone works), and later by working together. In every single meeting, you managed two small miracles: convincing me that the work I was doing was important, and convincing me that I could do it. You are a true mentor, and any future impact of my research will be, in large part, a testament to your guidance. E, a proposito, il mio cervello è infinitamente migliore dopo aver ti incontrato, ma il mio fegato... forse meno.

Gracias a **Marta Ligero**, por allanar el camino para los radiodoctorandos y por ser la primera mano amiga.

Gracias a **Alonso García**, por hacer del laboratorio un lugar más generoso y humano, y por cuidarme con palabras y memes a partes iguales.

Thank you to **Christina Zatse**, for consistently and kindly reminding me that I am loved.

Cuando una escoge hacer el doctorado, elige sobre todo proyecto y supervisión, pero no a sus compañeros de laboratorio. En este sentido, me considero extraordinariamente afortunada. Gracias de corazón al **grupo de Radiomics**, pasado y presente: **Adrià Marcos, Anna Voronova, Athanasios Grigoriou, Bente Gielen, Camilo Monreal, Carlos Macarro, Caterina Tozzi, Daniel Navarro, Ella Fokkinga, Humaira Abdul, Iker Zubietta, Ingrid Cayuela, Laia Coronas, Luz Atlagich, Maria Balaguer, Marta Buetas, Nikos Staikoglou, Óscar Llorian and Rosa Melano**. Thank you for the radiobeers, VHIOFindes, Valladolid, Amsterdam, Porto, Vienna, the cantine lunches, the Christmas dinners, and the endless dis-

cussions in lab meetings and project meetings, where much of this thesis was born. Every idea described here was the product of robust discussion between us. Your collective wisdom has been invaluable in shaping both my research and my growth as a scientist. Thank you for the million memories I will carry with me forever.

My work would not have been possible without many collaborators: **Emily Latacz, Garazi Serna, Javi Ros, Kate Connor, Maria Teresa Salcedo, Mireia Sanchís, Nadia Saoudi, Nuria Ruiz, Peter Vermeulen, Raquel Comas, Víctor Navarro, and Zynya Calixto**. Their contributions have been truly invaluable. Special thanks to **Francesc Salvà**, for always being available and for insisting that good research starts with the right clinical questions.

I am also deeply grateful to all the people at **VHIO** who influenced me through stimulating conversations and shared their vision, making it a truly special place to do a PhD. This includes **VHIO Academy**, especially **Imma Falero**, thank you for always listening and for the coffees at Noni. Thank you to **Alex Mur, Agustín Sánchez, Andreu Òdena, Ana Dueso, Andrea Herencia, Ariadna Grinyó, Carmen Escudero, Cayetano Galera, Débora Cabot, Ester Aguado, Iñigo González, María López, Marion Martínez, Mariona Cubells, Paula Carrillo, and Setareh Kompanian**, for friendship, knowledge, and adventure. It has been an honor and a privilege to get to know you all. Special thanks to **Ariadna** and **Marion** for the early days of the PhD council and for thinking together about how to build a better community.

Thank you to the **Mireia Crispin-Ortuzar lab** at the University of Cambridge for keeping me warm during the Cambridge autumn and winter. Thank you to **Inés Machado**, for the Mediterranean vibe; **Rebecca Wray**, for running under the rain while having most interesting conversations; and **Emanuela Greco**, for hosting me without me ever having to ask twice. And a major thank you to **Fabio Giuntini**, for turning a shared flat into a home and for feeding me for three months.

Gracias a la **Fundación “la Caixa”** por financiar la mayor parte de esta investigación y por darme la oportunidad de asistir a congresos donde tanto aprendí. Gracias también a todas las personas que me animaron y ayudaron a pedir la beca cuando estaba sin trabajo. En especial, gracias a mi amiga **Sara Barrera**, porque nunca tuviste ninguna duda.

Gracias a mi amiga más cara, **Ares Anfruns**, por hacer tan bien tu trabajo.

Thank you to my faraway friends for the visits: **Danielle Sher, Emma Lower, Fang Wen Lim, Gabe Velenosi, Guillaume Varvoux, Jorge Zaera, Lindsey Milisits, Marie Rajon and Rushabh Malde**. Special mention to **Rushabh, Fang, and Jorge**, for the cookies and the love... ten years and still counting.

Gracias a mis amigos de toda la vida por acompañarme sin condiciones y por las aventuras que siempre me recuerdan de dónde vengo: **Albert Lázaro, Alex Paredes, Blanca Fornesa, Blanca Montamat, Beltrán Jiménez, Laura Martín, Natalia Ollé, Núria Rodríguez, Pablo Jimeno, Pat Balaguer, Rafael Jaén, y Sara Alberch**.

Gràcies a la cinquena promoció d'Enginyeria Biomèdica de la Universitat de Barcelona: **Àlex Fajas, Alba Iruela, Anna Graell, Daniel Rodríguez, Enrique Vilalta, Genís Esplugas, Gerard Trias, Helena Briegas, Joan Ortiz, Judit Giró, Júlia Soler, Laura Escot, Mariona Ruiz, Marta Pérez, Marta**

**Trullols, i Umi Mir.** Vau convertir la universitat en una cosa molt més gran que una etapa acadèmica, i entre riures vau construir una base sòlida sobre la qual es sosté gran part del que sóc avui.

Por último, quiero agradecer especialmente esta tesis a mi familia.

Gràcies, **Padrí i Irene**, per totes les visites al Papi i per ser un model de la vida que vull tenir.

A **Laura y Pat**, por aprender a ser adultas juntas.

A **Camila y Nala**, por el amor incondicional.

A mis cuñadas y cuñados, **Mari Carmen, Adri, Juana Mari y Miguel Tercero**, gracias por hacerme tía y madrina y por enseñarme tanto. Si llego a saber que me haríais un sobrino por año de tesis, lo alargo todo un par de años más. Gracias **Héctor, Lluc y Leo**, por llenarnos la vida de alegría.

A mis suegros, **Mari y Miguel**, gracias por apoyarme en los buenos y malos momentos, por quererme tanto y tan bien, y por darme la oportunidad de empezar un hogar con vuestro hijo.

Al meu cunyat estimat, **Enric**, per totes les converses interessants i per arrencar-me somriures, sobretot amb cafès a Pamplona i dièsel a Saragossa.

Gracias a mi hermana **Virginia**, por ser siempre el lugar donde pensar en voz alta.

Gracias a mis padres, **María Gracia y Enrique**. Seguramente lo más relevante que haga en mi carrera profesional es ser hija vuestra. Gracias por inculcarme los valores de esfuerzo, respeto, honestidad y excelencia. Treballar dur en una cosa que m'agrada i intentar fer-ho bé és una manera de viure en la qual sempre em sento a prop teu, Papi.

Gracias a mi futuro (ex)marido, **Miguel**, por aguantarme en todas mis versiones. Me alegra saber que tengo el resto de mi vida para devolverte todo lo que te debo, porque necesitaré como mínimo todo ese tiempo.

No se me habría ocurrido hacer un doctorado si no hubiera tenido la suerte de descubrir la ciencia por primera vez de la mano de profesoras maravillosas. Qué importante es tener docentes en el colegio que guían, despiertan la curiosidad y plantan una semilla que crece con los años. Gracias a **Julie Connolly** (Science), **Hortènsia Mallén** (Biología), **Gloria López-Barrena** (Química) y **Ana Fuertes** (Física). No me olvido de vuestras clases.

Olivia Prior  
Barcelona, enero de 2026



# Contents

<b>Abstract</b>	vii
<b>Resumen</b>	ix
<b>Resum</b>	xi
<b>Acknowledgements</b>	xiii
<b>List of Figures</b>	xxi
<b>List of Tables</b>	xxiii
<b>I Introduction . . . . .</b>	<b>1</b>
<b>1 Motivation, Objectives, and Contributions</b>	<b>3</b>
1.1 Motivation . . . . .	3
1.2 Hypothesis and Objectives . . . . .	5
1.3 Contributions to Knowledge . . . . .	6
1.4 Thesis Structure . . . . .	8
<b>II Foundations . . . . .</b>	<b>9</b>
<b>2 Medical Image Techniques: An Overview</b>	<b>11</b>
2.1 Computed Tomography . . . . .	11
2.2 Magnetic Resonance Imaging . . . . .	13
2.2.1 Physical Principles . . . . .	13
2.2.2 Quantitative MRI . . . . .	14
2.3 Histopathologic Light Microscopy . . . . .	16
<b>3 Colorectal Liver Metastases</b>	<b>19</b>
3.1 Background . . . . .	19
3.2 Histopathological Heterogeneity . . . . .	20
3.3 Treatment Landscape and Selection Challenges . . . . .	22
3.3.1 Resectable disease . . . . .	22
3.3.2 Unresectable disease . . . . .	22
3.4 Response Assessment and Its Limitations . . . . .	24
<b>4 Imaging Heterogeneity: State of the Art</b>	<b>27</b>
4.1 Definition and Historical Context . . . . .	27

## CONTENTS

4.2	Qualitative and Conventional Quantitative Approaches . . . . .	29
4.3	Radiomics . . . . .	30
4.4	Habitat Imaging . . . . .	33
<b>III</b>	<b>Methodological Framework . . . . .</b>	<b>37</b>
<b>5</b>	<b>Data and Habitat Imaging Pipeline</b>	<b>39</b>
5.1	Patient Cohorts . . . . .	39
5.1.1	CT Image Preprocessing . . . . .	40
5.1.2	Multiparametric MRI Preprocessing (PREDICT) . . . . .	41
5.1.3	Histopathology Workflow (POEM) . . . . .	43
5.2	Habitat Imaging Pipeline . . . . .	43
5.2.1	Pipeline Overview . . . . .	44
5.2.2	Key Implementation Choices . . . . .	44
5.2.3	Habitat-Derived Quantitative Metrics . . . . .	47
<b>IV</b>	<b>A Data-Driven Approach to CT Habitats . . . . .</b>	<b>49</b>
<b>6</b>	<b>Identification of Precise Handcrafted Features</b>	<b>51</b>
6.1	Rationale . . . . .	52
6.2	Methods . . . . .	53
6.2.1	Patient Cohorts . . . . .	53
6.2.2	Image Segmentation and Perturbation . . . . .	53
6.2.3	Radiomics Feature Computation . . . . .	53
6.2.4	Precision Analysis: Repeatability and Reproducibility . . . . .	55
6.2.5	Habitat Computation and Stability Assessment . . . . .	55
6.2.6	Exploratory Biological Case Study . . . . .	56
6.2.7	Statistical Analysis . . . . .	56
6.3	Results . . . . .	57
6.3.1	Precision Analysis . . . . .	57
6.3.2	Identification of Precise Features . . . . .	58
6.3.3	Imaging Habitats Computed with Precise Features Show Increased Stability . . . . .	61
6.3.4	Exploratory Biological Case Study . . . . .	63
6.4	Discussion . . . . .	64
6.5	Summary . . . . .	66
<b>7</b>	<b>Development and Validation of an mpMRI-anchored CT Habitat Model</b>	<b>67</b>
7.1	Rationale . . . . .	68
7.2	Methods . . . . .	69
7.2.1	Study Design and Patient Cohorts . . . . .	69
7.2.2	CT-mpMRI Co-Registration . . . . .	70

7.2.3	CT Feature Extraction . . . . .	71
7.2.4	CT Habitat Model Development . . . . .	71
7.2.5	Histopathological Validation . . . . .	72
7.3	Results . . . . .	73
7.3.1	Handcrafted Radiomics Yields the Most Biologically Coherent CT Habitats . . . . .	73
7.3.2	CT Habitats Capture Distinct Vascular and Cellular Phenotypes . . . . .	75
7.3.3	Spatial Architecture and Histopathological Validation . . . . .	77
7.4	Discussion . . . . .	79
7.5	Summary . . . . .	80
<b>8</b>	<b>Clinical Relevance of CT Habitats</b>	<b>81</b>
8.1	Rationale . . . . .	81
8.2	Methods . . . . .	83
8.2.1	Patient Cohorts . . . . .	83
8.2.2	Habitat Computation . . . . .	83
8.2.3	Habitat-Derived Quantitative Metrics . . . . .	83
8.2.4	Statistical Analyses . . . . .	84
8.3	Results . . . . .	85
8.3.1	Patient characteristics . . . . .	85
8.3.2	Habitat Characteristics Across Cohorts . . . . .	85
8.3.3	TCIA: Neoadjuvant Chemotherapy Remodels Tumor Composition . . . . .	86
8.3.4	TCIA: Prognostic Value Depends on Treatment Context and Spatial Location . . . . .	88
8.3.5	VHIO: Prognostic Value Depends on Treatment and Molecular Context . . . . .	89
8.3.6	VHIO: Rim Entropy Change During Treatment . . . . .	93
8.4	Discussion . . . . .	95
8.5	Summary . . . . .	97
<b>V</b>	<b>Synthesis and Conclusions</b> . . . . .	<b>99</b>
<b>9</b>	<b>General Discussion</b>	<b>101</b>
9.1	Synthesis of Findings . . . . .	101
9.2	Biological Interpretation of CT-Derived Habitats . . . . .	102
9.3	Technical and Translational Considerations . . . . .	103
9.4	Final Reflections . . . . .	105
<b>10</b>	<b>Conclusions</b>	<b>107</b>
<b>References</b>		<b>109</b>
<b>A</b>	<b>Image Acquisition Details</b>	<b>123</b>
A.1	MRI Acquisition Details (PREDICT) . . . . .	123

## CONTENTS

A.2	mpMRI Maps Biological Ranges . . . . .	125
<b>B</b>	<b>Identification of Precise Handcrafted Features</b>	<b>127</b>
B.1	Cohort Characteristics . . . . .	127
B.2	Radiomics Features and Computation . . . . .	129
B.3	Image Perturbation . . . . .	131
B.4	Habitat Computation . . . . .	131
B.5	Intraclass Correlation Coefficient . . . . .	132
B.6	Justification for NGTDM Coarseness Inclusion . . . . .	135
B.7	Primary Tumor Has No Effect on Precision . . . . .	136
<b>C</b>	<b>Development and Validation of an mpMRI-anchored CT Habitat Model</b>	<b>139</b>
C.1	CT–mpMRI Co-Registration: Extended Methods . . . . .	139
C.2	Handcrafted Features - Correlation Analysis . . . . .	140
C.3	Sensitivity Analysis for Number of Habitats (K) . . . . .	143
C.4	Technical Validation: Extended Results . . . . .	143
C.5	mpMRI Characterization: Pairwise Comparison . . . . .	144
C.6	Habitat-Histology Correlations . . . . .	145
<b>D</b>	<b>Clinical Relevance of CT Habitats</b>	<b>147</b>
D.1	Habitat Metrics and Tumor Volume . . . . .	147
D.2	TCIA: Neoadjuvant Chemotherapy Remodels Tumor Composition . . . . .	148
D.3	Cox Regression Analyses by Treatment (VHIO) . . . . .	148
D.4	Longitudinal Analysis Summary Statistics . . . . .	149

# List of Figures

2.1	Hounsfield units and contrast enhancement in abdominal CT . . . . .	13
2.2	Examples of MRI and multiparametric MRI maps . . . . .	15
2.3	From radiological imaging to histopathological examination . . . . .	17
3.1	Histopathological heterogeneity in colorectal liver metastases . . . . .	21
3.2	Treatment landscape for colorectal liver metastases . . . . .	24
3.3	Response Evaluation Criteria in Solid Tumors (RECIST) . . . . .	25
4.1	Evolution of approaches to imaging heterogeneity . . . . .	28
4.2	Example of qualitative approach to imaging heterogeneity . . . . .	29
4.3	Schematic comparison of radiomics feature extraction methods . . . . .	32
4.4	Examples of MRI-based habitat imaging . . . . .	34
5.1	Overview of patient cohorts used in this thesis . . . . .	39
5.2	Habitat imaging pipeline overview . . . . .	45
5.3	Clustering scope: single-tumor, pooled, and hybrid modes . . . . .	46
6.1	Precision analysis workflow . . . . .	54
6.2	Repeatability of voxelwise radiomics features . . . . .	57
6.3	Reproducibility against kernel radius and bin size . . . . .	59
6.4	Summary of Precision Analysis . . . . .	60
6.5	Precise features increase habitat stability . . . . .	62
6.6	Exploration of the biological relevance of imaging habitats . . . . .	63
7.1	Schematic of the mpMRI-anchored CT habitat model development . . . . .	70
7.2	Selection of the optimal CT representation . . . . .	74
7.3	Biophysical characterization of CT habitats using mpMRI . . . . .	76
7.4	Spatial architecture and histopathological validation of CT habitats . . . . .	78
8.1	Clinical validation study design . . . . .	84
8.2	Spatial architecture and habitat-volume correlations across cohorts . . . . .	87
8.3	Neoadjuvant chemotherapy remodels tumor composition . . . . .	90
8.4	Treatment and molecular context modify prognostic value in VHIO . . . . .	92
8.5	Rim metrics differ between histopathological growth patterns . . . . .	94
8.6	Rim entropy dynamics during treatment . . . . .	95
9.1	Different representations capture different information . . . . .	105
B.1	Repeatability by primary tumor type . . . . .	136

## LIST OF FIGURES

B.2 Reproducibility by primary tumor type . . . . .	137
C.1 Exploratory correlations between CT habitat proportions and histological tissue percentages	141
C.2 Correlation matrix of the 26 precise handcrafted radiomics features . . . . .	142
C.3 Correlation matrix of the 6 selected precise handcrafted radiomics features . . . . .	146

# List of Tables

4.1	Summary of selected habitat imaging studies . . . . .	36
5.1	Summary of datasets and their role in this thesis . . . . .	40
5.2	Multiparametric MRI maps used for CT habitat development . . . . .	42
6.1	Reproducibility against kernel radius and bin size . . . . .	58
6.2	Precise 3D Radiomics Features in Liver and Lung Lesions . . . . .	61
7.1	Comparison of candidate CT feature representations . . . . .	74
7.2	CT Habitat Characterization with mpMRI metrics . . . . .	76
8.1	Clinical characteristics of the TCIA and VHIO cohorts . . . . .	86
8.2	Cox regression for overall survival in neoadjuvant-treated TCIA patients . . . . .	89
8.3	Context-dependent prognostic value of habitat metrics in VHIO . . . . .	93
A.1	mpMRI maps derived with biological ranges . . . . .	125
B.1	Total number of patients, images, and lesions per cohort and lesion location . . . . .	127
B.2	Image acquisition parameters per cohort . . . . .	128
B.3	List of primary tumor types included within the mixed cohort . . . . .	128
B.4	List of radiomics features analyzed in this study . . . . .	129
B.5	Image processing and radiomics feature computation parameters . . . . .	130
B.6	Two-way Mixed ANOVA Model . . . . .	134
C.1	Biophysical separation across different numbers of habitats . . . . .	143
C.2	Technical robustness of candidate CT feature representations . . . . .	144
C.3	Post-hoc pairwise comparisons for mpMRI metrics across CT habitats . . . . .	144
C.4	Spearman correlations between habitat proportions and histological tissue percentages .	145
D.1	Spearman correlations between habitat metrics and tumor volume . . . . .	147
D.2	Habitat metric comparisons: treatment-naive vs neoadjuvant-treated (TCIA) . . . . .	148
D.3	Cox regression: Chemotherapy alone . . . . .	148
D.4	Cox regression: Chemotherapy + Bevacizumab . . . . .	149
D.5	Cox regression: RAS-mutant patients . . . . .	149
D.6	Rim entropy change by RECIST category . . . . .	149





# Introduction

*The purpose of computing is insight, not numbers.*

— Richard Hamming



# Chapter 1

## Motivation, Objectives, and Contributions

### 1.1 Motivation

A patient with metastatic colorectal cancer undergoes a radiological examination after several cycles of therapy. The report indicates that her liver metastases have neither increased nor decreased in size. For the oncologist, this finding is ambiguous. The disease may be biologically controlled, with treatment-induced changes not yet translating into size reduction. Alternatively, the tumor may be adapting, with early biological resistance not yet visible as measurable growth. Size alone cannot distinguish between these possibilities. This uncertainty, encountered routinely in oncologic practice, motivates the work presented in this thesis.

Cancer is a leading cause of death worldwide [Bray et al., 2024], with most cancer-related mortality driven by metastatic disease rather than primary tumors [Chaffer and Weinberg, 2011]. The liver is one of the most common sites of metastasis across cancer types [Tsilimigras et al., 2021], and colorectal cancer is the leading source of liver metastases [Hess et al., 2006]. Approximately 25% of patients with colorectal cancer present with liver metastases at diagnosis, and up to 50% develop them during the course of disease. Surgical resection can be curative, but only a subset of patients are eligible. For the majority, systemic therapy is the main treatment option, with limited long-term survival [Engstrand et al., 2018].

Pathology studies have long recognized the biological heterogeneity of colorectal liver metastases and its clinical relevance. These tumors mostly contain varying proportions of viable tumor cells, fibrosis (i.e. scar tissue), and necrosis (i.e. dead cells) [Ozaki et al., 2022, Poulsides et al., 2012]. Importantly, both the presence of tissue components and their spatial location within the tumor have therapeutic implications. Fibrotic tissue at the tumor periphery following treatment has been associated with improved survival [Latacz et al., 2022], whereas extensive necrosis, particularly at the tumor core, has been linked to more aggressive disease and poorer prognosis [Rubbia-Brandt et al., 2007, Van den Eynden et al., 2013]. Despite their clinical relevance, such tissue phenotypes can only be assessed through surgical resection or biopsy. Resection provides comprehensive tissue but is limited to eligible patients at a single timepoint.

## MOTIVATION, OBJECTIVES, AND CONTRIBUTIONS

Biopsies are more widely performed, often to obtain molecular biomarkers for treatment selection, but they sample only a small fraction of the tumor. They cannot characterize spatial heterogeneity across the lesion or account for differences between metastases. Treatment decisions for most patients are therefore made without direct knowledge of tissue composition or its evolution over time.

CT is the standard imaging modality for the clinical management of colorectal liver metastases. CT scans are routinely acquired for staging and treatment monitoring, enabling repeated, non-invasive assessment of the whole liver over time. However, the information extracted from CT in clinical practice is largely size-based, focusing on lesion number and diameter. While size is simple to measure and interpret, it disregards changes in tissue composition within tumors (e.g. treatment-induced fibrosis or necrosis) that may precede or occur independently of volumetric shrinkage.

The field of radiomics treats medical images as quantitative data, computing radiomic features (i.e. numerical descriptors of intensity distributions and spatial texture patterns) to extract information beyond visual assessment [Lambin et al., 2012]. Most radiomics studies summarize each tumor using features averaged across the entire lesion, implicitly assuming that global aggregation is sufficient to characterize tumor biology despite known intratumor heterogeneity. Habitat imaging challenges this assumption [Gatenby et al., 2013]. Rather than collapsing tumors into summary statistics, habitat imaging partitions lesions into spatially distinct subregions based on imaging characteristics. These subregions, referred to as habitats, may correspond to different tissue phenotypes within the same tumor.

To date, habitat imaging has been applied predominantly to magnetic resonance imaging (MRI) [Li et al., 2024]. mpMRI combines multiple quantitative sequences acquired in the same imaging session, each sensitive to different tissue properties. For instance, the ADC reflects tissue cellularity [Le Bihan D, 1985], and the volume transfer constant,  $K^{\text{trans}}$ , provides information on vascular permeability and perfusion [Tofts and Kermode, 1991]. These biologically interpretable maps make MRI particularly suitable for studying intratumor heterogeneity in research settings. In contrast, CT-derived habitats are generally unexplored, despite CT being the dominant modality for managing colorectal liver metastases in routine clinical practice. This imbalance represents a translational limitation: methods developed for MRI are not readily applicable to clinical workflows where CT is standard.

Moreover, the habitat imaging literature lacks systematic evaluation of how data representation influences habitat computation. Tumor characteristics can be encoded using handcrafted (i.e. predefined) radiomics features [Van Griethuysen et al., 2017] or learned embeddings from deep learning models, in which image representations are optimized automatically rather than explicitly defined [Pai et al., 2025]. While learned representations are often assumed to be superior, their suitability for CT-derived habitats has not been established.

In addition to representation choices, important methodological and translational limitations remain. Feature robustness is often insufficiently assessed, biological validation is frequently qualitative or restricted to preclinical settings, and clinical relevance is rarely benchmarked against tumor volume, the current standard imaging biomarker. These limitations are particularly relevant for colorectal liver metastases, where systemic therapies may alter tissue composition before inducing measurable size change. Anti-angiogenic agents, for example, target tumor vasculature and can modify perfusion and tissue organization without producing immediate shrinkage. In such settings, biomarkers that capture tissue composition changes

could improve both treatment assessment and patient stratification.

These gaps motivate the central question of this thesis: can routine CT imaging capture biologically and clinically meaningful intratumor heterogeneity in colorectal liver metastases? Answering this requires assessing technical robustness of CT-derived features, evaluating which data representations produce biologically coherent habitats, characterizing what tissue phenotypes these habitats represent, and demonstrating clinical value beyond tumor volume. The following section formalizes these problems as five research questions.

## 1.2 Hypothesis and Objectives

This thesis aims to develop and validate a biologically grounded CT habitat imaging framework to characterize intratumor heterogeneity in colorectal liver metastases, and to assess its clinical relevance. This aim is based on the hypothesis that CT-derived habitats capture biologically meaningful heterogeneity that provides clinically relevant information beyond tumor volume. Towards this aim, we pose the following research questions:

1. Which CT handcrafted radiomics features are repeatable and reproducible enough to support stable habitat computation?
2. Which CT data representation produces habitats that best separate tissue with different cellularity and vascularity?
3. What tissue phenotypes do CT-derived habitats represent?
4. Do CT-derived habitats provide prognostic information independent of tumor volume?
5. If prognostic information exists, is it spatially localized (i.e. at the core or the rim), or is it uniformly distributed within a tumor?

These research questions correspond to five specific objectives:

1. **Objective 1:** To identify precise CT handcrafted radiomics features suitable for robust habitat computation. (*Chapter 6*)
2. **Objective 2:** To determine the optimal CT data representation for capturing biologically distinct tissue regions. (*Chapter 7*)
3. **Objective 3:** To characterize the biological meaning of CT-derived habitats using multiparametric MRI and histopathology. (*Chapter 7*)
4. **Objective 4:** To evaluate whether CT-derived habitats provide clinically relevant information beyond tumor volume. (*Chapter 8*)
5. **Objective 5:** To investigate the spatial localization of clinically relevant heterogeneity within tumors. (*Chapter 8*)

## 1.3 Contributions to Knowledge

This thesis contributes to the field of imaging-based tumor heterogeneity analysis by addressing key methodological, biological, and clinical limitations of existing habitat imaging approaches.

### Main Contributions

**1. An open-source, modality-agnostic habitat imaging pipeline was developed for the first time.**

The pipeline accepts any voxelwise feature maps as input and outputs cluster assignments and habitat-derived metrics. It is described in Chapter 5 and available at: <https://github.com/radiomicsgroup/imaging-habitats-pipeline>.

**2. The precision of voxelwise CT radiomics features for habitat imaging was evaluated.** This is the first comprehensive assessment of repeatability and reproducibility for 3D voxelwise features in liver and lung lesions. This work is described in Chapter 6 and was published as:

- **Prior O**, Macarro C, Navarro V, et al; Bernatowicz K, Perez-Lopez R. Identification of Precise 3D CT Radiomics for Habitat Computation by Machine Learning in Cancer. *Radiol Artif Intell.* 2024;6(2):e230118. DOI: [10.1148/ryai.230118](https://doi.org/10.1148/ryai.230118).

**3. Multiple CT data representations were compared for the first time for habitat imaging.** Raw Hounsfield units, handcrafted radiomics, and deep learning embeddings were evaluated using co-registered multiparametric MRI as biological reference. This work is described in Chapter 7.

**4. A biologically anchored CT habitat framework was developed.** Habitat definitions were constrained by tissue properties (cellularity, vascularity) measured with quantitative MRI rather than post-hoc outcome associations. This work is described in Chapter 7. Preliminary results were presented at international meetings as:

- **Prior O**; Grussu F; Garcia-Ruiz A; et al; Perez-Lopez R. Decoding liver intra-tumour heterogeneity with co-localized CT and multi-parametric MRI. *ISMRM Diffusion MRI Workshop* (Amsterdam, The Netherlands, 2022). Oral presentation.

- **Prior O**; Grussu F; Macarro C; et al; Perez-Lopez R. Dissecting heterogeneity in liver metastases: an mpMRI and CT approach. *ISMRM Iberian Chapter* (Valladolid, Spain, 2023). Poster.

- **Prior O**; Macarro C; Grigoriou A; et al; Perez-Lopez R. Non-invasive Characterization of Intratumor Heterogeneity: Comparing MRI-Radiomics and Histological Habitats. *ISMRM Iberian Chapter* (Porto, Portugal, 2024). Oral presentation.

**5. The clinical relevance of CT habitats was demonstrated to depend on context.** Habitat metrics were prognostic in anti-angiogenic and post-neoadjuvant settings but not with chemotherapy alone, and added information beyond tumor volume. This work is described in Chapter 8.

**6. Prognostic information was shown to concentrate at the invasive rim.** Rim-based metrics outperformed whole-tumor heterogeneity measures. This work is described in Chapter 8.

- *Manuscript in preparation (Contributions 3–6): Prior O, et al; Grussu F, Perez-Lopez R. Mapping Tumor Heterogeneity of Colorectal Liver Metastases via CT Habitats Improves Prognosis. In preparation.*

## Other Contributions During the PhD

### Publications

- Ligero M, Gielen B, Navarro V, **Prior O**, et al; Perez-Lopez R. A whirl of radiomics-based biomarkers in cancer immunotherapy, why is large scale validation still lacking? *npj Precis. Onc.* 2024;8(1):42. DOI: [10.1038/s41698-024-00534-9](https://doi.org/10.1038/s41698-024-00534-9).
- Bernatowicz K, Amat R, **Prior O**, et al; Perez-Lopez R. Radiomics signature for dynamic monitoring of tumor inflamed microenvironment and immunotherapy response prediction. *J Immunother Cancer.* 2025;13(1):e009140. DOI: [10.1136/jitc-2024-009140](https://doi.org/10.1136/jitc-2024-009140).
- de Grandis MC, Baraibar I, **Prior O**, et al; Perez-Lopez R. Differentiating low tumor burden from oligometastatic disease in colorectal cancer: a call for individualized therapeutic approaches. *ESMO Open.* 2025;10(8):105520. DOI: [10.1016/j.esmoop.2025.105520](https://doi.org/10.1016/j.esmoop.2025.105520).
- Voronova AK, **Prior O**, Grigoriou A, et al; Perez-Lopez R. Simulation-informed evaluation of microvascular parameter mapping for diffusion MR imaging of solid tumours. *medRxiv.* 2025. DOI: [10.1101/2025.08.27.25334553](https://doi.org/10.1101/2025.08.27.25334553).

### Conference Presentations

- **Prior O**; Bernatowicz K; Ligero M; et al; Perez-Lopez R. Artificial Intelligence for Predicting Response to Standard of Care Therapy in MSS RASmt mCRC Patients. *Cancer Core Europe Summer School* (Albufeira, Portugal, 2022). Oral presentation.
- **Prior O**; Gielen B; Ligero M; et al; Perez-Lopez R. Translating Imaging into Insight: Can Radiology and Machine Learning predict Immunotherapy Response in a Multi-Tumor Landscape? *ASEICA 40th Anniversary Congress* (A Coruña, Spain, 2023). Oral presentation.

### International Research Projects

- **COLOSSUS (EU-funded H2020 project, led by the Royal College of Surgeons in Ireland):** Contribution to the development of candidate imaging biomarkers for colorectal cancer patients.
  - *Manuscript in preparation:* Connor K\*, **Prior O.\***, Shiels LP, et al; Byrne AT. “Co-clinical CT radiomics pipeline to establish candidate imaging biomarkers for colorectal cancer.” (\*equal contribution)
- **POEM Study (led by the Oncology Center Antwerp, Belgium):** Participation in the prospective characterization of histological growth patterns of liver metastases.

## MOTIVATION, OBJECTIVES, AND CONTRIBUTIONS

- Latacz E, Prior Palomares O., Ruiz Roig N, et al. "Prospective complete histopathological characterization of liver metastases from colorectal and breast carcinoma to predict the histopathological growth patterns by medical imaging (POEM)." *Eur J Surg Oncol.* 2024. DOI: [10.1016/j.ejso.2024.109241](https://doi.org/10.1016/j.ejso.2024.109241).

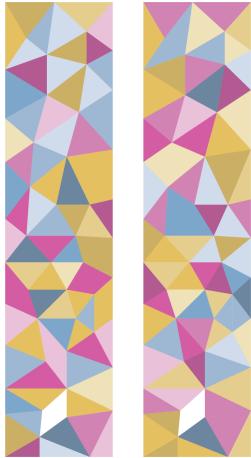
### Internal Pipelines - Radiomics Group at VHIo

- **CT preprocessing pipeline:** Developed and implemented standardized CT preprocessing pipelines for the Radiomics Group at VHIo, ensuring reproducibility in feature extraction across imaging datasets from multiple clinical studies.
- **Machine learning framework:** Developed a reusable machine learning framework for the Radiomics Group with emphasis on nested cross-validation to ensure model generalizability and prevent data leakage.

## 1.4 Thesis Structure

This thesis consists of five parts across ten chapters. Part I introduces the problem and research objectives (**Chapter 1**). Part II provides foundations: medical imaging techniques (**Chapter 2**), oncological context with focus on colorectal liver metastases and treatment response assessment (**Chapter 3**), and imaging heterogeneity methods including the state of the art in habitat imaging (**Chapter 4**). Part III describes data sources, the general methodological framework, and the habitat imaging pipeline developed for this work (**Chapter 5**). Part IV presents the core experimental contributions: precision analysis of radiomics features (**Chapter 6**), development and biological validation of CT habitats using mpMRI (**Chapter 7**), and clinical relevance assessment across treatment contexts (**Chapter 8**). Part V synthesizes findings and draws conclusions: general discussion including limitations and future directions (**Chapter 9**), and main conclusions (**Chapter 10**).

Each experimental chapter is self-contained and can be read independently. Sequentially, they follow a logical progression: from identifying reliable imaging features, to constructing biologically-informed habitats, to demonstrating their clinical utility. Extended methods and supplementary results are provided in the Annexes.



# Foundations

*But suddenly the tumor tells its owner: end of mystery, the veil is lifted.  
And like in a game of Cluedo that ends, one of the players reveals all  
the details of the crime and interrupts the turns with a killer  
declaration: “I accuse cancer, with its metastases, in the hospital  
room”.*

— Delphine Horvilleur, *Vivre avec nos morts*



# Chapter 2

## Medical Image Techniques: An Overview

Imaging guides clinical decisions at nearly every stage of cancer care. The modalities most relevant to this thesis are computed tomography (CT), magnetic resonance imaging (MRI), and histopathology. This chapter describes how each technique generates images and captures tumors at different scales. A note on terminology: because radiological images are 3D volumes, the basic unit of measurement is the voxel (a pixel extended into three dimensions). A single CT or MRI voxel typically represents a tissue volume on the order of  $1\text{mm}^3$ . Histopathological images, by contrast, are 2D sections at micrometer resolution so the basic unit is the pixel.

### 2.1 Computed Tomography

CT produces cross-sectional images of the body using X-rays [Sprawls, 1995]. The physics begins with how X-ray photons interact with matter. When photons pass through tissue, they can be absorbed (the photoelectric effect) or deflected (Compton scattering). Both processes remove photons from the X-ray beam reducing its intensity. This reduction is known as attenuation and we quantify it with the linear attenuation coefficient,  $\mu$ . In our bodies, different tissues attenuate X-rays to different degrees depending on their composition. For instance, bone, with its high calcium content, attenuates strongly (i.e. has a high  $\mu$ ); air attenuates almost none; soft tissues like liver fall in between.

A CT scanner consists of an X-ray tube and a rotating detector, with the patient lying on a table that passes through. The X-ray tube emits a narrow beam that traverses the patient while detectors on the opposite side count the photons that emerge. Each detector reading represents the sum of attenuation along that beam's path—mathematically, a line integral of  $\mu$ . Fewer emerging photons means more attenuation occurred along that path. Over a full rotation, the scanner collects hundreds of such measurements from different angles, effectively sampling line integrals of  $\mu(x)$  through the body from all directions. The relationship follows Lambert-Beer's law: a beam from origin  $s$ , with unit direction vector  $\theta$  and initial intensity  $N_0$  is reduced to  $N$  as it passes through tissue, where:

$$N = N_0 e^{-\int d\lambda \mu(s + \lambda\theta)}$$

Turning these raw measurements into an image is an inverse problem: recovering the spatial distribution of attenuation coefficients from their line integrals. The standard solution is filtered back-projection, which combines a sharpening filter with projection along original beam paths to reconstruct each slice. The result is a 3D map of attenuation coefficients, reconstructed as a stack of slices. Each slice represents a cross-section of the body at a given thickness, typically 0.5–5 mm depending on the clinical application. Each element in this 3D map (voxel) represents the average attenuation within a small tissue volume. All voxel values are rescaled to Hounsfield units (HU)<sup>1</sup>, a dimensionless scale defined so that water is 0 HU and air is –1000 HU:

$$\text{HU} = 1000 \times \frac{\mu_{\text{tissue}} - \mu_{\text{water}}}{\mu_{\text{water}}}$$

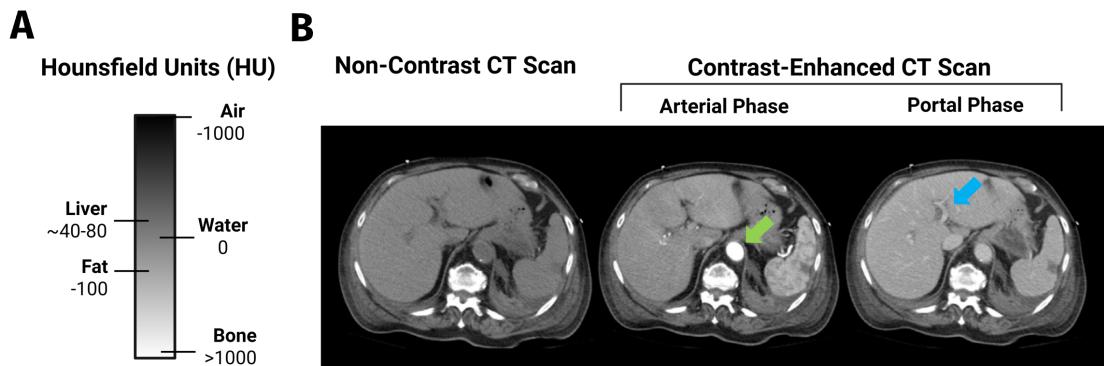
On this scale (Figure 2.1), soft tissues typically fall around 20–80 HU, bone exceeds +400 HU, and fat is slightly negative (around –100 HU). Radiologists can adjust the range of HU values (known as window) to optimize visualization for different tissues. Importantly, HU values depend on scanner model, acquisition settings (tube voltage, reconstruction kernel), and whether contrast has been administered. This variability is why only broad HU ranges exist for tissues rather than a precise atlas. When contrast is used we refer to the image as a CECT scan. Contrast agents are usually iodine-based. This is because iodine has a high atomic number, which makes it a strong X-ray absorber. Thus, tissues containing the contrast show increased attenuation. Since contrast is administered intravenously, the resulting enhancement reflects blood flow. A CECT scan can be obtained in different phases related to where the contrast is travelling through the body at that point:

1. Arterial phase (~20–30 seconds post-injection): contrast fills the arterial system, highlighting arteries and hypervascular lesions that receive blood directly from arterial supply.
2. Portal venous phase (~60–70 seconds): contrast fills the portal venous system and liver tissue. This phase provides the best tumor-to-liver contrast for detection of liver metastases.
3. Delayed phase (several minutes): contrast has diffused into extracellular spaces and is washing out.

CT is one of the most widely used imaging modalities due to its speed (the total acquisition time is less than 15 minutes), availability, and quantitative nature. The main disadvantage is ionizing radiation. For primary liver tumors such as hepatocellular carcinoma, MRI is preferred because its superior soft tissue contrast enables more accurate diagnosis [American Cancer Society, 2025]. For colorectal liver metastases, however, CT remains the standard for diagnosis and response assessment.

---

<sup>1</sup>The method was developed in the early 1970s by Sir Godfrey Hounsfield, building on mathematical work by Allan Cormack on image reconstruction, for which they shared the 1979 Nobel Prize in Physiology or Medicine.



**Figure 2.1: Hounsfield units and contrast enhancement in abdominal CT.** (A) Hounsfield unit (HU) scale with typical values for air ( $-1000$  HU), fat ( $\approx -100$  HU), water (0 HU), soft tissue such as liver ( $\approx 40$ – $80$  HU), and bone ( $>1000$  HU). HU provide a quantitative measure of X-ray attenuation, with values depending on scanner and protocol. (B) Axial abdominal CT images acquired without contrast and after contrast administration in the arterial and portal venous phases, adapted from [Hartung et al., 2024]. Images are displayed according to radiological convention (viewed from the patient's feet). Arrows indicate the liver's vasculature (blue) and the aorta (green) to highlight phase-specific enhancement. Created with BioRender.com.

## 2.2 Magnetic Resonance Imaging

Magnetic resonance imaging<sup>2</sup> (MRI) generates images by exploiting the magnetic properties of atomic nuclei, primarily hydrogen. Unlike CT, MRI uses no ionizing radiation. Instead, it's based on radiofrequency pulses and strong magnetic fields to induce and detect signals from water protons in tissue. Figure 2.2 shows examples of MRI images of patients with liver metastases.

### 2.2.1 Physical Principles

A hydrogen nucleus consists of a single proton, which possesses a property called spin, an intrinsic angular momentum that gives it a small magnetic moment. Normally, these magnetic moments are randomly oriented. When placed in a strong external magnetic field (typically 1.5 or 3 Tesla in clinical scanners), they align preferentially with the field direction, creating a net magnetization [Lauterbur, 1973]. A radiofrequency pulse at the resonant frequency (known as the Larmor frequency) tips this magnetization away from equilibrium. When the pulse ends, the system relaxes back to equilibrium, releasing energy as a radiofrequency signal that can be detected by receiver coils. This signal is used to construct the MR image.

Two relaxation processes govern signal behavior [Pooley, 2005]. T1 relaxation (spin-lattice relaxation) describes the recovery of longitudinal magnetization as protons release energy to their surroundings. T2 relaxation (spin-spin relaxation) describes the decay of transverse magnetization as protons lose phase coherence due to interactions with each other. Different tissues have different T1 and T2 values depend-

<sup>2</sup>MRI was developed for clinical use around 1980, building on foundational work by Lauterbur (who described the first MR image in 1973) and Mansfield (who developed gradient-based spatial encoding). Both were awarded the Nobel Prize in Physiology or Medicine in 2003.

ing on their molecular environment, which is what creates contrast in MR images. By manipulating the timing of pulses and signal acquisition, different contrasts can be generated. T1-weighted images emphasize differences in T1 (fat appears bright, fluid dark) while T2-weighted images emphasize differences in T2 (fluid appears bright). This flexibility allows MRI to highlight different tissue properties depending on clinical need. Spatial localization is achieved through magnetic field gradients, which are small, controlled variations in field strength across the imaging volume. These gradients encode spatial position into the frequency and phase of the MR signal, enabling reconstruction of a spatially resolved image through Fourier transformation.

## 2.2.2 Quantitative MRI

MRI can provide quantitative measurements of tissue properties through specialized sequences and biophysical modeling. In this thesis, two types of quantitative MRI are relevant: diffusion-weighted imaging and dynamic contrast-enhanced imaging.

### Diffusion-Weighted Imaging

Diffusion-weighted MRI (DWI) measures the random motion of water molecules in tissue [Le Bihan D, 1985]. A magnetic field gradient is applied that labels water protons based on their position. After a short interval, an opposite gradient is applied. If protons were stationary, the second gradient would perfectly reverse the effect of the first and full signal would be recovered. But water molecules undergo Brownian motion, so protons move between the two gradients. Those that have diffused further experience incomplete refocusing, leading to signal loss. The more freely water can diffuse, the greater the signal attenuation.

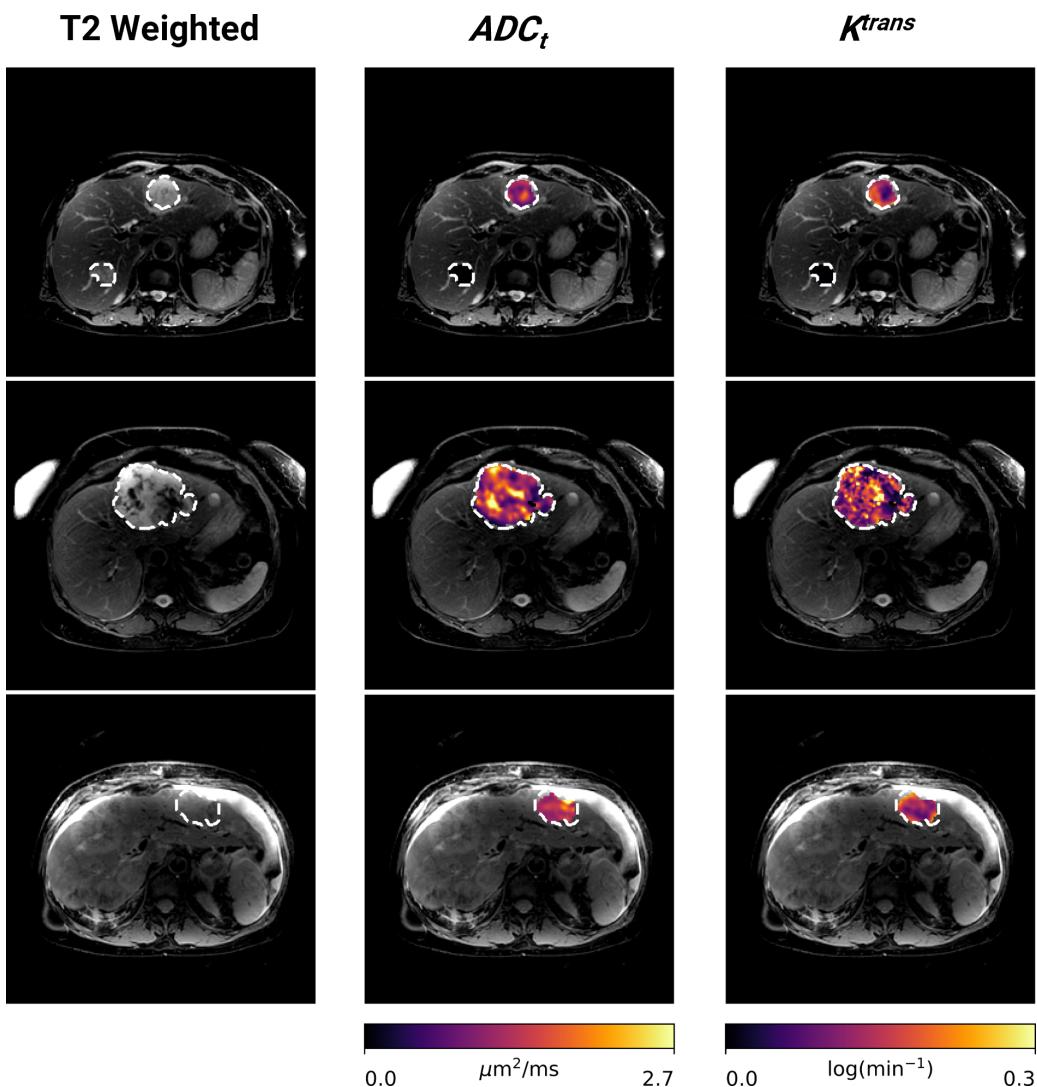
The strength and timing of the diffusion gradients are combined into a single parameter called the b-value (with units s/mm<sup>2</sup>). At higher b-values, the sequence becomes more sensitive to diffusion and signal attenuation increases. The ADC quantifies water mobility by fitting the signal decay across b-values. For two b-values,  $b_1$  and  $b_2$ , and corresponding signal intensities  $S(b_1)$  and  $S(b_2)$ , the ADC is:

$$\text{ADC} = \frac{1}{b_2 - b_1} \ln \left( \frac{S(b_1)}{S(b_2)} \right)$$

Low ADC indicates restricted diffusion (water has diffused less so there is less signal attenuation). Biologically, this can mean high cellularity since cells obstruct water movement. Following the same reasoning, high ADC indicates freer diffusion (water has diffused more and therefore there is more signal attenuation), which means that the tissue in question has low cellularity (e.g. necrosis or edema). While ADC is used as a proxy for cellularity, we must acknowledge that it is a measurement conflating multiple factors including cell density, cell size, membrane permeability, and vascular perfusion. These factors affect ADC in competing ways, which explains why correlations between ADC and histological cellularity have been inconsistent across studies [Chen et al., 2013, Surov et al., 2017, Yoshikawa et al., 2008].

More advanced diffusion models attempt to disentangle these factors by assuming a certain tissue mi-

crostructure. In these models, the signal is represented as a mixture of compartments with distinct water diffusivity, such as vascular, intracellular, and extracellular components [Novikov et al., 2018]. Many such models have been proposed, and while they remain mostly investigational, they have shown promising results for clinical applications. A well known example of a model is the intravoxel incoherent motion (IVIM) model [Le Bihan et al., 1988], which assumes that tissue water is present in two different compartments: vascular (pseudo-diffusing water inside blood vessels) and nonvascular (diffusing water in and around cells). In this thesis, when diffusion–relaxation modeling is employed, we therefore distinguish between a tissue-related apparent diffusion coefficient (ADC<sub>t</sub>) and a vascular-related apparent diffusion coefficient (ADC<sub>v</sub>). The specific models and fitting strategies used are described in Chapter 5.



**Figure 2.2: Examples of MRI and multiparametric MRI maps.** Axial T2-weighted MRI slices from three representative patients with liver metastases. The example mpMRI maps overlays are the apparent diffusion coefficient ( $ADC_t$ ) and the volume transfer constant ( $K_{trans}$ ), derived from DWI and DCE, respectively. In  $ADC_t$ , darker regions indicate restricted diffusion (higher cellularity). In the  $K_{trans}$  maps, brighter regions indicate higher vascular permeability.

## Dynamic Contrast-Enhanced MRI

DCE captures the flow of a contrast agent through tissue over time. Unlike CT, the contrast agent is gadolinium-based. Following the injection of the contrast, MRI differentiates tissues based on how gadolinium shortens the T1 relaxation. Therefore, in tissues where gadolinium accumulates, the signal is enhanced. Pharmacokinetic models analyze these enhancement curves. The most common model used is the Tofts model [Tofts and Kermode, 1991] which assumes that the movement of the contrast agent in the tissue is from the blood vessels to the extracellular space (outside the cells). Thus, we can obtain a parameter related to the contrast agent leakage (perfusion),  $K^{trans}$ , and a parameter related to the volume of the extracellular space,  $ve$ .

Within a single examination, multiple MRI sequences can be acquired to obtain functional information beyond anatomy. We refer to this combination of sequences as mpMRI. After preprocessing and model fitting, we can obtain the different quantitative parameters such as ADC,  $f_v$ , and  $K^{trans}$ . We refer to these as mpMRI maps. In this thesis, mpMRI serves as the biological reference for developing and validating CT-derived habitats (Chapter 7). The specific parameters used are detailed in Chapter 5.

## 2.3 Histopathologic Light Microscopy

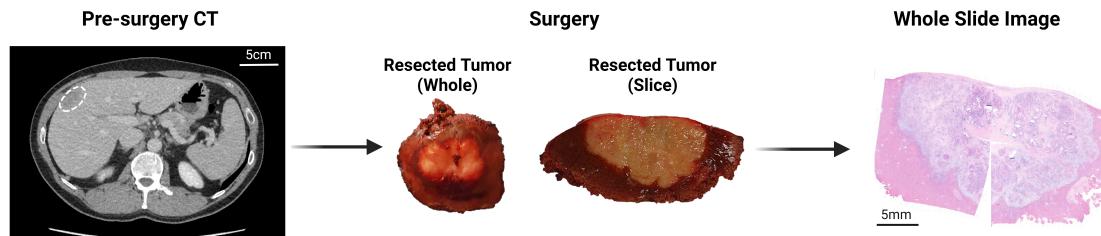
Unlike the non-invasive imaging techniques described above, histopathological assessment requires the surgical removal of tissue. The extracted sample must then be processed through a series of preparation steps [Paxton et al., 2003] prior to microscopic examination.

First, to prevent tissue decay, the sample requires chemical fixation, which is done by applying formalin, which inactivates enzymes and kills bacteria that would otherwise degrade the tissue. Next, to enable processing into thin slices, the tissue's mechanical stability must be increased. This is achieved by first dehydrating the sample with alcohol and then embedding it in paraffin wax. After this the tissue is a hardened block which is cut by a microtome (i.e. a slicer) that produces sections typically around 5  $\mu\text{m}$  thick. Once the tissue sections are mounted on glass microscope slides, dyes are applied to enhance visual contrast of cellular components. The most common staining is Haematoxylin and eosin (H&E), in which cell nuclei appear purple and most other components different shades of pink (Figure 2.3).

The stained and mounted sections are then examined through conventional light microscopy, which achieves resolutions down to approximately 200 nm. In the last few years pathologists have shifted from using conventional microscopy to digitized images. This is known as WSI [Pantanowitz et al., 2015]. Here, automated scanners that combine a light microscope and a digital camera digitize whole tissue sections. WSI images are usually obtained at several magnifications (stored as an image pyramid), with 40 $\times$  magnification producing pixels that are roughly 0.25  $\mu\text{m}$  in size. The resulting files grow to several gigabytes per slide, which presents storage issues but also makes computational analysis possible. Algorithms that quantify details that are impossible for the human eye to count can be fed the same image that a pathologist examines.

Understanding the scale gap between radiology and pathology will allow us to contextualize the results

in Chapter 7. A single CT voxel may contain tens of thousands of hepatocytes (liver cells) while in a whole-slide image a single hepatocyte covers thousands of pixels. The resolution gap is approximately nine orders of magnitude in volume.<sup>3</sup> As an analogy, we can picture Barcelona's beach. Histopathology is like examining individual grains of sand while CT imaging is like viewing the beach from an airplane. From altitude we can tell fine sand from rocky areas, but individual grains are invisible.



**Figure 2.3: From radiological imaging to histopathological examination.** Pre-operative CT reveals a colorectal liver metastasis (dashed contour, ~4 cm). Following surgical resection, the specimen is sectioned and processed for whole-slide imaging (H&E), enabling examination at cellular resolution.

---

<sup>3</sup>A CT voxel is approximately 1 mm<sup>3</sup> volume ( $=10^9 \mu\text{m}^3$ ). Assuming an average hepatocyte diameter of  $\sim 30 \mu\text{m}$  (volume  $\approx 14,000 \mu\text{m}^3$  for a sphere), this is roughly  $10^4$ – $10^5$  hepatocytes per voxel. Whole-slide imaging at 40 $\times$  magnification has a pixel size of  $\sim 0.25 \times 0.25 \mu\text{m}$  ( $0.0625 \mu\text{m}^2$ ). Approximating a hepatocyte cross-section as a circle of diameter  $30 \mu\text{m}$  (area  $\approx 700 \mu\text{m}^2$ ), one cell spans  $\sim 10^4$  pixels. Comparing volumes: one CT voxel ( $10^9 \mu\text{m}^3$ ) versus one histology pixel extended to a 5  $\mu\text{m}$  section thickness ( $0.0625 \mu\text{m}^2 \times 5 \mu\text{m} \approx 0.3 \mu\text{m}^3$ ) gives a ratio of  $\sim 10^9$ .



# Chapter 3

## Colorectal Liver Metastases

This chapter provides the clinical and biological context for understanding the importance of imaging heterogeneity in colorectal liver metastases.

### 3.1 Background

Liver metastases are tumors that originate from a primary cancer elsewhere in the body and establish growth in hepatic tissue [Tsilimigras et al., 2021]. The liver is one of the most common metastatic sites regardless of primary tumor type [Budczies et al., 2015]. There are two main theories explaining this organotropism (i.e. preference of cancer for specific organs). On one hand, Paget [Paget, 1889] proposed that metastasis requires compatibility between tumor cells (the "seed") and the host organ (the "soil"). The liver is highly vascularized, rich in growth factors, and immunologically more tolerant than other organs [Valderrama-Treviño et al., 2017]. These features make it a hospitable environment for circulating cancer cells. On the other hand, Ewing [Ewing, 1919] emphasized the role of blood flow. The liver receives a dual blood supply: oxygenated blood from the hepatic artery and blood from the portal vein, which drains the intestines and other abdominal organs. For tumor cells that enter the circulation from an abdominal primary, the liver is often the first organ they encounter.

Among solid tumors, colorectal cancer is the most common source of liver metastases, followed by pancreatic cancer, breast cancer, lung cancer, and melanoma [Hess et al., 2006]. Colorectal cancer originates in the epithelial lining (i.e. tissue walls) of the colon or rectum. It is the third most common malignancy worldwide, with approximately 1.9 million new cases and over 900,000 deaths annually [Morgan et al., 2023]. Around 25% of patients present with liver metastases at diagnosis, and half will develop them during their disease course [Maher et al., 2017]. This results in a significantly reduced life expectancy with a 5-year overall survival of 17% compared to patients without liver metastases (5-years survival rate of 70%) [Engstrand et al., 2018]. Tumors that originate in the left side of the colon are more likely to metastasize to the liver, whereas right-sided tumors more commonly spread to the peritoneum (i.e. membrane that covers abdominal organs) [Siegel et al., 2020].

## 3.2 Histopathological Heterogeneity

Colorectal liver metastases are not uniform masses. Different tissue types coexist within the same lesion, and their proportions vary from tumor to tumor (Figure 3.1A). The three main tissue types in colorectal liver metastases are viable tumor, fibrosis, and necrosis [Bird et al., 2006]. Viable tumor consists of densely packed cancer cells, usually forming glandular patterns inherited from the primary colorectal tumor. These regions have high blood supply to match their metabolic demands. Fibrosis is the deposition of collagen and extracellular matrix by stromal cells. Necrosis represents cell death, typically resulting from inadequate blood supply or treatment effect. Necrotic regions lack cellular architecture and contain dying cells or cellular debris.

In the histopathology literature of colorectal liver metastases, we can find two types of necrosis: usual necrosis and ILN [Chang et al., 2012]. Usual necrosis is characterized by cell debris mixed with some viable cells while ILN presents as large confluent areas with absent or minimal cell debris. It is hypothesized that usual necrosis is the result of the tumor outgrowing its blood supply while ILN is a consequence of treatment and may represent an intermediate stage in the evolution from viable tumor to fibrosis [Loupakis et al., 2013]. These two definitions have been relevant in research settings but are not routinely distinguished in clinical practice.

These tissue types do not distribute randomly within tumors, as can be observed in Figure 3.1. A common spatial pattern is concentric organization: necrosis or fibrosis predominates at the center while viable tumor glands are located at the periphery. The presence of viable cells at the periphery may be explained by the periphery's proximity to healthy liver vasculature. A scattered distribution of tissue types throughout the metastasis, or a pattern with viable cells at the center, is less commonly observed.

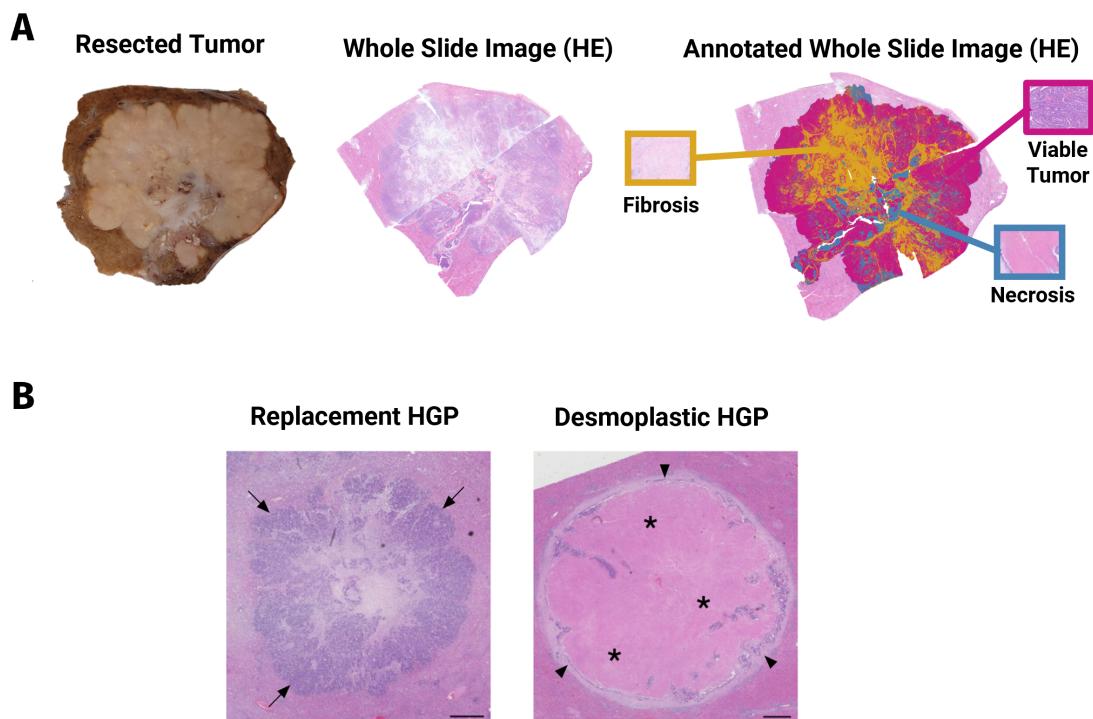
Importantly, necrosis, fibrosis and viable tumor carry prognostic significance, though interpretation depends on context. Baseline necrosis (i.e. present before treatment, usual necrosis) suggests aggressive biology and predicts worse outcomes [Van den Eynden et al., 2013], perhaps due to poor drug penetration into the tumor core [Wong and Neville, 2007]. Fibrosis, in contrast, is linked to improved outcomes after treatment, following the hypothesis that when tumor cells are killed, the liver replaces them with scar tissue. As a matter of fact, fibrosis is the only tissue type that is formally measured. This only happens when patients have been selected for surgery and have received presurgical (i.e. neoadjuvant) treatment. After surgical resection, pathologists evaluate resected tumors using tumor regression grading (TRG), a histological scoring system that quantifies the extent of residual viable tumor versus fibrosis [Rubbia-Brandt et al., 2007].

In this system, TRG 1 indicates complete regression with no residual tumor and dense fibrosis, while TRG 5 indicates no response, with abundant viable tumor and no fibrosis. This scoring system illustrates the prognostic importance of fibrosis. TRG, however, can only be assessed in patients who undergo surgery, which represents a minority of patients with colorectal liver metastases. For the majority, tissue composition remains unknown throughout treatment.

Another tissue type that may be present in colorectal liver metastases is mucin (i.e. acellular collections of mucus). This occurs in mucinous adenocarcinomas, which represent 10–15% of colorectal cancers

[Hugen et al., 2016]. Colorectal liver metastases with mucinous histology are not studied in this thesis.

Beyond internal composition, the interface between tumor and liver carries prognostic significance. This interaction has been characterized through histopathological growth patterns (HGP), described not only for colorectal liver metastases but for liver metastases from other primaries as well [Latacz et al., 2022] (Figure 3.1B). Liver metastases present three main HGP: desmoplastic, replacement and less common pushing pattern. Desmoplastic, also known as encapsulated, growth is characterized by a fibrous rim separating tumor from liver parenchyma. It is not clear whether this rim is a reaction from the liver or a consequence of the tumor composition determining this specific growth. Tumors with desmoplastic HGP associate with improved outcomes. Replacement growth occurs when tumor cells infiltrate along existing liver architecture, co-opting host vasculature rather than inducing new vessel formation. This pattern is associated with worse prognosis and has been hypothesized to be the default pattern of successful aggressive tumor invasion [Fernández Moro et al., 2023]. Pushing growth, which is less common, describes tumors that compress but do not infiltrate surrounding liver. The tumor-liver boundary is well-defined but lacks a fibrous capsule, and prognosis is intermediate.



**Figure 3.1: Histopathological heterogeneity in colorectal liver metastases.** (A) Representative histology sections of a colorectal liver metastasis showing the three dominant tissue compartments: viable tumor (cellular, vascularized), necrosis (central, avascular), and fibrosis (often treatment-induced). (B) The two main histopathological growth patterns at the tumor-liver interface: desmoplastic (fibrous rim indicated by arrowheads) and replacement (infiltration along liver architecture indicated by arrows). Asterisks indicate areas of infarct-like necrosis. Adapted from Frentzas et al., 2016.

The distinction between these patterns has therapeutic implications. Solid tumors require blood supply to grow beyond a few millimeters. They can acquire this supply through two mechanisms: angiogenesis or vascular co-option. Angiogenesis is the formation of new blood vessels, typically driven by hypoxia. When tumors outgrow their blood supply, hypoxic regions upregulate vascular endothelial growth factor (VEGF), which promotes new vessel formation. Vascular co-option occurs when tumor cells ‘hijack’ ex-

isting vasculature without inducing new vessels. Replacement-pattern tumors rely on vascular co-option: they grow along the liver's vascular network rather than building their own vessels. Anti-angiogenic therapy (e.g. bevacizumab) targets VEGF-driven new vessel formation, so tumors that co-opt existing vessels may be inherently resistant [Frentzas et al., 2016].

### 3.3 Treatment Landscape and Selection Challenges

The prognosis for patients with metastatic colorectal cancer (mCRC) is poor, but outcomes vary substantially depending on whether the tumors can be resected or not. This distinction, primarily determined by CT imaging, separates patients into two fundamentally different treatment paths: surgery or systemic therapy.

#### 3.3.1 Resectable disease

Hepatic resection is considered to be the only potentially curative treatment. The goal is complete resection with negative margins (i.e. resection margins free of tumor cells) while preserving enough functional liver volume. Eligibility for hepatic resection depends on patient fitness, absence of other distant metastases, tumor number, size, and location. CT imaging plays a central role in this assessment. Five-year survival rates after resection are 58% compared with 27% for patients without resection [Zeineddine et al., 2023].

For patients whose disease is initially unresectable due to tumor burden or location, neoadjuvant chemotherapy can sometimes reduce tumor size sufficiently to enable surgery. This conversion strategy has expanded the population eligible for potentially curative resection. The response to neoadjuvant treatment also provides prognostic information: patients whose tumors respond favorably have better long-term outcomes even after complete resection. As discussed in Section 3.2, neoadjuvant therapy does not merely shrink tumors but remodels their internal composition, and pathological response (assessed via TRG) correlates with survival.

#### 3.3.2 Unresectable disease

The majority of patients with colorectal liver metastases are not candidates for surgery. For these patients, systemic therapy is given with the goal of controlling the disease, prolonging survival, and maintaining quality of life. There are three main types of treatment: chemotherapy, targeted therapies and immunotherapy. Treatment selection is guided by molecular biomarkers and primary tumor sidedness (Figure 3.2).

Standard chemotherapy regimens combine fluoropyrimidines (5-fluorouracil or capecitabine) with either oxaliplatin (FOLFOX) or irinotecan (FOLFIRI). These agents target rapidly dividing cells by disrupting DNA synthesis and repair. When effective, they kill tumor cells, which is why we refer to chemotherapy as being a cytotoxic (i.e. causing damage or death to cells) treatment. Unlike chemotherapy, targeted

therapies and immunotherapy are cytostatic (i.e. inhibiting cell growth and proliferation, which may lead to cell death eventually). In immunotherapy, the patient's immune system is enhanced to recognize and control cancer cells by blocking inhibitory signals, thereby slowing tumor growth without immediately killing cells outright. Immunotherapy has transformed outcomes for mCRC patients, with median overall survival being more than twice as long as in patients treated with chemotherapy [André et al., 2025]. Only 5% of mCRC patients are eligible for approved immunotherapy [Bari et al., 2025]. These are patients with MSI status, which indicates defective DNA mismatch repair system.

Targeted therapies are treatments that target specific proteins involved in tumor growth and survival. In metastatic colorectal cancer, the main targets are the EGFR and VEGF. EGFR is a receptor on the cell surface that, when activated, triggers signaling pathways promoting cell proliferation and survival. The RAS proteins (KRAS and NRAS) are downstream components of this pathway. In tumors with RAS mutations, the pathway is active regardless of whether EGFR is blocked. This is why anti-EGFR therapy benefits only patients with RAS wild-type tumors. In RAS-mutant disease, blocking EGFR has no effect because the downstream signal is already permanently switched on [Goldberg and Kirkpatrick, 2005]. RAS mutation status is therefore a predictive biomarker: it identifies patients who will not benefit from anti-EGFR therapy.

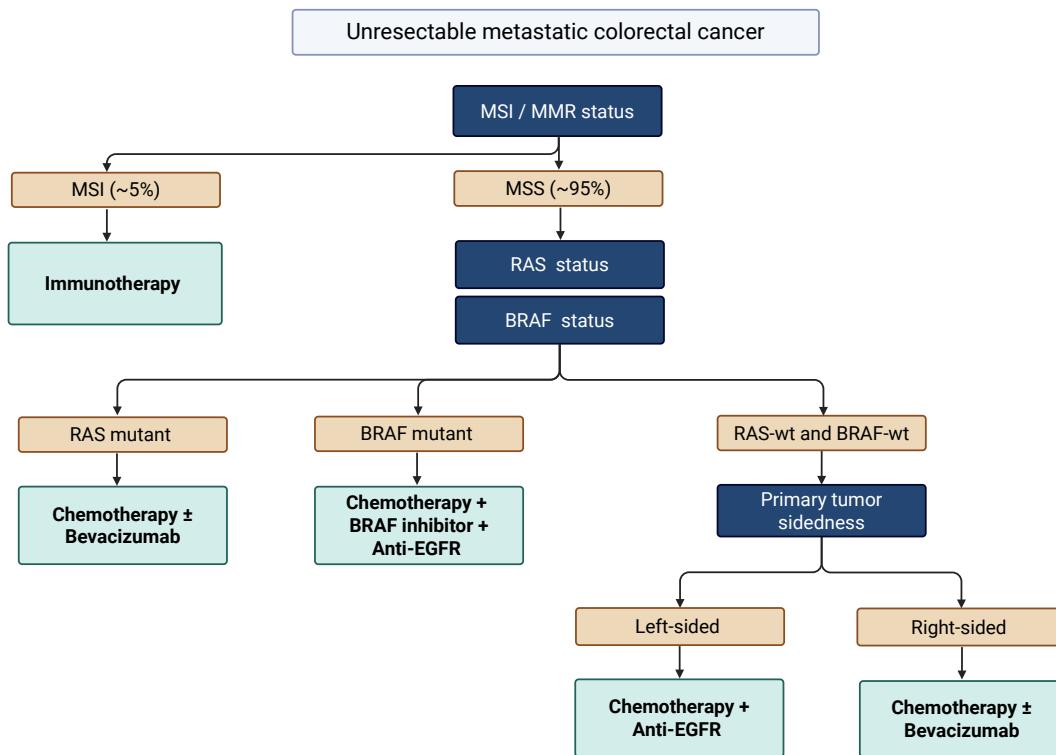
BRAF is another protein in the same signaling cascade, downstream of RAS. About 10% of mCRC patients have BRAF mutations, which significantly worsens prognosis with standard chemotherapy. These patients are now treated with a triplet combination of chemotherapy, BRAF inhibitors and anti-EGFR therapy, following clinical trials demonstrating improved outcomes compared to chemotherapy alone [Kopetz et al., 2024].

VEGF is a growth factor that promotes angiogenesis. Anti-angiogenic treatments target VEGF to prevent angiogenesis. The best known treatment is bevacizumab, an antibody that binds VEGF and prevents it from activating its receptor, thereby inhibiting angiogenesis. Unlike anti-EGFR therapy, there is currently no validated biomarker to predict which patients will benefit from bevacizumab. It is added to chemotherapy regimens empirically, and while it modestly improves survival in unselected populations, individual responses vary considerably [Loupakis et al., 2014].

Primary tumor sidedness has emerged as an additional factor guiding treatment selection. Left-sided tumors (descending colon, sigmoid, rectum) have better prognosis and respond more favorably to anti-EGFR therapy than right-sided tumors (cecum, ascending colon) [Cervantes et al., 2023, Morris et al., 2023]. The biological basis is not fully understood but likely relates to differences in molecular subtypes and embryological origin.

These criteria for treatment selection have been established through a series of landmark clinical trials over the past two decades [Di Nicolantonio et al., 2021]. The result is that metastatic colorectal cancer has become one of the success stories of precision oncology. We now define patient subtypes based on RAS status, BRAF status, MSI status, and primary tumor sidedness, and treatment is tailored accordingly. Despite this progress, many patients with favorable biomarker profiles do not respond as expected, and many with unfavorable profiles outlive predictions.

## COLORECTAL LIVER METASTASES



**Figure 3.2: Treatment landscape for colorectal liver metastases.** Simplified flowchart showing treatment decision-making based on resectability and molecular profiling. Resectable disease proceeds to surgery with or without perioperative chemotherapy. Unresectable disease is treated with systemic therapy, with regimen selection guided by RAS, BRAF, and MSI status. Anti-angiogenic therapy (bevacizumab) is added to chemotherapy in appropriate patients. Adapted from [Cervantes et al., 2023]. Created with BioRender.com.

## 3.4 Response Assessment and Its Limitations

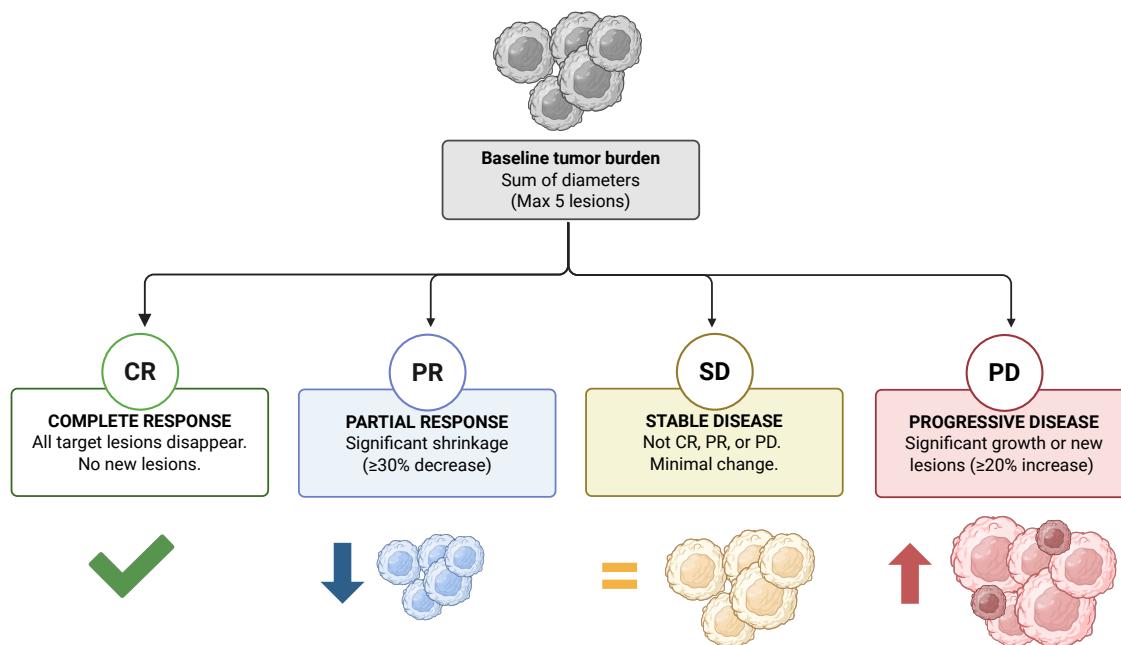
In routine clinical practice, treatment response assessment is primarily based on two sources of information: how the patient feels (symptoms, performance status, tolerance of treatment) and what imaging shows. For solid tumors, the imaging assessment is dominated by size. CT is the standard imaging modality for monitoring colorectal liver metastases. Scans are acquired at baseline and at regular intervals during treatment. Radiologists measure tumor dimensions, count lesions, and look for new disease. In day-to-day clinical practice, response assessment is often straightforward: if tumors are smaller, treatment is working; if tumors are larger or new lesions appear, it is not.

For clinical trials, this assessment is formalized through the Response Evaluation Criteria in Solid Tumors (RECIST). The original criteria were published in 2000 [Therasse et al., 2000] and were updated to version 1.1 in 2009 [Eisenhauer et al., 2009]. RECIST provides standardized definitions that enable comparison of treatment effects across studies. Under RECIST 1.1, up to five target lesions (maximum two per organ) are selected at baseline, each measuring at least 10 mm in longest diameter. The sum of these diameters constitutes baseline tumor burden. At subsequent assessments, percentage change determines the response category out of four options (Figure 3.3). CR requires disappearance of all target lesions. PR requires at least 30% decrease in sum of diameters. PD is defined as at least 20% increase

with a minimum 5 mm absolute increase, or the appearance of new lesions. SD encompasses everything in between.

RECIST has served clinical trials well by providing common language and enabling multi-center studies. However, it has well-recognized limitations [Burton, 2007, Nishino, 2018]. By focusing exclusively on size, RECIST creates a system where treatments that shrink tumors appear successful and treatments that alter tumor biology without changing size appear ineffective. RECIST was designed for a different era of oncology, one dominated by cytotoxic chemotherapy where treatment either killed cells and tumors shrank or it did not. As discussed in Section 3.3, modern oncology includes targeted agents, immunotherapies, and combination regimens that affect tumor composition in ways size alone cannot capture.

The gaps found in treatment selection and response assessment motivate the search for imaging biomarkers that capture heterogeneity [Gerwing et al., 2019]. The following chapter reviews how imaging heterogeneity has been defined and measured, from qualitative visual assessment through radiomics to habitat imaging, establishing the scientific context for the methods developed in this thesis.



**Figure 3.3: Response Evaluation Criteria in Solid Tumors (RECIST).** Illustration of RECIST 1.1 response categories based on percentage change in sum of longest diameters from baseline. Complete response (CR) requires disappearance of all target lesions. Partial response (PR) requires  $\geq 30\%$  decrease. Progressive disease (PD) is defined as  $\geq 20\%$  increase (with minimum 5 mm absolute increase) or appearance of new lesions. Stable disease (SD) encompasses changes that do not meet criteria for PR or PD. Created with BioRender.com.



# Chapter 4

## Imaging Heterogeneity: State of the Art

This chapter reviews how imaging heterogeneity has been approached, from qualitative visual assessment through radiomics to habitat imaging. The goal is to establish the scientific context for the work in this thesis and to identify the methodological gaps that motivate it. This chapter largely follows the comprehensive reviews carried out by O'Connor and colleagues [[O'Connor, 2017](#), [O'Connor et al., 2008, 2015](#)].

### 4.1 Definition and Historical Context

At the macroscopic scale of radiology, intratumor heterogeneity refers to variations in imaging phenotype observed within tumor. These are differences in intensity, texture, enhancement patterns, and spatial organization that may reflect underlying biological variation. In the literature, the term has been used inconsistently. Different authors have referred to imaging heterogeneity when describing abstract radiological phenotypes linked to CT density, global texture features computed across whole tumors, or subregions (i.e. habitats). In this thesis, we focus on the latter. We define imaging heterogeneity as the spatial variation in tissue composition that can be detected and quantified from imaging and that corresponds to known histopathological structures such as viable tumor, necrosis, and fibrosis.

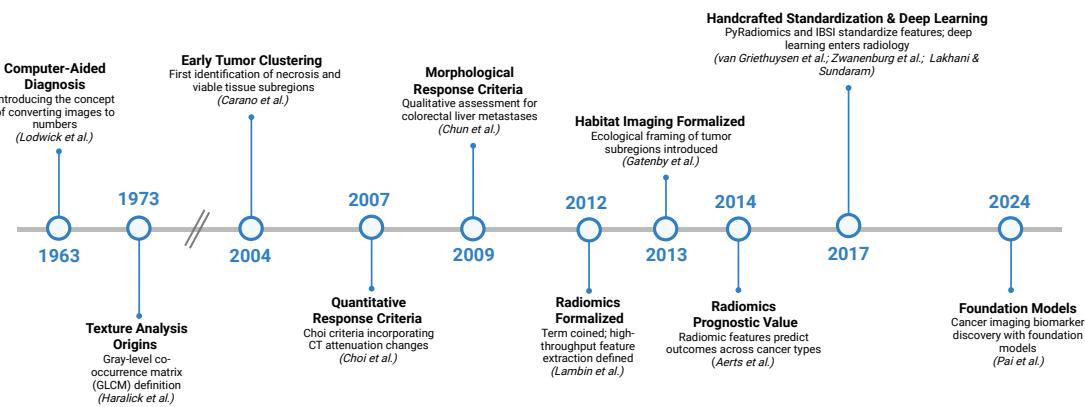
What the literature agrees on is twofold. First, clinical imaging is currently underutilized. We extract only a fraction of the information contained in routine scans, primarily size, location, and number of lesions. Second, tumor heterogeneity is clinically important. Heterogeneity is thought to drive treatment resistance and tumor adaptation. In the era of precision medicine, where patients receive targeted therapies requiring monitoring, non-invasive assessment of spatial and temporal heterogeneity could inform treatment decisions.

Approaches to quantifying imaging heterogeneity have evolved over time (Figure 4.1). The idea of using computers to extract quantitative information from medical images dates back to the 1960s, when Lodwick first applied computational analysis to chest radiographs and introduced the concept of computer-aided diagnosis [[Lodwick et al., 1963](#)]. A decade later, Haralick developed the gray-level co-occurrence matrix

## IMAGING HETEROGENEITY: STATE OF THE ART

(GLCM), establishing the mathematical foundation for texture analysis that would later support radiomics [Haralick et al., 1973]. Early applications of clustering to identify tumor subregions appeared in the 2000s [Carano et al., 2004], predating the formal concept of habitat imaging. Around the same time, quantitative approaches to response assessment emerged, including the Choi criteria incorporating CT attenuation changes [Choi et al., 2007] and morphological response criteria for colorectal liver metastases [Chun, 2009].

The field was formalized in the early 2010s. The term radiomics was coined in a seminal paper of 2012 [Lambin et al., 2012], rebranding texture analysis as a systematic, high-throughput approach to extracting quantifiable data from medical images. A year later, the concept of imaging habitats was introduced [Gatenby et al., 2013]. The landmark study by Aerts and colleagues [Aerts et al., 2014] demonstrated that radiomic features could predict clinical outcomes across multiple cancer types, catalyzing widespread interest in the field. Standardization followed with PyRadiomics [Van Griethuysen et al., 2017] and the Image Biomarker Standardization Initiative (IBSI) [Zwanenburg et al., 2020].



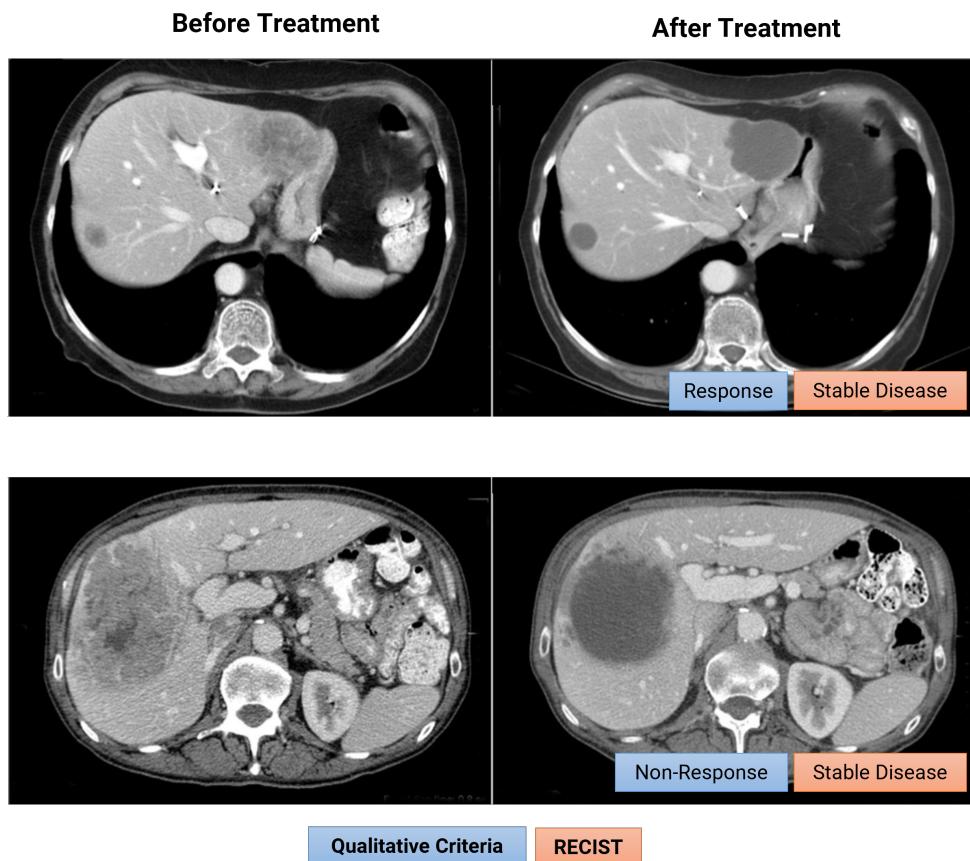
**Figure 4.1: Evolution of approaches to imaging heterogeneity.** Timeline illustrating the methodological progression in quantifying tumor heterogeneity from medical images. Early foundations include computer-aided diagnosis [Lodwick et al., 1963] and texture analysis mathematics [Haralick et al., 1973]. The 2000s saw early subregion analysis [Carano et al., 2004] and quantitative response criteria [Choi et al., 2007, Chun, 2009]. The field was formalized in the early 2010s with the definition of radiomics [Lambin et al., 2012] and habitats [Gatenby et al., 2013], followed by the demonstration of prognostic value across cancers [Aerts et al., 2014] and standardization initiatives [Van Griethuysen et al., 2017, Zwanenburg et al., 2020]. The late 2010s marked the entry of deep learning into radiology [Lakhani and Sundaram, 2017] and foundation models for cancer imaging biomarkers have emerged most recently [Pai et al., 2024]. The trajectory shows a shift from qualitative assessment and handcrafted features toward learned representations. Created with Biorender.com.

More recently, deep learning approaches have begun to complement or replace handcrafted features. Lakhani and Sundaram [2017] demonstrated the application of deep convolutional neural networks to chest radiograph interpretation, marking the entry of deep learning into mainstream radiology research [Lakhani and Sundaram, 2017]. Foundation models are now emerging in medical imaging, including models specifically designed for cancer imaging biomarker discovery [Pai et al., 2024]. These developments represent a shift from handcrafted features designed by humans to learned representations discovered by neural networks.

## 4.2 Qualitative and Conventional Quantitative Approaches

The simplest approach to imaging heterogeneity is qualitative visual assessment, the subjective evaluation of tumor characteristics by radiologists without explicit measurement. These are the classic features noted in radiology reports: morphology, enhancement patterns, and relationship with surrounding tissues. A radiologist might describe margins as irregular versus well-defined, enhancement as homogeneous versus heterogeneous, or note the presence of rim enhancement.

For colorectal liver metastases specifically, qualitative morphological criteria were proposed to improve response assessment [Chun, 2009]. The authors based the criteria on overall attenuation (heterogeneous, mixed, or homogeneous), tumor-liver interface (ill-defined, variable, or sharp), and peripheral rim of enhancement (present or absent). They observed that metastases treated with anti-angiogenic therapy often transform from heterogeneous, irregular lesions with ill-defined borders into homogeneous, hypoattenuating lesions with well-defined borders. Figure 4.2 shows an example of pre- and post-treatment CT scans of two patients response.



**Figure 4.2: Example of qualitative approach to imaging heterogeneity.** Pre- and posttreatment CT scans showing RECIST stable disease and morphological optimal response (top) and incomplete response (bottom). Optimal response is characterized by homogeneous attenuation and sharp tumor-liver interface. Adapted from [Chun, 2009].

Conventional quantitative approaches use objective numerical values derived from standard imaging. The most direct quantitative measure from CT is the Hounsfield unit value itself. In CECT, the difference in HU between pre-contrast and post-contrast images reflects contrast agent accumulation, which is related

to tissue vascularity. Studies have shown that changes in tumor attenuation can detect early anti-vascular effects of bevacizumab before size changes occur [Koukourakis et al., 2007]. A seminal study in this area is the work of Choi et al. [2007], who extended RECIST by incorporating HU changes for gastrointestinal stromal tumors (GIST) treated with imatinib, a targeted therapy. Similar to the therapies discussed in Section 3.3, imatinib is a cytostatic drug, inhibiting growth and division of cancer cells rather than killing them outright. In the study, a 15% reduction in HU on CECT was included alongside size criteria, acknowledging that for targeted therapies, compositional changes captured by attenuation matter. The use of Choi criteria has been suggested for metastatic colorectal cancer, but current evidence is insufficient to recommend them routinely [Zamboni et al., 2024].

However, these approaches have limitations. First, both approaches provide single-point measurements or global descriptions that fail to capture spatial complexity within tumors. Second, qualitative assessment suffers from inter-observer variability and HU values are sensitive to acquisition parameters (tube voltage, reconstruction kernel, contrast timing). For an imaging measurement to serve as a biomarker, it must be quantifiable and reproducible [O’Connor et al., 2017]. The qualitative observations and conventional quantitative measures described above showed that imaging heterogeneity contains prognostic information, but they lacked the standardization required for clinical use. The field needed a systematic effort to convert these observations into proper quantitative imaging biomarkers with demonstrated reproducibility. This came in the 2010s with the formalization of radiomics.

## 4.3 Radiomics

The idea of extracting quantitative features from medical images is not new. Historically, this work was known as texture analysis and included different methods for quantifying spatial patterns in images. The most widely used was the gray-level co-occurrence matrix (GLCM) [Haralick et al., 1973], which counts how often specific pairs of intensity values appear next to each other in an image. Texture analysis was applied to medical imaging throughout the 1990s and 2000s. Early studies demonstrated that CT texture features could predict survival in colorectal cancer [Miles et al., 2009], decode gene expression patterns in hepatocellular carcinoma [Segal et al., 2007], and correlate with histopathological characteristics in non-small cell lung cancer [Kawata et al., 2012].

What changed in the 2010s was not the underlying mathematics but the framing and infrastructure. The term *radiomics* was coined by Lambin and colleagues [Lambin et al., 2012], rebranding texture analysis as a systematic, high-throughput approach to extracting quantifiable data from medical images. The real catalyst for widespread adoption was PyRadiomics [Van Griethuysen et al., 2017], an open-source and user-friendly Python package that standardized the computation of handcrafted radiomic features. An additional effort to push radiomic features towards clinical biomarkers was the Image Biomarker Standardization Initiative (IBSI) [Zwanenburg et al., 2020]. The IBSI established consensus definitions for feature computation, addressing the problem that different groups had previously implemented features differently, making comparison across studies difficult.

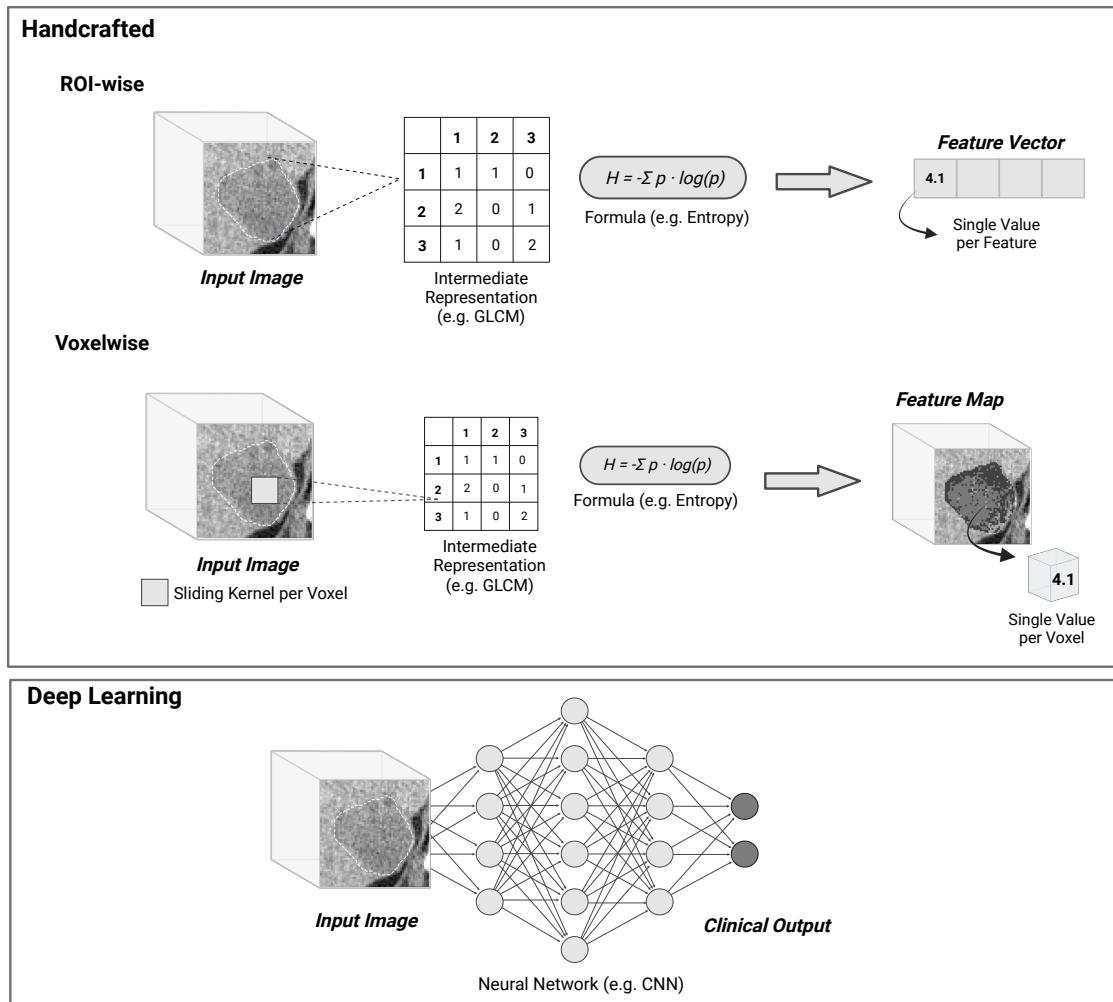
While radiomic features are broadly defined as any quantitative measures derived from radiological im-

ages, the term is most commonly used in the literature to denote *handcrafted* radiomic features (a convention largely shaped by the influence of PyRadiomics). Handcrafted features are derived from explicitly defined mathematical formulas or matrices designed by humans, such as the GLCM (Figure 4.3). They are typically grouped into three categories: shape features that describe tumor geometry (e.g., volume, surface area, sphericity), first-order features that summarize the distribution of voxel intensities (e.g., mean, variance, skewness), and texture features that quantify spatial relationships between voxels (e.g., entropy, contrast, homogeneity derived from GLCM and similar matrices).

The defining characteristic of handcrafted features is that they are designed a priori, based on domain knowledge, before observing the data. This is also their main limitation: their usefulness depends on whether the predefined features capture the patterns that are relevant for a given task. Traditional machine learning methods cannot adapt the feature representation itself [Goodfellow et al., 2016]: when applied directly to raw pixel data without carefully engineered inputs, they tend to perform poorly. Deep learning [LeCun et al., 2015] overcomes this limitation by learning hierarchical feature representations directly from data, progressively building complex concepts from simpler ones across multiple layers. Rather than requiring humans to specify which image patterns matter, deep neural networks learn representations optimized for the task at hand. More recently, foundation models [Bommasani et al., 2021] have emerged as a further development: large neural networks pretrained on vast datasets that learn general-purpose representations and can be adapted to specific tasks with minimal additional training. In the context of this thesis, which began in 2021, deep learning and foundation model approaches for radiomics were not yet sufficiently mature. The methods developed here therefore focus on handcrafted features, although comparisons with deep learning-based embeddings are presented in Chapter 7.

A relevant distinction in radiomics is how features are computed spatially (Figure 4.3). Computing features over the entire tumor, producing a single value per feature, is known as region-of-interest-wise (ROI-wise) or standard radiomics. This is the dominant approach because most radiomics studies use supervised learning, which requires fixed-length feature vectors as input. Voxelwise radiomics computes features locally, within a sliding kernel centred on each voxel. For example, to compute voxelwise GLCM entropy, the algorithm defines a neighbourhood around each voxel, computes entropy within that neighbourhood, and assigns the result to the central voxel. This produces a feature map rather than a single number, preserving where heterogeneity is located within the tumor. The standard radiomics pipeline is straightforward: extract features from segmented tumors, select or reduce features to avoid overfitting, and train a supervised model to predict a clinical endpoint.

The limitation of ROI-wise radiomics is that it implicitly assumes tumors are sufficiently homogeneous that averaging captures their essential properties. As discussed in Chapter 3, colorectal liver metastases are histologically heterogeneous, with viable tumor, necrosis, and fibrosis distributed in spatial patterns. Averaging over this heterogeneity may discard clinically relevant information.



**Figure 4.3: Schematic comparison of radiomics feature extraction methods.** Handcrafted Radiomics: Features are derived from explicit intermediate mathematical representations, such as the Gray-Level Co-occurrence Matrix (GLCM). In the ROI-wise approach (top), a single large GLCM is accumulated over the entire tumor mask, and a formula (e.g., for entropy) is applied once to yield a single global value. In the Voxelwise approach (bottom), a small, local GLCM is computed within a sliding kernel around each voxel. The formula is applied to this local matrix, and the resulting value is assigned to the center voxel, generating a spatial feature map. Deep Learning: The raw input image is passed directly through a neural network (e.g., a CNN) to learn abstract, hierarchical feature representations (embeddings) optimized for a specific task, without explicit human-defined formulas or intermediate matrix representations. Created with BioRender.com.

## 4.4 Habitat Imaging

The idea of partitioning tumors into spatially distinct subregions based on imaging characteristics predates the term *habitat*. In the 2000s and early 2010s, several studies already proposed the use of clustering algorithms to study intratumor heterogeneity (mainly necrosis and viable tumor) for several cancer types in preclinical models [Carano et al., 2004, Divine et al., 2016, Henning et al., 2007] using PET and MRI imaging modalities.

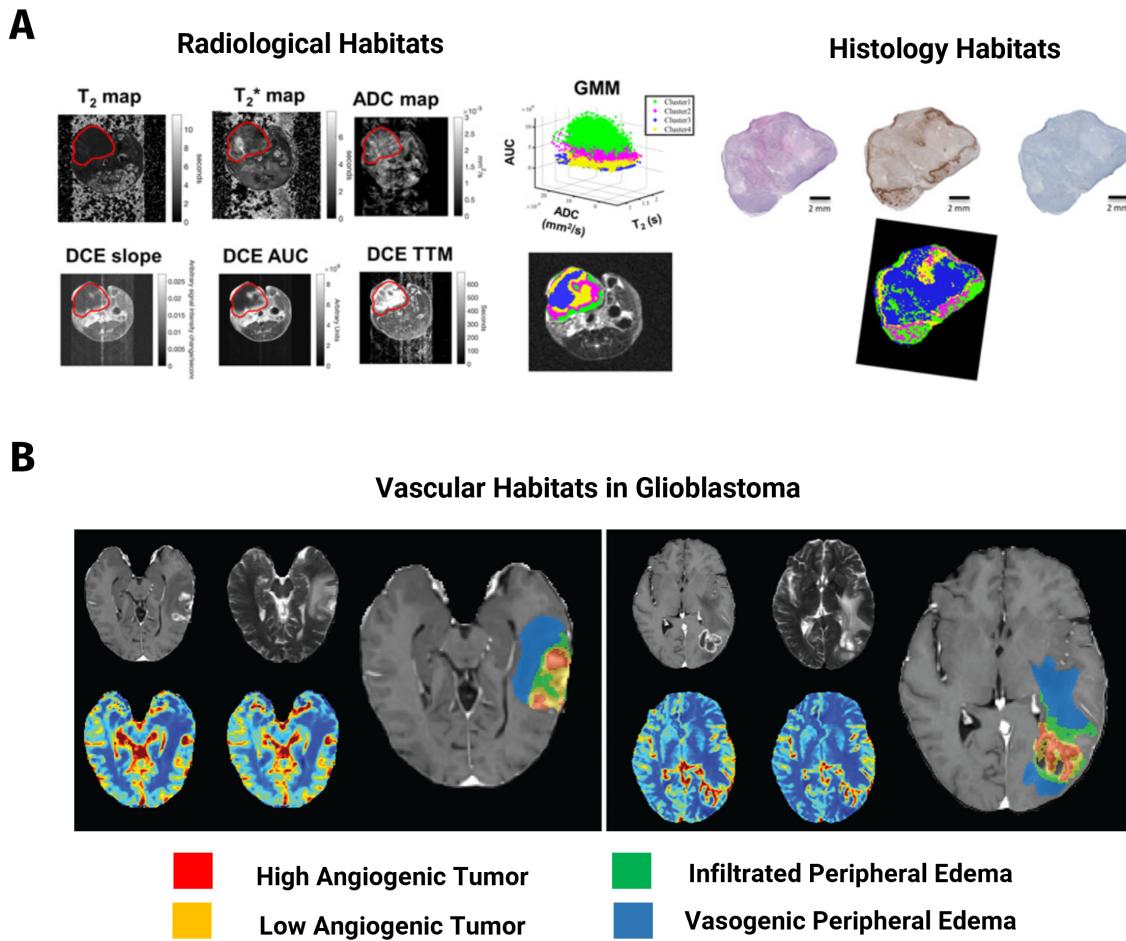
The term habitat and its ecological framing were introduced in 2013 [Gatenby et al., 2013]. Drawing on landscape ecology, they proposed that tumors could be understood as ecosystems containing distinct habitats with different environmental conditions. Just as a forest contains distinct microenvironments with different species compositions, a tumor contains regions with different blood flow, oxygenation, and cellular populations. According to the authors, the heterogeneous environment may drive tumor evolution (e.g., hypoxic habitats select for cells with aggressive, therapy-resistant phenotypes) and therefore habitats might be worth studying. This ecological framing was subsequently applied to breast cancer DCE-MRI by the same group [Chaudhury et al., 2015], establishing the conceptual foundation for the field. The field remained relatively small throughout the 2010s, with most radiomics research focusing on supervised learning approaches that required whole-tumor feature vectors.

To date, most habitat imaging studies have used MRI rather than CT. This is mainly due to the richer biological information available from multiparametric MRI: diffusion-weighted imaging provides the apparent diffusion coefficient (ADC), which correlates with cellularity; dynamic contrast-enhanced sequences yield perfusion parameters such as  $K^{trans}$  and  $v_e$ ; and T2-weighted imaging reflects tissue water content and edema. These parameters have known relationships to tissue biology, making habitat interpretation more straightforward.

This advantage was first demonstrated in glioblastoma, where Katiyar and colleagues [Katiyar et al., 2017] applied spatially regularised spectral clustering to mpMRI in mouse xenografts, achieving strong correlation with histological tissue fractions of necrosis and viable tumor. This was later shown in breast cancer mouse models [Jardim-Perassi et al., 2019], where habitats derived from six MRI parameters were validated against co-registered histology using 3D-printed tumor moulds (Figure 4.4). The resulting habitats corresponded to viable-normoxic, viable-hypoxic, nonviable-hypoxic, and nonviable-normoxic tissue types, validated against H&E staining, pimonidazole (a hypoxia marker), and CD31 (a vasculature marker). Beyond biological validation, imaging habitats have been shown to correlate with clinical outcomes, mainly in glioblastoma [Waqar et al., 2022]. The most prominent study computed vascular habitats from perfusion MRI and demonstrated that they were highly prognostic at presurgery [Juan-Albarracín et al., 2018].

CT-based habitat imaging has received less attention across all tumor types, likely because CT provides fewer biologically interpretable parameters. CECT reflects vascularity, and Hounsfield units correlate broadly with tissue density, but the relationship between CT intensity and specific tissue types is less direct than for MRI-derived parameters.

Two studies have applied habitat imaging to colorectal liver metastases, both using MRI rather than CT



**Figure 4.4: Examples of MRI-based habitat imaging.** (A) Preclinical breast cancer model illustrating biological validation of imaging habitats. Habitats were defined by clustering six MRI parameters ( $T_2$ ,  $T_2^*$ , ADC, and DCE kinetic metrics) using a Gaussian Mixture Model. The resulting spatial maps (bottom center) were validated against co-registered histology (right), where habitats correlated with specific biological microenvironments: viable-normoxic (green), viable-hypoxic (magenta/yellow), and nonviable/necrotic (blue) regions, confirmed by H&E, pimonidazole (hypoxia), and CD31 (vasculature) staining. Adapted from Jardim-Perassi et al., 2019. (B) Glioblastoma illustrating vascular habitats derived from perfusion MRI. Tumors are partitioned into four distinct vascular habitats based on perfusion characteristics: High Angiogenic Tumor (HAT, red), Low Angiogenic Tumor (LAT, yellow), Infiltrated Peripheral Edema (IPE, green), and Vasogenic Peripheral Edema (VPE, blue). These habitats have been shown to differ in survival prognosis. Adapted from Juan-Albarracín et al., 2018.

as well. In one study, researchers used a semi-automated clustering approach based on PCA of DCE-MRI contrast uptake curves to define regions of viable and non-viable tumor in 14 patients with CRLM scheduled for hepatic resection [Franklin et al., 2020]. Imaging-derived subregions showed good concordance with spatially matched histology (mean Dice similarity coefficient 0.74 for viable tumor), and pharmacokinetic parameters differed significantly between viable and non-viable subregions. In [Katiyar et al., 2023], the authors applied a cross-species approach, training multi-view spectral clustering on preclinical PET-MRI and applying it to six CRLM lesions from four patients undergoing resection, with tissue fractions correlating with histology. These studies provide proof-of-concept that imaging habitats can reflect meaningful biological phenotypes in CRLM, but sample sizes were small and the imaging modalities (DCE-MRI, PET-MRI) are not routinely used for CRLM management.

Unlike supervised radiomics, where the pipeline is relatively standardised (feature extraction, feature selection, model training), there is no consensus pipeline for habitat imaging. Published studies vary considerably in their methodological choices, and these differences affect results in ways that are not fully understood (Table 4.1).

Studies differ in the features used for clustering: some use raw intensity values from single or multiple imaging sequences, others use derived parameters such as ADC or Ktrans, and still others compute voxelwise texture features in local neighbourhoods before clustering. K-means clustering dominates the literature due to its computational efficiency and interpretability, though Gaussian mixture models offer probabilistic assignments that may better reflect gradual transitions between tissue types. The choice of cluster number ( $k$ ) remains a central challenge, with some studies fixing  $k$  a priori based on expected tissue types and others using data-driven criteria such as the Calinski and Harabasz score or the Bayesian Information Criterion [Schubert, 2023]. Once habitats are defined, studies extract different metrics for downstream analysis, ranging from simple volume fractions to texture features within each habitat or spatial diversity indices. This methodological heterogeneity makes comparison across studies difficult and complicates efforts to establish habitat imaging as a reliable tool.

The lack of standardisation would be less problematic if habitat methods were rigorously validated. However, validation remains the most significant gap in the literature. The imaging biomarker roadmap for cancer studies [O'Connor et al., 2017] describes three types of validation: technical (is the measurement reproducible?), biological (does it reflect true tissue properties?), and clinical (does it predict outcomes independently of established factors?). Most habitat studies fall short on all three criteria.

Technical validation, such as test-retest reproducibility or multi-site stability, is reported in only a minority of studies. Biological validation against histopathology requires resected specimens and careful co-registration, which few studies achieve; most rely on indirect evidence such as correlation with known physiology or gene expression patterns. Clinical validation typically consists of survival analysis, but external validation in independent cohorts is rare, and few studies benchmark habitat metrics against tumor volume, a strong prognostic factor that any new imaging biomarker should demonstrate added value beyond.

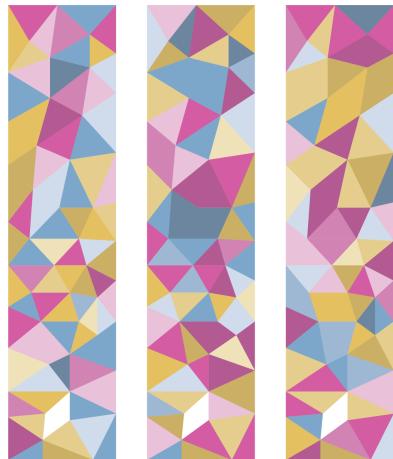
Habitat imaging is virtually unexplored for colorectal liver metastases despite strong biological rationale: the histopathological heterogeneity of CRLM (viable tumor, necrosis, fibrosis) and its prognostic significance suggest that spatial imaging phenotypes may carry clinical information. The absence of CT-based

## IMAGING HETEROGENEITY: STATE OF THE ART

habitat studies in CRLM, combined with the broader validation gaps, motivated the work in this thesis. The following chapters address these gaps through a staged approach: Chapter 5 develops a standardised pipeline for CT-based habitat imaging in CRLM, Chapter 6 establishes technical precision through feature stability analysis, Chapter 7 provides biological validation using multiparametric MRI as a reference, and Chapter 8 assesses clinical utility by benchmarking habitat metrics against volume across treatment contexts.

**Table 4.1: Summary of selected habitat imaging studies.** Studies are listed chronologically. Technical validation refers to test-retest reproducibility or multi-site stability; biological validation refers to correlation with histopathology or known tissue biology; clinical validation refers to outcome prediction. GBM: glioblastoma; GMM: Gaussian mixture model.

Study	Cancer type	Modality	Method	k	Tech. Val.	Biol. Val.	Clin. Val.
[Carano et al., 2004]	Colorectal (preclinical)	MRI	K-means	4	No	Yes (histo.)	No
[Henning et al., 2007]	Sarcoma (preclinical)	MRI	K-means	4	No	Yes (histo.)	No
[Chaudhury et al., 2015]	Breast	MRI	K-means	4	No	No	Yes (molecular)
[Divine et al., 2016]	Breast (preclinical)	PET/ MRI	GMM	5	No	Yes (histo.)	No
[Katiyar et al., 2017]	GBM (preclinical)	MRI	Spectral	3	No	Yes (histo.)	No
[Juan-Albarracín et al., 2018]	GBM	MRI	K-means	4	No	Yes (vascular)	Yes (survival)
[Jardim-Perassi et al., 2019]	Breast (preclinical)	MRI	GMM	4	Yes (repeat.)	Yes (histo.)	No
[Franklin et al., 2020]	CRLM	MRI	K-means	2	No	Yes (histo.)	No
[Katiyar et al., 2023]	CRLM	PET/ MRI	Spectral	2	No	Yes (histo.)	No



## Methodological Framework

*“Science is a bit like the joke about the drunk who is looking under a lamppost for a key that he has lost on the other side of the street, because that’s where the light is. It has no other choice.”*

— Noam Chomsky

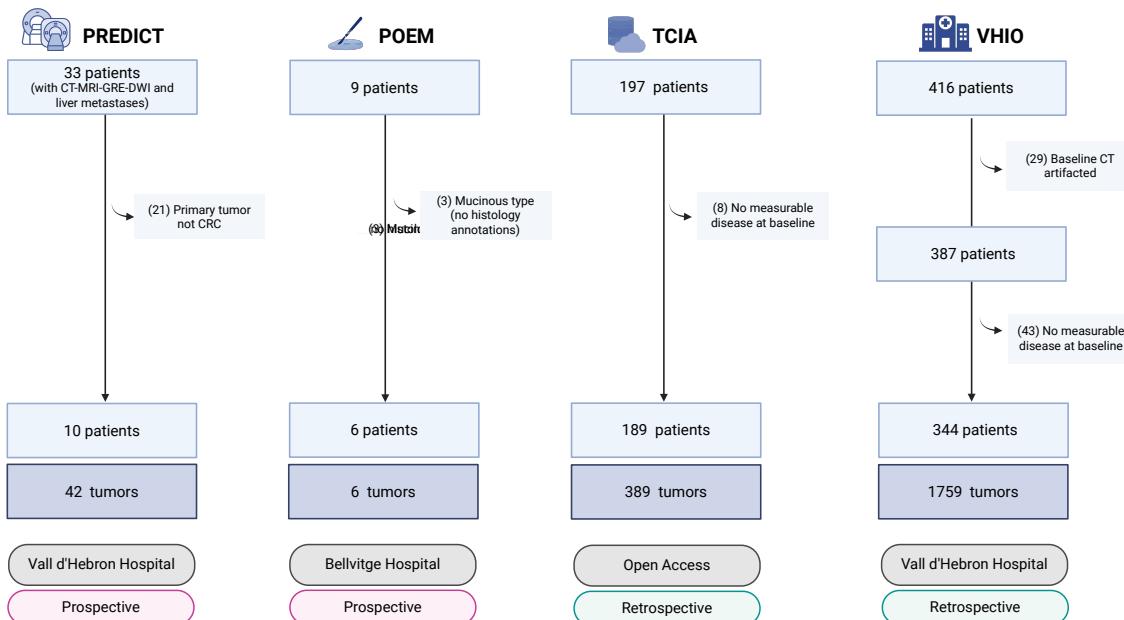


# Chapter 5

## Data and Habitat Imaging Pipeline

This chapter describes the datasets and methods that support the experiments carried out in this thesis. We begin by introducing the four main patient cohorts and their role in habitat model development, validation, and clinical application (Section 5.1). We then present the habitat imaging pipeline (Section 5.2), our end-to-end framework for computing tumor habitats from medical images.

### 5.1 Patient Cohorts



**Figure 5.1: Overview of patient cohorts used in this thesis.** Prospective cohorts provide multiparametric imaging (PREDICT) and histology (POEM) for developing and validating the biologically-informed habitat model. Retrospective cohorts (blue) provide large-scale CT imaging with clinical outcomes for applying the habitat model and assessing its clinical relevance. Numbers indicate patients and tumors remaining after exclusions.

This thesis is based on four main datasets of colorectal cancer patients with liver metastases, each serv-

## DATA AND HABITAT IMAGING PIPELINE

ing a distinct role in the development, validation, and clinical application of CT-derived tumor habitats (Table 5.1, Figure 5.1). For the development and biological validation of the habitat model we used two small prospective cohorts: PREDICT (n=10 patients, 42 tumors) provides co-registered multiparametric MRI (mpMRI) to anchor CT-derived habitats to quantitative imaging proxies of vascularity and cellularity, while POEM (n=6 patients, 6 tumors) provides whole-tumor histopathology for independent validation. Two large retrospective cohorts—TCIA (n=189 patients, 389 tumors) and VHIO (n=343 patients, 1759 tumors)—provide portal-phase CT imaging alongside clinical outcomes and are used to apply the biologically-informed habitat model (developed in Chapter 7) and assess its prognostic value (Chapter 8). Detailed demographic and clinical characteristics of TCIA and VHIO are provided in Chapter 8.

Chapter 6 presents a broader technical study assessing the precision of 3D voxelwise radiomics features using additional retrospective data from both liver and lung lesions. That analysis established which handcrafted features are sufficiently robust for habitat computation and informed the feature selection strategy used in Chapter 7. The datasets described here—PREDICT, POEM, TCIA, and VHIO—form the foundation for the biologically-informed habitat model and its clinical translation, which are the central contributions of this thesis.

**Table 5.1: Summary of datasets and their role in this thesis.** PREDICT provides co-registered mpMRI for biologically-informed habitat model development; POEM provides histology for independent validation. TCIA and VHIO provide large-scale CT imaging with survival outcomes for applying the habitat model and testing its prognostic value. OS: overall survival; PFS: progression-free survival; DFS: disease-free survival; mpMRI: multiparametric MRI; HE: hematoxylin and eosin.

Dataset	N patients	N tumors	Imaging modalities	Clinical outcome	Primary use
PREDICT	10	42	CT (Portal Phase), MRI	—	Habitat model development (mpMRI-anchored) and validation (Chapter 7)
POEM	6	6	CT (Portal Phase), Histology (HE)	—	Biological validation of habitat model (Chapter 7)
TCIA	189	389	CT (Portal Phase)	OS, DFS	Application of habitat model and clinical outcome modeling (Chapter 8)
VHIO	343	1759	CT (Portal Phase)	OS, PFS	Application of habitat model and clinical outcome modeling (Chapter 8)

All cohorts were collected as part of ongoing studies approved by the Vall d'Hebron University Hospital Ethics Committee: PR(AG)29/2020 (PREDICT), PR(AG)305/2024 (POEM), PR(AG)335/2018 and PR(AG)500/2023 (VHIO). Patients provided written informed consent to participate in such studies.

### 5.1.1 CT Image Preprocessing

All CT images were acquired in the portal venous phase with intravenous contrast, tube voltage of 100-130kV, slice thickness of 1.25-5mm and pixel size 0.7-1.1mm<sup>2</sup>, and stored in DICOM format (*Digital Imaging and Communications in Medicine (DICOM) Standard* [National Electrical Manufacturers As-

sociation (NEMA), 2025]). PREDICT and VHIO scans were retrieved from the Vall d’Hebron Hospital imaging archive (Picture Archiving and Communication System, PACS), anonymized with an identifiable patient ID, and transferred to a password-protected VHIO storage server. TCIA scans were downloaded from the public repository (<https://www.cancerimagingarchive.net/>). POEM scans were provided through a collaborative project with Hospital Universitari de Bellvitge (see Section 5.1.3.) DICOM files were converted to NIfTI (Neuroimaging Informatics Technology Initiative) format using dcsm2niix [Li et al., 2016],

Images underwent qualitative assessment by both the author and the group’s experienced radiologists to ensure acquisition quality and suitability for analysis. Scans were excluded if they were non-contrast-enhanced, acquired outside the portal venous phase, did not include the entire liver, or showed no measurable disease at baseline (defined as liver lesions <1 cm in diameter or absent liver metastases). Figure 5.1 summarizes exclusion counts for each cohort.

For the majority of the scans, tumor segmentation was performed manually using 3D Slicer [Fedorov et al., 2012]. For PREDICT and POEM, all well-defined target lesions per patient were delineated by the group’s experienced radiologists. For VHIO cohort, a third of the tumors (~600) were manually delineated by the author and supervised by an experienced radiologist. The rest were automatically delineated by SALSA, an automated liver tumor segmentation tool developed in our lab during the course of this thesis, with manual correction when needed. For TCIA, manual segmentations were performed by radiologists as part of the original data collection [Simpson et al., 2024]

Voxel resampling and intensity preprocessing were performed during feature extraction (see Methods Section of Chapters 6 and 7), ensuring consistent spatial resolution and intensity normalization across all voxelwise feature maps.

## 5.1.2 Multiparametric MRI Preprocessing (PREDICT)

The PREDICT cohort included contrast-enhanced CT alongside mpMRI acquired on two scanners: a 1.5T Siemens Avanto and a 3T GE SIGNA Pioneer. The MRI protocol included high-resolution anatomical imaging (T2-weighted and T1-weighted), advanced diffusion MRI with multiple b-values and echo times, variable flip angle spoiled gradient echo (SGrE) imaging for T1 mapping, and dynamic contrast-enhanced (DCE) MRI. Full acquisition parameters are provided in Appendix A.

MRI preprocessing was performed by MRI physicist F. Grussu, and included denoising (MP-PCA) [Veraart et al., 2016], Gibbs unringing [Kellner et al., 2016], motion correction via affine coregistration [Ourselin et al., 2001], and EPI distortion correction using reversed phase-encoding acquisitions [Astrand et al., 2003]. Anatomical images underwent N4 bias field correction based on ANTs [Avants et al., 2008]. Tumor segmentations were drawn manually by an experienced radiologist on the T2-weighted scan, aided by visual inspection of all MRI contrasts.

From the preprocessed MRI data, thirteen multiparametric (mpMRI) maps were derived using biophysical models (Table 5.2). Diffusion-relaxation MRI was modeled using two approaches: (1) a two-pool system capturing tissue and vascular water contributions, yielding apparent diffusion coefficients ( $ADC_t$ ,

## DATA AND HABITAT IMAGING PIPELINE

$ADC_v$ ), corresponding to tissue and vascular water contributions, transverse relaxation times ( $T_{2t}$ ), vascular signal fraction ( $f_v$ ), and tissue kurtosis excess ( $K_t$ ); and (2) an advanced microstructural model [Grussu et al., 2025] incorporating intra- and extracellular compartments, yielding intrinsic diffusivity ( $D_0$ ), volume-weighted cell size ( $vCS$ ), intracellular fraction ( $f_{in}$ ), and cell density. Variable flip angle imaging provided total longitudinal relaxation time ( $T_1$ ). DCE MRI was fitted using a two-compartment Tofts model [Tofts, 1997], yielding the capillary permeability constant ( $K^{trans}$ ), plasma volume fraction ( $v_p$ ), and extracellular-extravascular volume fraction ( $v_e$ ). Model fitting was performed using custom Python code available at <https://github.com/fragrussu/MyRelax> and <https://github.com/fragrussu/bodymritools>.

**Table 5.2: Multiparametric MRI maps used for CT habitat development.** Thirteen quantitative maps were derived from diffusion-relaxation MRI, variable flip angle T1 mapping, and dynamic contrast-enhanced MRI. ADC: Apparent Diffusion Coefficient; T2: Transverse Relaxation Time; EES: Extravascular Extracellular Space; SGrE: Spoiled Gradient Echo; DCE: Dynamic Contrast-Enhanced.

mpMRI metric	Computed from	Units	Description
$ADC_t$	Tissue ADC	$\mu\text{m}^2/\text{ms}$	Apparent diffusivity of water in tissue (excluding vasculature)
$ADC_v$	Vascular ADC	$\mu\text{m}^2/\text{ms}$	Apparent diffusivity in vascular compartment (IVIM effect)
$K_t$	Tissue kurtosis excess	Dim.less	Quantifies non-Gaussian diffusion due to heterogeneity or restriction
$f_v$	Vascular signal fraction	Norm.	Signal arising from vascular compartment (IVIM)
$T_{2t}$	Tissue $T_2$	ms	Transverse relaxation time of tissue (excluding vasculature)
$D_0$	Intrinsic diffusivity	$\mu\text{m}^2/\text{ms}$	Intrinsic intracellular diffusivity
$vCS$	Vol-weighted cell size	$\mu\text{m}$	Average cell size weighted by cell volume
$f_{in}$	Intracellular fraction	Norm.	Fraction of signal arising from intracellular water
$CD$	Cell density	Cells/ $\text{mm}^3$	Estimated cell density (derived from $f_{in}$ and $vCS$ )
$T_1$	$T_1$	ms	Total longitudinal relaxation time
$T_2^*$	$T_2^*$	ms	Effective transverse relaxation time
$K^{trans}$	Capillary permeability	$\text{min}^{-1}$	Influx mass transfer rate of contrast agent
$v_e$	EES volume	Norm.	Volume of extracellular, extravascular space per unit tissue volume

We harmonized mpMRI metrics across scanners with a custom-written ComBat [Fortin et al., 2018] implementation, rescaling metrics obtained on the 1.5T system to the 3T range. Quality control ensured

that all mpMRI values were biologically plausible. For each metric, we established a valid range based on empirical knowledge (Appendix A). Voxels within this range were included, those marginally outside ( $\pm 10\%$ ) were clipped to the boundary, and those beyond this tolerance were excluded.

### 5.1.3 Histopathology Workflow (POEM)

The POEM cohort (PrOspEctive Metastases Imaging) was collected as part of a multicenter observational study led by Dr. Peter Vermeulen (GZA Hospitals, Belgium) in collaboration with Hospital Universitari de Bellvitge (Spain). The study aims to develop radiomic signatures of histopathological growth patterns [Latacz et al., 2021] in liver metastases from colorectal and breast cancer using standard-of-care CT and MRI alongside complete histopathological sampling [Latacz et al., 2024]. Our analysis focused on colorectal liver metastases from patients treated at Hospital Universitari de Bellvitge who underwent complete tumor resection.

The histopathology protocol was designed to preserve spatial orientation for comparison with preoperative imaging. Surgeons applied orientation marks to the resected specimen, allowing the pathologist to slice the tissue in transverse (axial) planes corresponding to CT and MRI imaging planes. Slices were cut at 0.5–1 cm thickness, photographed, and sampled systematically. Tissue samples were formalin-fixed, paraffin-embedded, and stained with hematoxylin and eosin (H&E). Whole-slide images were acquired using a 3DHISTECH Pannoramic-series scanner, producing MRXS files.

Histological annotations of necrosis, viable cancer and fibrosis/stroma were automatically performed using QuPath [Bankhead et al., 2017], an open-source digital pathology platform, and supervised by an experienced pathologist. An experienced pathologist manually delineated regions of necrosis, fibrosis, and viable tumor on digitized whole-slide images. These annotations provided qualitative ground truth for validating CT-derived habitats (Chapter 7). We attempted spatial coregistration between histology slices and corresponding CT slices but found it impractical due to tissue deformation during resection and processing. As a result, histology was used for qualitative biological validation rather than voxel-by-voxel spatial comparison.

## 5.2 Habitat Imaging Pipeline

Habitat imaging is becoming increasingly popular in oncology as a way to segment tumors into meaningful radiological subregions. Yet despite a growing literature, there is no standard tool for computing habitats. Each study implements its own custom-built codes, which are rarely shared and in the case they are, poorly documented. Any researcher interested in habitat imaging faces the same challenges repeatedly: how to preprocess voxelwise features, how to choose the clustering algorithm, how to ensure habitat labels across patients are consistent so that “Habitat 1” means the same thing in every case, etc.

These are not trivial implementation choices. A pipeline that ignores voxel spacing differences will produce habitats that reflect acquisition parameters rather than biology. A pipeline that clusters all tumors together may dilute tumor-specific heterogeneity into noise. The experiments carried out in Chapter 6

## DATA AND HABITAT IMAGING PIPELINE

made these challenges concrete for us. Computing habitats for the precision analysis requires solving each of these problems from scratch, and the solutions were tied to the specific dataset and analysis.

To enable the work in Chapters 7 and 8, and to support future research, we developed a modular, documented pipeline that addresses these issues systematically. The pipeline is image-modality agnostic, being applicable to any voxelwise imaging data (CT, MRI, PET, etc.) and works with both 2D and 3D data. It also includes key design choices as configurable parameters rather than hardcoded assumptions. We are currently preparing the pipeline for public release as an open-source tool, available at <https://github.com/radiomicsgroup/imaging-habitats-pipeline>. We share it with the hope that it saves other researchers the substantial time and effort required to build habitat imaging workflows from scratch, and that it helps standardize methods across studies to improve reproducibility and comparability in the field.

### 5.2.1 Pipeline Overview

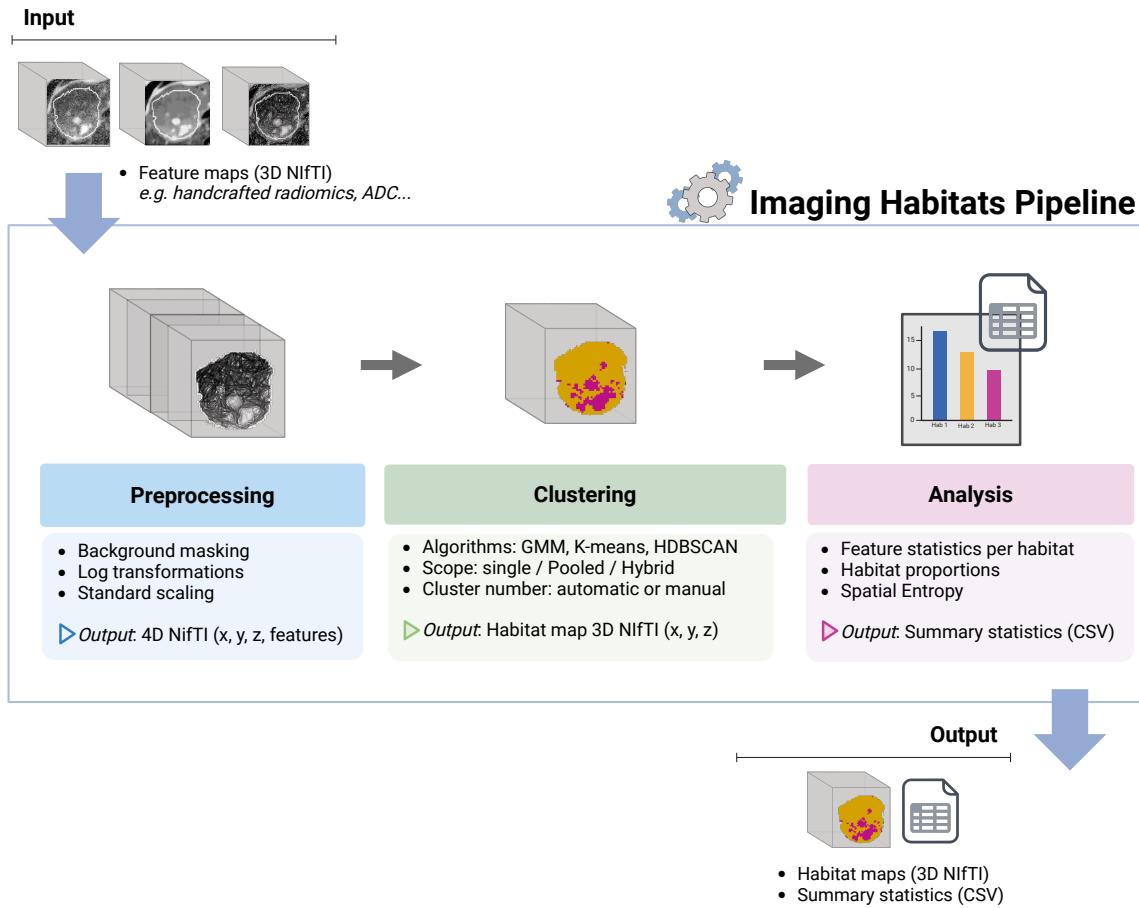
The pipeline consists of three modules: Preprocessing, Clustering, and Analysis (Figure 5.2):

- **Preprocessing** takes voxelwise feature maps (e.g., Kurtosis, ADC, Coarseness) as 3D NIfTI files, removes background voxels, enforces consistent array shapes across maps, and applies feature scaling. The output is a 4D NIfTI volume per tumor (dimensions: x, y, z, features), with both scaled and unscaled versions preserved for transparency. Background voxels are automatically detected and masked out, and all feature maps are checked for spatial alignment.
- **Clustering** takes the preprocessed 4D volumes and assigns each voxel to a habitat. The module supports multiple algorithms (Gaussian Mixture Models, K-means, Bayesian GMM, HDBSCAN) and allows the number of clusters to be set manually or selected via data-driven criteria (BIC, silhouette score). Cluster labels are integers starting at 1; background is reserved as 0.
- **Analysis** computes habitat-derived metrics (described in Section 5.2.4) and exports summary statistics as CSV files for downstream analysis.

The pipeline is implemented in Python and can be run from the command line. All random seeds are exposed for reproducibility, and outputs are organized by experiment name to support version control.

### 5.2.2 Key Implementation Choices

**Voxel resampling.** Texture features describe spatial relationships between neighboring voxels, which means their values depend on voxel dimensions. A feature computed on a  $0.7 \times 0.7 \times 1.0$  mm grid is not directly comparable to the same feature computed on a  $0.9 \times 0.9 \times 2.5$  mm grid. For cross-tumor comparisons to be meaningful, input images should be resampled to a common isotropic resolution before feature extraction. The pipeline itself does not yet perform resampling—it expects preprocessed inputs—but users working with heterogeneous acquisitions should resample upstream since results may reflect acquisition differences rather than biology.

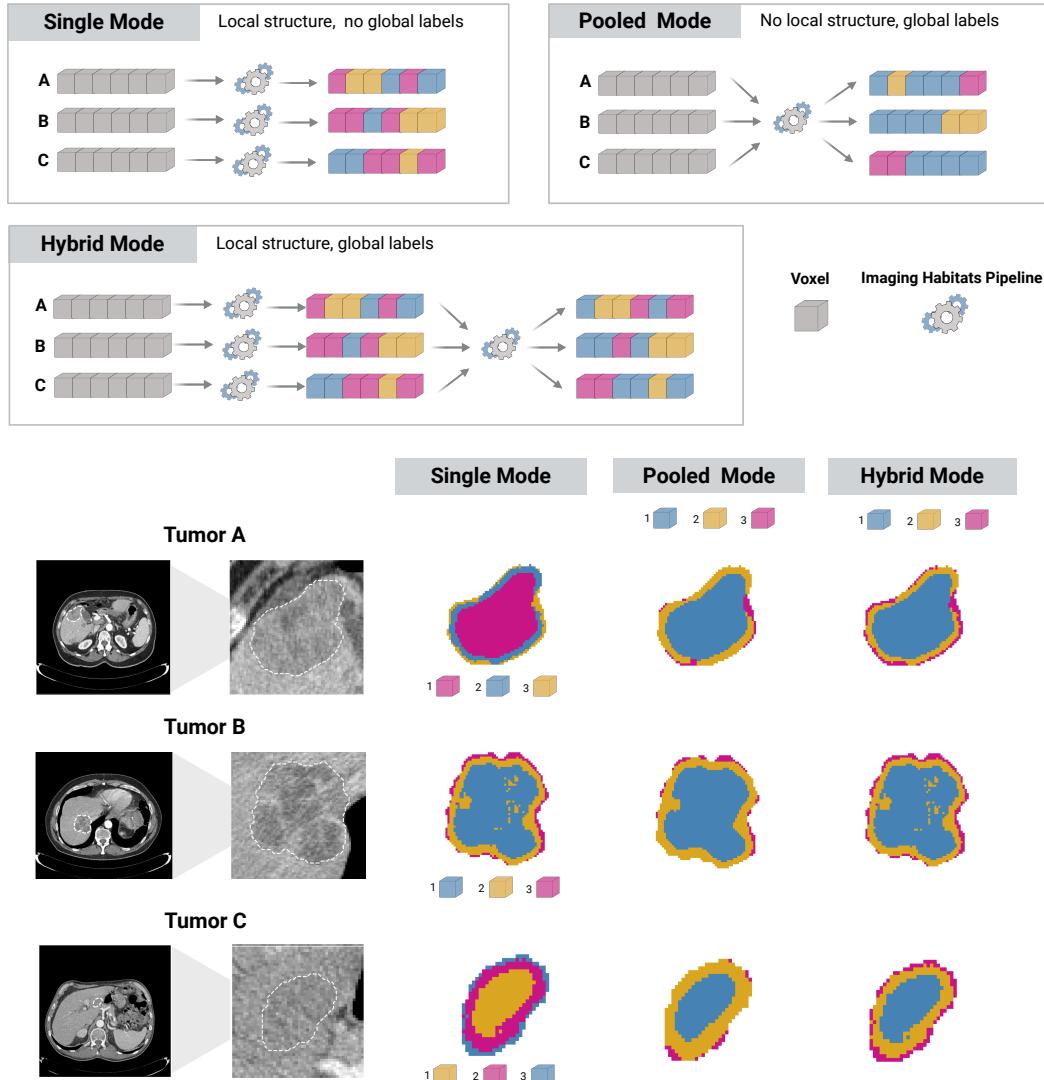


**Figure 5.2: Habitat imaging pipeline overview.** The pipeline consists of three modules. Preprocessing removes background voxels, applies optional log transformation, and performs standard scaling, outputting a 4D NIFTI volume per tumor. Clustering assigns voxels to habitats using GMM, K-means, Bayesian GMM, or HDBSCAN, with automatic or manual selection of cluster number. Three clustering scopes are available: single-tumor, pooled, or hybrid (Figure 5.3). Analysis computes habitat-derived metrics (proportions, entropy, rim/core statistics) and exports summary tables.

**Feature scaling.** Clustering algorithms are sensitive to feature magnitudes. If one feature ranges from 0 to 1 and another from 0 to 10,000, the latter will dominate distance calculations regardless of its biological relevance. We apply standard scaling (zero mean, unit variance) to all features before clustering. The scaler can be fit per tumor, per patient, or across the full cohort, depending on the analysis goal. For the studies in this thesis, we fit a single scaler across all voxels being processed, ensuring that habitat definitions are comparable across tumors.

**Log transformation.** Some imaging metrics span several orders of magnitude and exhibit strong right skew—ADC values, T1 relaxation times, and permeability coefficients are common examples. Clustering such features in their raw form can produce habitats driven by outliers rather than typical tissue properties. We assess each feature’s distribution (skewness, dynamic range, coefficient of variation) and apply log transformation where appropriate, excluding bounded fractions (e.g., volume fractions constrained to [0,1]) to preserve their probabilistic interpretation. This is not a universal preprocessing step; it is a decision that must be made feature by feature, with the goal of presenting the clustering algorithm with distributions that reflect meaningful biological variation rather than measurement artifacts.

## DATA AND HABITAT IMAGING PIPELINE



**Figure 5.3: Clustering scope: single-tumor, pooled, and hybrid modes.** (A) Single-tumor mode. Each tumor is clustered independently. Labels are arbitrary and not comparable across tumors. (B) Pooled mode. All voxels are clustered together, producing globally consistent labels. However, local heterogeneity may be diluted. (C) Hybrid mode. Tumors are first clustered locally, then local cluster centroids are pooled and meta-clustered to define global prototypes. Each local cluster is mapped to its nearest prototype, yielding consistent labels while preserving local structure.

**Clustering algorithm and number of clusters.** The pipeline supports multiple clustering algorithms: Gaussian Mixture Models (GMM), K-means, and HDBSCAN. GMM is our default because it provides probabilistic (soft) assignments that better capture gradual transitions between tumor phenotypes, which are biologically plausible given partial volume effects and mixed tissue interfaces. However, users may prefer hard assignments or density-based clustering depending on their application. The number of clusters can be specified manually or determined automatically using different criteria such as the Bayesian Information Criterion (BIC) or the Silhouette Score.

**Clustering Scope:** A critical design choice is whether to cluster each tumor individually or pool all tumors together. The pipeline supports three modes (Figure 5.3):

*Single-tumor mode* clusters each tumor independently. This preserves local heterogeneity patterns but pro-

duces arbitrary label IDs—"Cluster 1" in Tumor A might represent necrosis, while "Cluster 1" in Tumor B might represent highly perfused tissue. Cross-patient comparisons require post-hoc label alignment.

*Pooled mode* stacks all voxels from all tumors and clusters them in a single run. Labels are globally consistent by construction: every voxel in the cohort is assigned to one of K shared habitats. However, pooling can dilute tumor-specific patterns. If one tumor is predominantly necrotic and another highly vascular, the algorithm may find "average" habitats that represent neither tumor well. Inter-tumor variability in baseline intensities can dominate the clustering, obscuring biologically meaningful intra-tumor structure.

*Hybrid mode* combines both strategies through a two-level approach:

1. *Local clustering*: Each tumor is clustered independently (as in single-tumor mode), preserving fine-grained heterogeneity.
2. *Centroid extraction*: The centroid (mean feature vector) of each local cluster is computed, producing a compact representation of each tumor's habitat structure.
3. *Meta-clustering*: All centroids from all tumors are pooled and clustered again. The resulting cluster centers define *global prototypes*—canonical imaging phenotypes that recur across the cohort.
4. *Label assignment*: Each local cluster is mapped to its nearest global prototype, producing globally consistent habitat labels while respecting local tumor structure.

**Label harmonization.** Regardless of clustering mode, local cluster labels are arbitrary—determined by algorithm initialization, not biology. The pipeline provides an optional harmonization step that reorders labels based on a biologically meaningful reference feature (e.g., median Hounsfield units). After harmonization, Habitat 1 consistently represents the lowest-intensity phenotype across all tumors, Habitat K the highest. This can be applied to any mode.

### 5.2.3 Habitat-Derived Quantitative Metrics

Once habitats are defined, the pipeline computes metrics that summarize tumor composition and heterogeneity. For each tumor, we calculate:

- **Habitat proportions**: The fraction of tumor volume assigned to each habitat. A tumor might be 60% Habitat 1, 30% Habitat 2, and 10% Habitat 3. Differences in these proportions across patients may correlate with clinical outcomes.
- **Shannon entropy**: A single number quantifying habitat diversity, calculated as

$$H(p) = - \sum_{i=1}^K p_i \log_2 p_i$$

## DATA AND HABITAT IMAGING PIPELINE

$p_i$  is the proportion of habitat  $i$ . Entropy is maximized when habitats are equally distributed (high heterogeneity) and minimized when one habitat dominates (low heterogeneity).

The pipeline outputs clustered NIfTI files, allowing users to compute additional metrics as needed. In our studies of liver metastases (Chapters 7 and 8), we observed that tumors often exhibit a viable rim surrounding a necrotic core—a spatial pattern with potential prognostic relevance. To capture this, we extended the metrics to separate tumor subregions:

- **Rim:** The outer shell within a specified distance from the tumor boundary, computed using signed distance transforms.
- **Core:** The interior region remaining after excluding the rim.

The rim thickness is a user-defined parameter. In this thesis, we use a default of 2 mm based on biological considerations: tumor cells cannot survive more than approximately 2 mm from a blood vessel due to the diffusion limits of oxygen and nutrients [Thng, 2010]. This distance defines the maximum viable thickness of tissue without direct vascular supply, and is the rationale underlying anti-angiogenic therapies that target tumor neovascularization. For each subregion, we compute habitat proportions and entropy separately. High rim entropy, for example, may reflect an active tumor-host interface with mixed viable tumor and vascular remodeling, while low core entropy may indicate uniform necrosis. These spatially-resolved metrics are used as imaging biomarkers in the clinical analyses of Chapter 8.



## A Data-Driven Approach to CT Habitats

*There are no straight lines or sharp corners in nature.*

— Antoni Gaudí



# Chapter 6

## Identification of Precise Handcrafted Features

Robustness is a prerequisite for useful representations. If a radiomic feature's value changes meaningfully when a patient shifts position by a few millimeters, or when we adjust a minor computational parameter, then any habitat computed with that feature will be measuring noise, not biology. In this Chapter we describe a comprehensive study of precision analysis of handcrafted features, with the goal of identifying optimal features for stable habitat computation.

### Contributions:

- We assess repeatability by perturbing images to simulate a test-retest dataset collected under routine clinical conditions.
- We assess reproducibility against two key parameters for computing handcrafted radiomics: kernel radius ( $R$ , defining the neighborhood for feature calculation) and bin size ( $B$ , defining intensity discretization).
- Combining repeatability and reproducibility results, we identify a subset of precise voxelwise features for both lung and liver lesions, and test whether habitats computed with only precise features are more stable.
- We explore the biological meaning of habitats in an independent cohort of 13 liver metastases with mpMRI and histology.

This chapter reproduces many parts of the publication [Prior et al., 2024], adapted to the thesis format. A note on scope: lung lesion results are not directly relevant to the questions addressed in this thesis and therefore are not discussed in depth, they have just been kept for comparison.

## 6.1 Rationale

CT-based habitat imaging offers a non-invasive, volumetric approach to quantify tumor heterogeneity. Unlike biopsy, it samples the entire tumor. Unlike whole-region summary statistics (e.g. mean kurtosis), it preserves spatial structure. This makes habitats potentially valuable for treatment planning, response monitoring, and understanding resistance patterns—but only if the habitats themselves are stable.

For habitats to be stable, and therefore clinically useful, the underlying radiomics features (RFs) [Lambin et al., 2017] must be precise. This means that RFs should be both *repeatable* (i.e., exhibiting measurement precision under the same set of computation conditions, also known as test-retest) and *reproducible* (i.e., exhibiting measurement precision under different computation conditions) [Kessler et al., 2015, Sullivan et al., 2015].

Most existing studies of radiomics precision, however, focus on features extracted from entire ROIs and evaluate them as independent predictive biomarkers. Many such studies also fail to provide critical information such as whether the features were computed in 2D (pixelwise) or 3D (voxelwise) [Pfaehler et al., 2021, Traverso et al., 2018]. This is especially relevant for habitat computation since pixelwise features are computed slice-by-slice and ignore out-of-plane texture information, making them less representative of true 3D tumor architecture. Few studies address the precision of voxelwise 3D features specifically, and fewer still in the context of clustering-based methods like habitat imaging [Ng et al., 2013, Xu et al., 2019]. Thus, there is a lack of knowledge on precision of 3D RF for CT tumor habitat computation.

We therefore set out to answer: Which 3D radiomics features are precise enough to support stable habitat computation? We evaluated precision under three sources of variability:

1. Repeatability: Do features remain consistent when the same patient is scanned twice (test-retest)?
2. Reproducibility against kernel radius (R): Does changing the size of the neighborhood used for feature calculation alter feature values significantly?
3. Reproducibility against bin size (B): Does changing the intensity discretization design alter feature values significantly?

Features that “passed” all three tests were identified as *precise* and suitable for habitat modeling. We then asked whether using only precise features resulted in more stable habitats, using a Gaussian mixture model (GMM) clustering approach [Divine et al., 2016, Jardim-Perassi et al., 2019]. Finally, we explored the biological plausibility of the resulting CT habitats by comparing them to mpMRI and histopathology in an independent liver metastasis cohort.

## 6.2 Methods

### 6.2.1 Patient Cohorts

We analyzed 1,861 liver lesions and 575 lung lesions from 318 patients (mean age 64.5 years  $\pm$  10.1 SD; 185 male) imaged at multiple timepoints between November 2010 and December 2021 (Figure 6.1A). All patients had advanced cancer with liver metastases. Intravenous contrast-enhanced CT scans were acquired as part of routine clinical care at Vall d’Hebron University Hospital. The analysis of anonymized imaging data was approved by the Vall d’Hebron Ethics Committee with waiver of informed consent. The cohort comprised four primary tumor types: (1) colorectal, (2) lung, (3) gastrointestinal neuroendocrine tumors, and (4) a mixed group of other cancers. Patients with neuroendocrine tumors were enrolled in the multicenter phase II TALENT trial (NCT02678780). Full cohort characteristics and imaging protocols are detailed in Appendix B.

A small independent cohort of 13 patients with liver metastases, enrolled in the PREDICT prospective trial (PR[AG]29/2020), underwent CT, mpMRI, and image-guided biopsy. This cohort was used solely for exploratory biological validation. All participants provided written informed consent for acquisition and analysis of imaging and tissue samples.

### 6.2.2 Image Segmentation and Perturbation

A radiologist with more than 10 years of experience in oncologic imaging manually segmented all measurable lesions (maximal diameter  $\geq$ 10 mm per RECIST 1.1 [Eisenhauer et al., 2009]) in 3D using 3D Slicer (v4.11.20210226) [Fedorov et al., 2012].

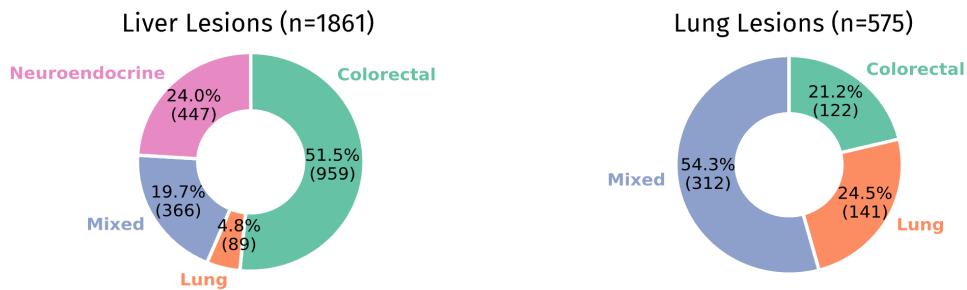
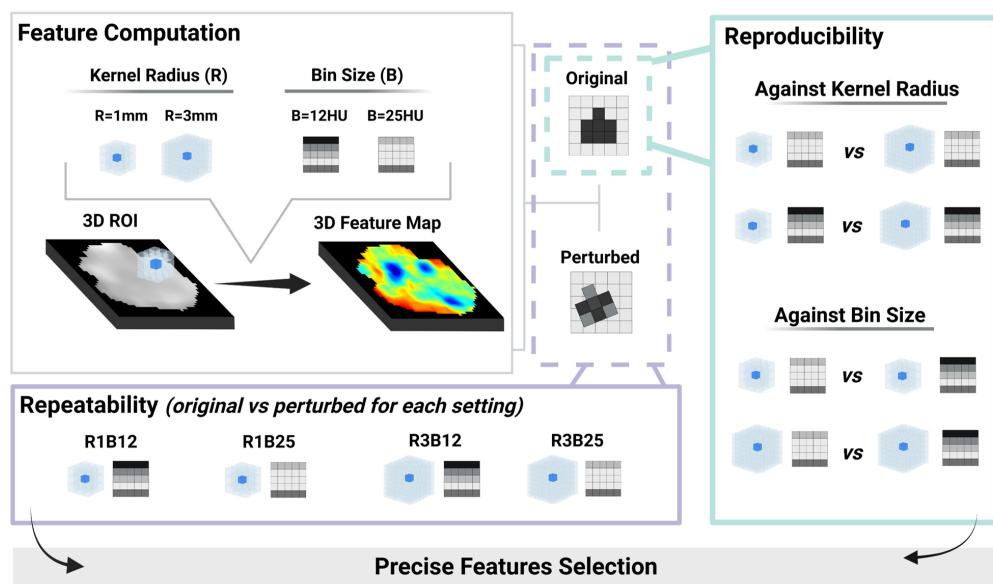
To assess repeatability, we simulated test-retest scenarios by applying controlled image perturbations, implemented using the [Medical Image Radiomics Processor \(MIRP\) toolkit](#) (v1.2.0) [Zwanenburg et al., 2019]. Details of the perturbation protocol are provided in Appendix B.

### 6.2.3 Radiomics Feature Computation

We computed 91 voxelwise radiomics features using PyRadiomics (v3.0.1) [Van Griethuysen et al., 2017]. The full feature list is provided in Appendix B. Features were calculated on original (unfiltered) images at each voxel within the segmented lesion volume. For each lesion we computed features four times, using the different combinations of the two computational parameteres we’re studying:

- **Kernel radius (R):** the size of the neighborhood used for texture computation (1 mm or 3 mm).
- **Bin size (B):** the intensity discretization used before feature calculation (12 HU or 25 HU).

The four parameter combinations—denoted R1B12, R1B25, R3B12, and R3B25—were chosen based on common practice in the radiomics literature. PyRadiomics defaults to R=1 mm and B=25 HU; we

**A****Precision Analysis Cohorts****B****Precision Analysis Design**

**Figure 6.1: Precision analysis workflow.** (A) Distribution of lung and liver lesions across different cohorts for precision analysis. (B) Precision analysis design. 3D radiomic features were computed from both original and perturbed images, four times per image, each time with a different combination of kernel radius, R (1mm/3mm), and bin size, B (12HU/25HU). To study repeatability, original-perturbed feature pairs were evaluated for every combination of extraction settings (R1B12, R1B25, R3B12 and R3B25). To study reproducibility against extraction parameters, we compared original feature pairs extracted with different extraction settings: for reproducibility against R, we first compared features extracted with fixed B=12HU and different kernel radius, and then repeated for features extracted with fixed B=25HU; analogously, for reproducibility against B, we compared features extracted from original images with fixed R=1mm and different bin size, and then repeated with those extracted with fixed R=3mm. Precise features were selected by linking reproducibility and repeatability results. R, kernel size; B, bin size.

selected R=3 mm and B=12 HU as alternative values frequently reported in prior studies. All computations followed the Image Biomarker Standardization Initiative (IBSI) guidelines [Zwanenburg et al., 2020]. Computation details are provided in Appendix B.

### 6.2.4 Precision Analysis: Repeatability and Reproducibility

Figure 6.1B summarizes the precision analysis workflow. We evaluated:

- **Repeatability** (test-retest stability): We compared feature values from original and perturbed images for each of the four parameter settings (R1B12, R1B25, R3B12, R3B25).
- **Reproducibility against kernel radius (R)**: We compared features computed with R=1 mm versus R=3 mm, holding bin size constant. Two experiments were conducted: one at B=12 HU (R1B12 vs. R3B12) and one at B=25 HU (R1B25 vs. R3B25).
- **Reproducibility against bin size (B)**: We compared features computed with B=12 HU versus B=25 HU, holding kernel radius constant. Two experiments were conducted: one at R=1 mm (R1B12 vs. R1B25) and one at R=3 mm (R3B12 vs. R3B25).

All experiments were performed on the full dataset, on liver and lung lesions separately, and stratified by primary tumor type. This allowed us to assess whether precision varied by lesion location or primary cancer.

We quantified precision using the ICC [Koo and Li, 2016]. For repeatability, we used a single-measurement, absolute-agreement, two-way mixed-effects model. For reproducibility, we used a single-measurement, consistency, two-way mixed-effects model. Following Koo and Li, we classified features based on the LCL of the ICC: poor ( $LCL < 0.50$ ), moderate ( $0.50 < LCL < 0.75$ ), good ( $0.75 < LCL < 0.90$ ), and excellent ( $LCL \geq 0.90$ ). A feature was selected as precise if the 95% LCL of the ICC was  $\geq 0.50$  across the three relevant experiments: repeatability (R3B12), reproducibility against R (B=12HU), and reproducibility against B (R=3mm).

### 6.2.5 Habitat Computation and Stability Assessment

Habitats were computed using Gaussian mixture models (GMMs). We clustered voxels based on their radiomics feature values, treating each lesion independently. To avoid redundancy, we first removed highly correlated features (Spearman's  $r \geq 0.70$ ,  $p < 0.001$ ) [Schwarz, 1978]. The optimal number of clusters (habitats) for each lesion was determined using the Bayesian Information Criterion (BIC) [Cohen, 1992]. We computed habitats four times per lesion:

- Using all 91 features, on original images
- Using all 91 features, on perturbed images
- Using only precise features, on original images

- Using only precise features, on perturbed images

Habitat stability was quantified by comparing the spatial overlap of habitats derived from original versus perturbed images, using the DSC. Higher DSC indicates greater stability. We tested the hypothesis that habitats derived from precise features would be more stable than those derived from all features. Details of the GMM implementation and BIC optimization are provided in Appendix B.

### 6.2.6 Exploratory Biological Case Study

To assess whether CT habitats capture biologically meaningful tissue compartments rather than imaging artifacts, we conducted an exploratory analysis in a subset of 13 patients from the PREDICT cohort (see Section 5.1) who had co-acquired CT, multiparametric MRI (mpMRI) and digitized hematoxylin-eosin (HE) histopathology from biopsy.

CT habitats were computed using the precise liver features identified in Section 6.2.4, clustered with GMM and BIC as described above. For comparison, mpMRI habitats were computed by voxelwise clustering of nine quantitative maps—warped non-linearly onto a high-resolution T2-weighted anatomical scan—including tissue T2 ( $T_{2t}$ ), longitudinal relaxation time ( $T_1$ ), tissue and vascular apparent diffusion coefficients ( $ADC_t$ ,  $ADC_v$ ), tissue apparent kurtosis ( $K_t$ ), vascular fraction ( $f_v$ ), capillary permeability constant ( $K^{trans}$ ), extracellular extravascular volume fraction ( $v_e$ ), and plasma volume fraction ( $v_p$ ). We then compared the number of habitats selected by BIC for both CT and mpMRI habitats in each lesion and the qualitative correspondence between imaging habitats and tissue phenotypes identified by a pathologist on HE slides.

This analysis was deliberately exploratory. Given the sampling bias, our goal was not to establish definitive biological ground truth, but to confirm that CT habitats have a plausible relationship to known tissue phenotypes such as necrosis and viable tumor.

### 6.2.7 Statistical Analysis

Differences in feature reproducibility across parameter settings, lesion locations, and habitat stability were assessed using paired two-sided Wilcoxon signed-rank tests. Effect sizes were calculated using Cohen’s d and classified as small ( $d \geq 0.20$ ), medium ( $d \geq 0.50$ ), or large ( $d \geq 0.80$ ) [Cohen, 1992]. Statistical significance was defined as  $p < 0.05$ . All statistical analyses were reviewed by a statistician and performed in Python (v3.7.10). Code for reproducibility and repeatability analysis is publicly available at <https://github.com/radiomicsgroup/precise-habitats>.

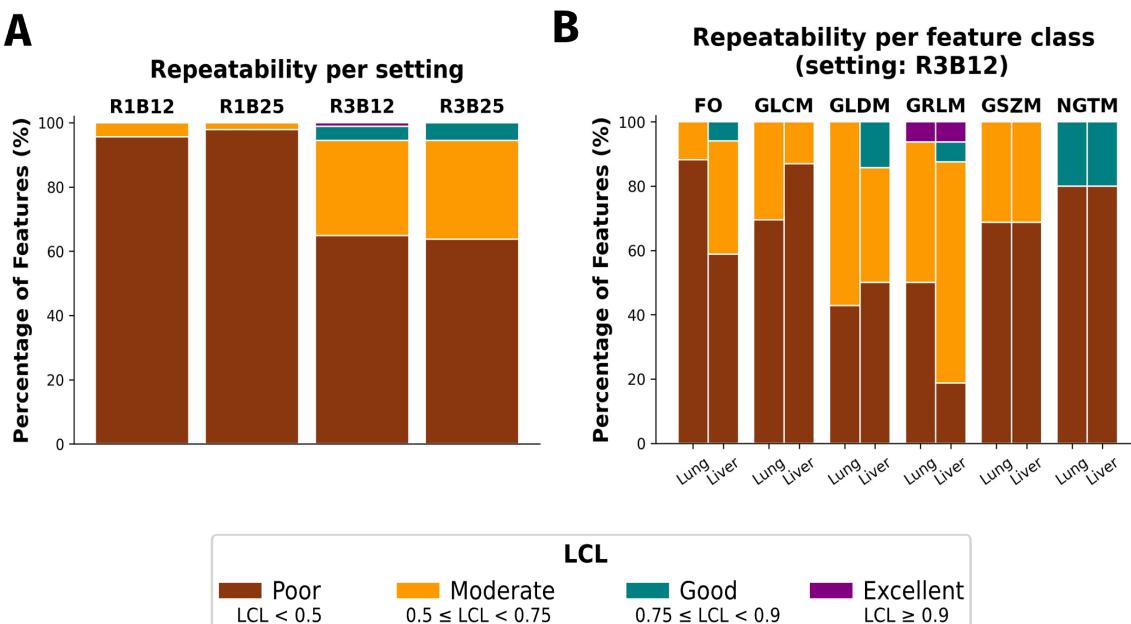
## 6.3 Results

### 6.3.1 Precision Analysis

**Repeatability.** Figure 6.2 shows the distribution of repeatability across the four parameter settings. Most features exhibited poor repeatability ( $\text{ICC LCL} < 0.50$ ) regardless of settings, but clear differences emerged between parameter choices.

Features computed with a kernel radius of 3 mm were substantially more repeatable than those computed with 1 mm, regardless of bin size. The R3B12 setting yielded the highest repeatability, with a median ICC LCL of 0.442 (IQR: 0.312–0.516) across all features, compared to 0.191 (0.116–0.382) for R1B12, 0.199 (0.103–0.344) for R1B25, and 0.415 (0.306–0.516) for R3B25. Bin size had minimal impact: changing B from 12 HU to 25 HU did not meaningfully alter repeatability for either kernel radius.

The type of repeatable features varied by lesion location (Figure 6.2B). In liver lesions, first-order and Gray-Level Run-Length Matrix (GLRLM) features were more repeatable. Primary tumor type (colorectal, lung, neuroendocrine, other) had no detectable effect on repeatability (Appendix B).



**Figure 6.2: Repeatability of voxelwise radiomics features.** (A) Repeatability distribution of radiomics features per setting. Most radiomic features exhibit poor repeatability. Features extracted with kernel radius (R) of 3mm were more repeatable than those extracted with R=1mm. Bin size changes didn't affect repeatability. (B) Repeatability distribution of radiomics features extracted with setting R3B12 per feature class for lung and liver lesions separately. First order and GLRLM features were more repeatable in liver lesions while GLCM features were more repeatable in lung lesions. LCL, 95% lower confidence limit of the ICC; R3B12, features extracted with kernel radius 3mm and bin size 12HU; FO, First-Order; GLCM, Grey Level Co-occurrence Matrix features; GLDM, Grey Level Dependence Matrix; GLRLM, Grey Level Run Length Matrix; GSZM, Grey Level Size Zone Matrix; NGTDM, Neighboring Grey Tone Difference Matrix Features.

**Reproducibility.** Features were far more sensitive to changes in kernel radius than to changes in bin size (Figure 6.3A-B). Changing R from 1 mm to 3 mm substantially altered feature values: median ICC LCL

## IDENTIFICATION OF PRECISE HANDCRAFTED FEATURES

was only 0.440 (IQR: 0.330–0.526) when  $B=12$  HU and 0.437 (0.355–0.524) when  $B=25$  HU. In contrast, changing  $B$  from 12 HU to 25 HU had little effect: median ICC LCL was 0.929 (0.853–0.988) when  $R=3$  mm and 0.833 (0.706–0.946) when  $R=1$  mm.

**Table 6.1: Reproducibility against kernel radius and bin size.** Median (IQR) lower confidence limit of the ICC reported.  $B$  = bin size, HU = Hounsfield units,  $R$  = kernel radius.

	Reproducibility against R				Reproducibility against B			
	Fixed B = 12 HU		Fixed B = 25 HU		Fixed R = 1 mm		Fixed R = 3 mm	
	Liver	Lung	Liver	Lung	Liver	Lung	Liver	Lung
LCL								
Median (IQR)	0.422 (0.346-0.513)	0.573 (0.403-0.701)	0.407 (0.291-0.536)	0.573 (0.443-0.696)	0.805 (0.672-0.919)	0.929 (0.823-0.997)	0.921 (0.821-0.982)	0.967 (0.93-0.999)

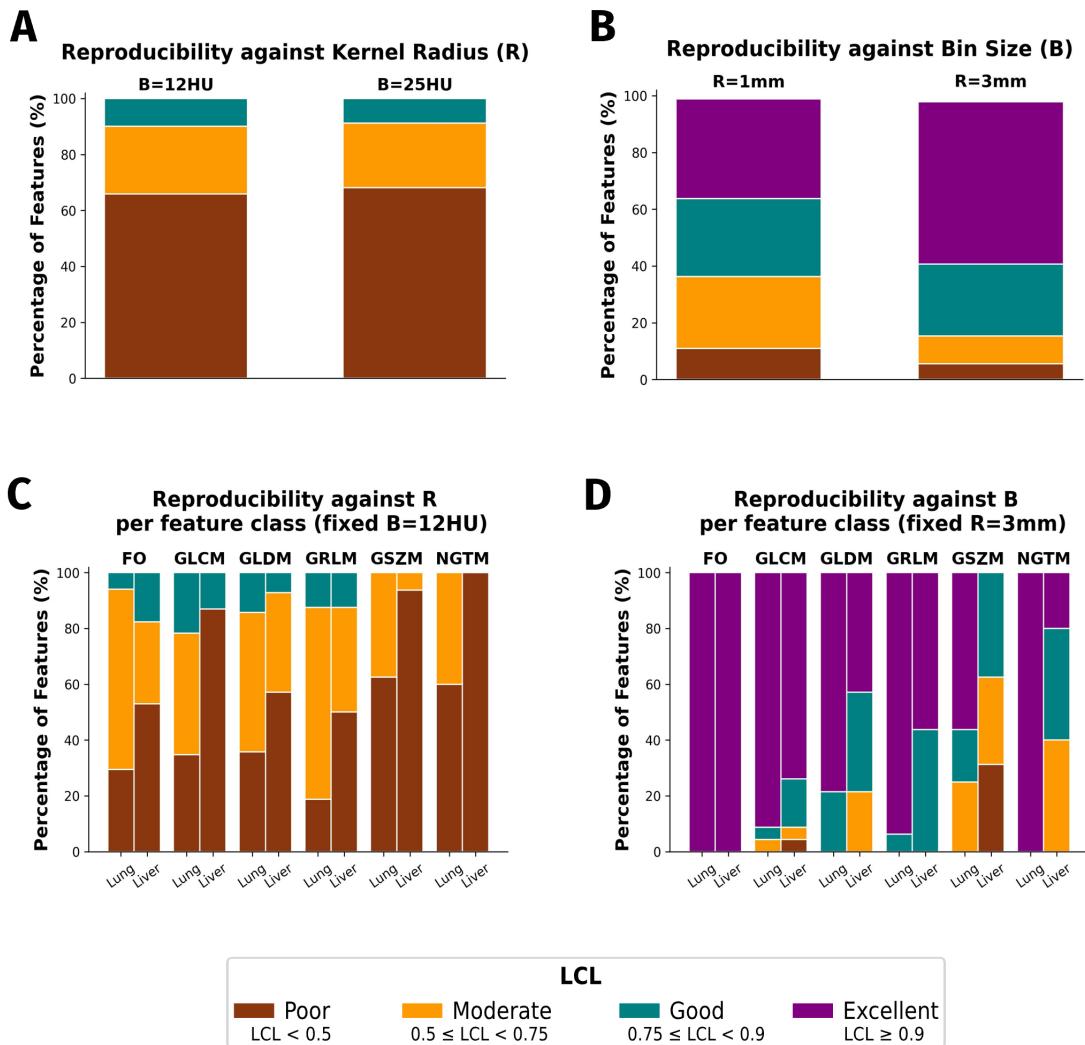
In other words, most features were reproducible against bin size (good or excellent agreement), but poorly reproducible against kernel radius. The choice of neighborhood size matters more than the choice of discretization bin. Features computed with  $B=12$  HU were slightly more reproducible against kernel radius than those with  $B=25$  HU ( $p <0.001$ ). Features computed with  $R=3$  mm were more reproducible against bin size than those with  $R=1$  mm ( $p <0.001$ ).

As with repeatability, lesion location affected reproducibility (Figure 6.3C-D, Table 6.1). Features from lung lesions were more reproducible than those from liver lesions, both against R and against B ( $p <0.001$ ), particularly for GLCM and GLRLM feature classes. Primary tumor type again had no detectable effect (Appendix B).

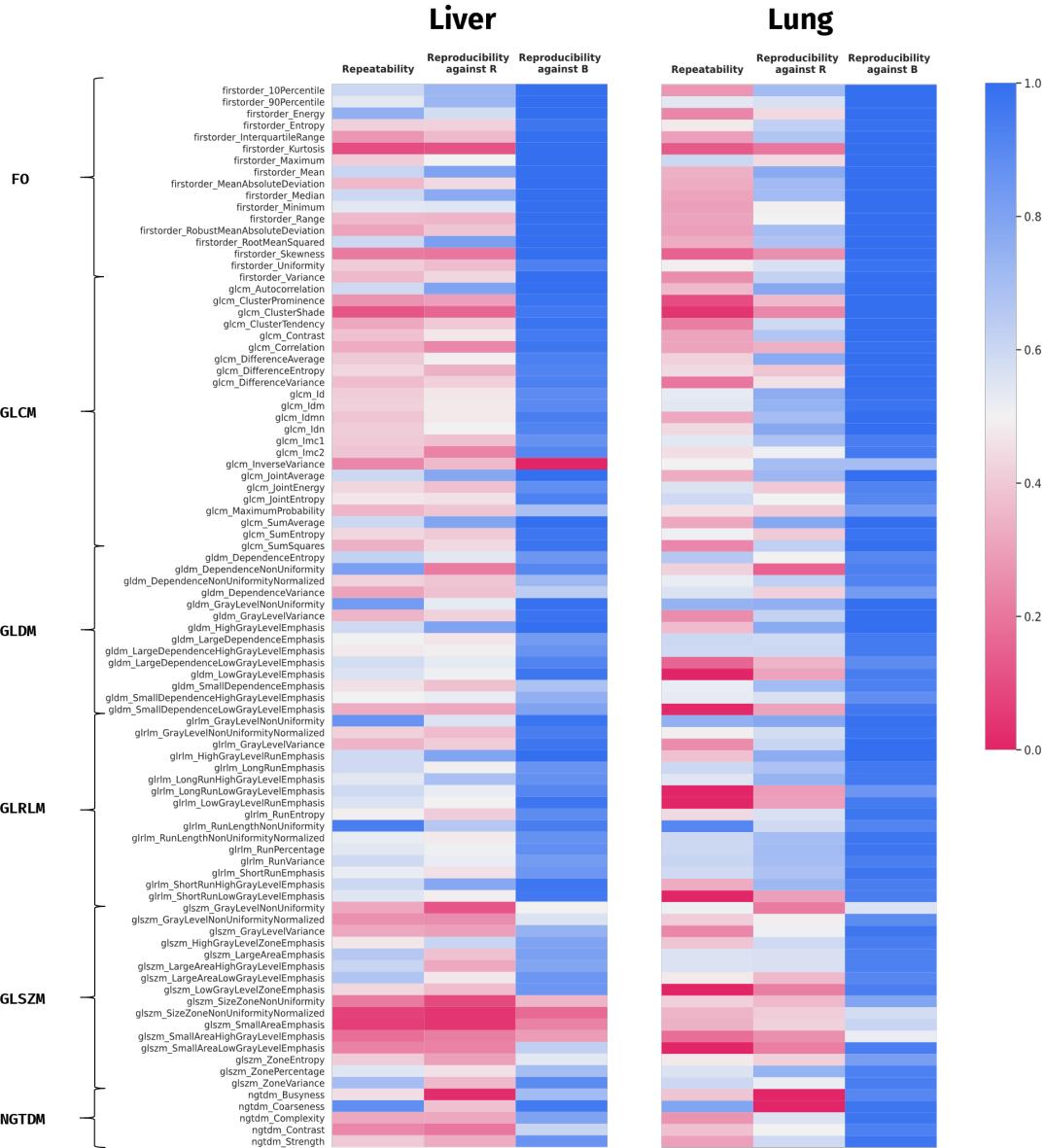
### 6.3.2 Identification of Precise Features

We defined a feature as precise if it met three criteria: ICC LCL  $\geq 0.50$  for (1) repeatability (R3B12 setting), (2) reproducibility against kernel radius (at  $B=12$  HU), and (3) reproducibility against bin size (at  $R=3$  mm). These thresholds represent the minimum acceptable stability for clinical use.

Of the 91 features tested, 25 met all three criteria for liver lesions. A 26th feature—NGTDM Coarseness—was added based on its consistently high performance across experiments (top 3 most repeatable and reproducible against B; see Appendix B for justification). The final set thus comprised 26 precise features for liver lesions. The same analysis for lung lesions also identified 26 precise features (Section 6.3.2). Figure 6.4 displays the ICC LCL values for all 91 features across the three experiments, for liver and lung lesions separately.



**Figure 6.3: Reproducibility against kernel radius and bin size.** (A) Reproducibility distribution against kernel radius, R, for features extracted with fixed bin size of 12HU and bin size 25HU. Most features present poor reproducibility against R. Features extracted with B=12HU were more reproducible ( $p<.001$ ). (B) Reproducibility distribution against bin size, B, for features extracted with fixed kernel radius 3mm and fixed kernel radius 1mm. Most features present good or excellent reproducibility against B. Features extracted with R=3mm were more reproducible ( $p<.001$ ). (C) Reproducibility distribution against kernel radius for features extracted with fixed bin size of 12HU per feature class for lung and liver lesions separately. Features extracted from lung lesions are more reproducible against R, especially for features belonging to GLCM and GRLM classes. (D) Reproducibility distribution against bin size for features extracted with fixed kernel radius 3mm per feature class for lung and liver lesions separately. Features extracted from lung lesions are more reproducible against B, especially for features belonging to GLCM, first-order and NGTDM classes. LCL, 95% lower confidence limit of the ICC; FO, First-Order features; GLCM, Grey Level Co-occurrence Matrix features; GLDM, Grey Level Dependence Matrix; GRLM, Grey Level Run Length Matrix; GSZM, Grey Level Size Zone Matrix; NGTDM, Neighboring Grey Tone Difference Matrix Features.



**Figure 6.4: Summary of Precision Analysis.** Heatmap displaying the lower 95% confidence limit (LCL) of the ICC results obtained in the three experiments used to identify precise features: repeatability (setting R3B12), reproducibility against R (fixed B=12HU), and reproducibility against B (fixed R=3mm), for lung and liver lesions separately. FO, First-Order features; GLCM, Grey Level Co-occurrence Matrix features; GLDM, Grey Level Dependence Matrix; GLRLM, Grey Level Run Length Matrix; GLSzm, Grey Level Size Zone Matrix; NGTDM, Neighboring Grey Tone Difference Matrix Features.

**Table 6.2: Precise 3D Radiomics Features in Liver and Lung Lesions.** A precise radiomic feature was defined as lower confidence limit  $\geq 0.50$ . FO = first-order; GLCM = Grey Level Co-occurrence Matrix; GLDM = Grey Level Dependence Matrix; GLRLM = Grey Level Run Length Matrix; GLSZM = Grey Level Size Zone Matrix; NGTDM = Neighboring Grey Tone Difference Matrix.

Liver lesions	Lung lesions
FO_10Percentile	FO_90Percentile
FO_90Percentile	GLCM_Id
FO_Energy	GLCM_Idm
FO_Mean	GLCM_Imc1
FO_Median	GLCM_InverseVariance
FO_Minimum	GLCM_JointEntropy
FO_RootMeanSquared	GLDM_DependenceEntropy
GLCM_Autocorrelation	GLDM_DependenceNonUniformityNorm.
GLCM_JointAverage	GLDM_GrayLevelNonUniformity
GLCM_SumAverage	GLDM_LargeDependenceEmphasis
GLDM_DependenceEntropy	GLDM_LargeDependenceHighGrayLevelEmphasis
GLDM_GrayLevelNonUniformity	GLDM_SmallDependenceEmphasis
GLDM_HighGrayLevelEmphasis	GLDM_SmallDependenceHighGrayLevelEmphasis
GLDM_LargeDependenceLowGrayLevelEmphasis	GLRLM_GrayLevelNonUniformity
GLDM_LowGrayLevelEmphasis	GLRLM_LongRunEmphasis
GLDM_SmallDependenceHighGrayLevelEmphasis	GLRLM_LongRunHighGrayLevelEmphasis
GLRLM_GrayLevelNonUniformity	GLRLM_RunLengthNonUniformity
GLRLM_HighGrayLevelRunEmphasis	GLRLM_RunLengthNonUniformityNorm.
GLRLM_LongRunHighGrayLevelEmphasis	GLRLM_RunPercentage
GLRLM_LongRunLowGrayLevelEmphasis	GLRLM_RunVariance
GLRLM_LowGrayLevelRunEmphasis	GLRLM_ShortRunEmphasis
GLRLM_RunLengthNonUniformity	GLSZM_LargeAreaEmphasis
GLRLM_RunPercentage	GLSZM_LargeAreaHighGrayLevelEmphasis
GLRLM_RunVariance	GLSZM_ZonePercentage
GLRLM_ShortRunHighGrayLevelEmphasis	GLSZM_ZoneVariance
NGTDM_Coarseness	NGTDM_Coarseness

### 6.3.3 Imaging Habitats Computed with Precise Features Show Increased Stability

Having identified which features are precise, we asked whether using only these features improves habitat stability. For each lesion, we computed habitats twice: once using all 91 features, and once using only the 26 precise features. In both cases, highly correlated features (Spearman's  $r \geq 0.70$ ) were removed before clustering. Habitats were generated using Gaussian mixture models (GMMs), with the number of

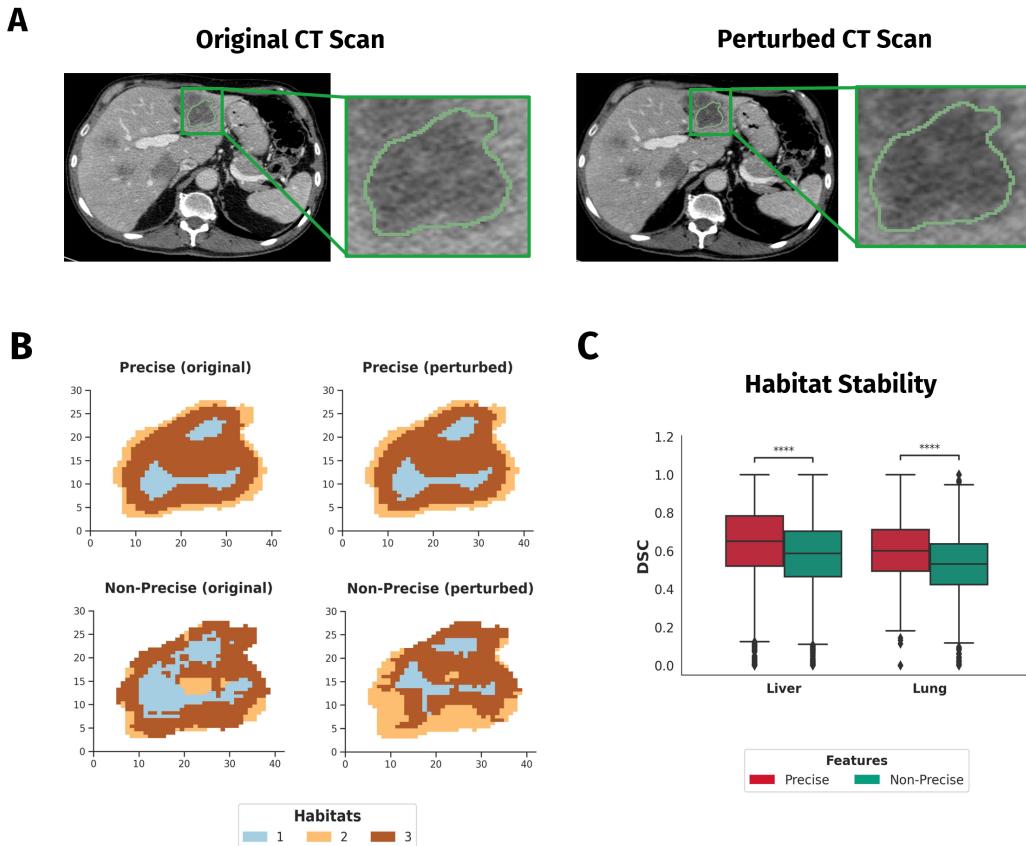
## IDENTIFICATION OF PRECISE HANDCRAFTED FEATURES

clusters determined by the Bayesian Information Criterion (BIC). We then compared the spatial overlap of habitats derived from original versus perturbed images using the Dice Similarity Coefficient (DSC). Higher DSC indicates greater stability.

Habitats computed from precise features were significantly more stable than those computed from all features ( $p < 0.001$ , Wilcoxon signed-rank test; Figure 6.5C). This held for both liver and lung lesions, with small-to-medium effect sizes (Cohen's  $d = 0.34$  for lung,  $0.43$  for liver).

For liver lesions, the median DSC was  $0.587$  (IQR:  $0.465$ – $0.703$ ) when using all features, compared to  $0.651$  ( $0.520$ – $0.784$ ) when using only precise features. For lung lesions, median DSC was  $0.532$  ( $0.424$ – $0.637$ ) with all features versus  $0.601$  ( $0.494$ – $0.712$ ) with precise features.

Figure 6.5B illustrates this difference for a single liver lesion. Habitats derived from precise features maintained their spatial structure across test-retest (DSC:  $0.976$ ,  $0.891$ ,  $0.915$  for the three habitats), whereas habitats derived from all features exhibited substantial variation (DSC:  $0.751$ ,  $0.328$ ,  $0.570$ ).

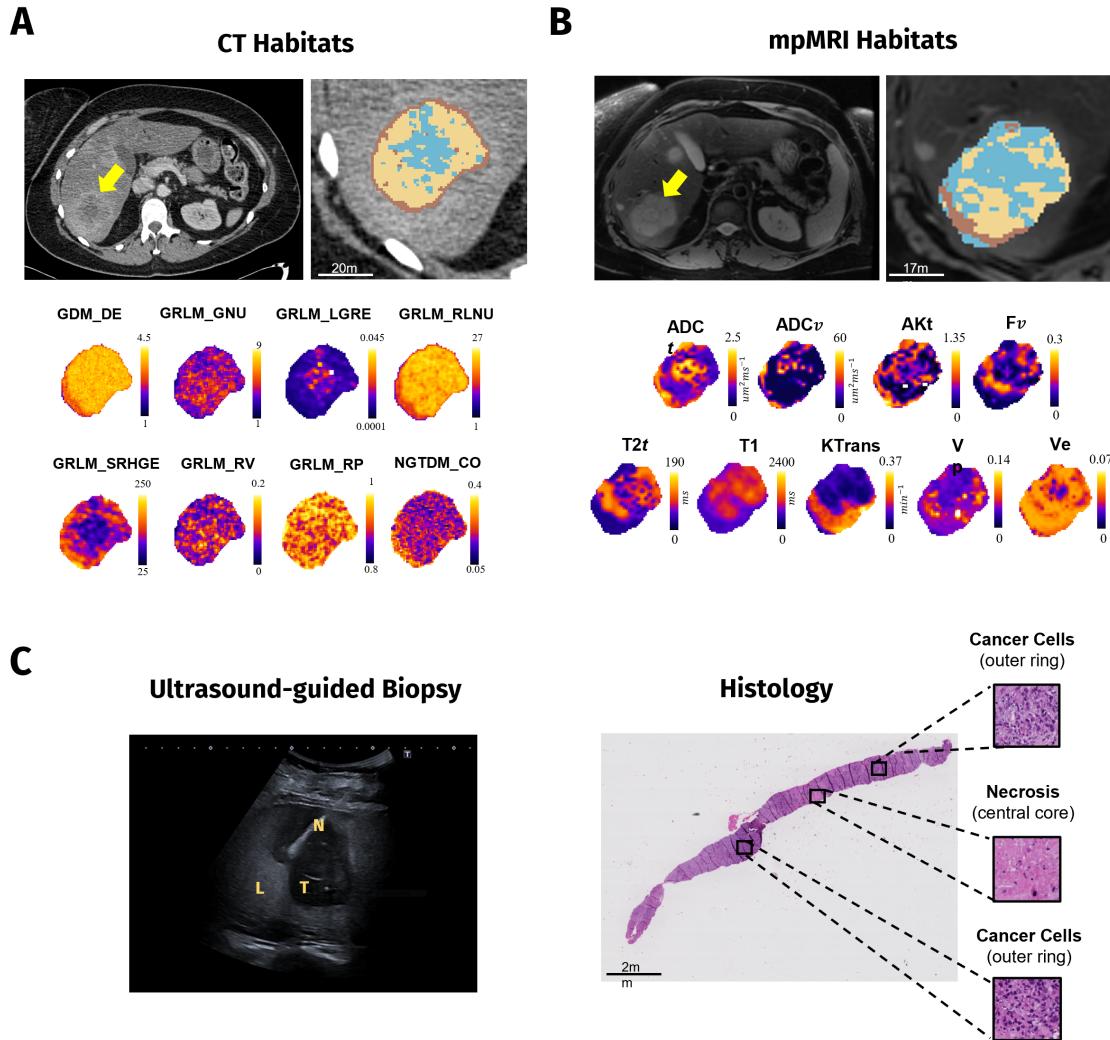


**Figure 6.5: Precise features increase habitat stability.** (A) Original and perturbed CT scans for one liver lesion. (B) Example of habitats obtained for the same lesion. Habitats computed with precise features show higher stability (measured via DSC of original-perturbed habitat pairs). Top row: habitats obtained with precise features computed from original image (left) and perturbed image (right). DSC scores for habitats 1, 2 and 3 are  $0.976$ ,  $0.891$  and  $0.915$ , respectively. Bottom row: habitats obtained with non-precise (i.e., all computed features) features computed from original image (left) and perturbed image (right). DSC scores for habitats 1, 2 and 3 are  $0.751$ ,  $0.328$  and  $0.570$ , respectively. (C) Quantification of habitat stability computed with precise features and non-precise features for all lung and liver lesions. Habitats computed with precise features present higher stability (Wilcoxon signed rank test,  $p < 0.0001$ ). DSC: Dice Similarity Coefficient.

### 6.3.4 Exploratory Biological Case Study

As a preliminary check of biological plausibility, we examined CT habitats in 13 PREDICT patients with multiparametric MRI (mpMRI) and histopathology (biopsy, not whole tumor resection).

Figure 6.6 shows an example from a patient with liver metastasis from melanoma.



**Figure 6.6: Exploration of the biological relevance of imaging habitats.** One exemplificatory patient (liver metastasis of melanoma). (A) CT scan with visible lesion (yellow arrow) and resulting CT habitats computed with precise liver radiomic features (also shown). (B) Anatomical MRI T2 scan with visible lesion (yellow arrow) and resulting mpMRI habitats computed with the 9 mpMRI maps (also shown). mpMRI and CT habitats presented comparable distributions. (C) Image guided biopsy with needle (N) and liver (L) tumor lesion (T), and resulting histologic image (H&E stain) with observable tissue types, annotated by a pathologist. The H&E-stained histological material reveals areas of necrosis in the core of the lesion.  $ADC_t$ : tissue apparent diffusion coefficient,  $ADC_v$ : vascular apparent diffusion coefficient,  $T_{2t}$ : tissue transverse relaxation time,  $K_t$ : tissue apparent kurtosis coefficient,  $f_v$ : vascular signal fraction,  $T_1$ : total longitudinal relaxation time,  $K^{trans}$ : capillary permeability constant,  $v_p$ : plasma volume fraction,  $v_e$ : extravascular and extracellular volume fraction.

CT habitats (computed from precise features) displayed spatial distributions qualitatively consistent with

habitats derived from nine mpMRI maps (Figure 6.6A–B). For example, a CT habitat depicted in blue—spatially concordant with an mpMRI habitat in the lesion core—showed elevated  $T_{2t}$ ,  $T_1$ , and  $ADC_t$  (consistent with lower cellularity) and reduced  $f_v$ ,  $K^{trans}$ ,  $v_e$ , and  $v_p$  (consistent with lower vascularization). This pattern is compatible with necrosis. Histopathology confirmed a necrotic core surrounded by viable tumor (Figure 6.6C), matching the spatial organization suggested by both imaging modalities.

This exploratory analysis relied on visual comparison rather than voxel-wise co-registration or quantitative spatial metrics. Its purpose was not to validate the biological meaning of habitats definitively, but to rule out the possibility that they reflect pure imaging noise or reconstruction artifacts. The qualitative agreement between CT habitats, mpMRI habitats, and pathologist-identified tissue types (necrosis, viable tumor) suggests that CT habitats encode biologically relevant information about tumor microenvironment heterogeneity.

## 6.4 Discussion

To achieve an effective clinical translation of CT habitats, excellent robustness of the underlying imaging features is necessary. In this chapter we examined 3D radiomics’ repeatability in a simulated test-retest scenario and reproducibility against kernel radius (R) and bin size (B), two relevant computation parameters.

Of the 91 features we evaluated, only 26 met our criteria for acceptable repeatability and reproducibility in liver lesions. These precise features were not a random subset—they exhibited consistent patterns. Texture features generally outperformed first-order (histogram) features, likely because histogram features depend on absolute intensity values and are more vulnerable to outliers. Features computed with a larger kernel radius (R=3 mm) were more repeatable than those with a smaller radius (R=1 mm), suggesting that averaging over a larger neighborhood reduces sensitivity to noise. Features were far more sensitive to changes in kernel radius than to changes in bin size, meaning the choice of neighborhood matters more than the choice of discretization.

Using only precise features produced more stable habitats. When we clustered voxels based on all 91 features, habitats shifted substantially between test and retest (median DSC: 0.587 for liver lesions). When we used only the 26 precise features, habitat boundaries remained far more consistent (median DSC: 0.651). The improvement was statistically significant ( $p < 0.001$ ) and, while modest in absolute terms, represents a meaningful gain in reliability. Without ground truth, we cannot say which habitat map is “correct,” but we can say which is more robust—and robustness is a prerequisite for clinical utility.

The set of precise features differed between liver and lung lesions. This was true both in number and in type: GLCM features were more precise in lung, while first-order and GLRLM features were more precise in liver. One plausible explanation is the difference in contrast-to-noise ratio between the two tissue types. Lung lesions, surrounded by air, exhibit high contrast on CT; liver lesions, embedded in soft tissue, do not. This suggests that radiomics models are not universally portable—features that work in one anatomical site may not work in another. Primary tumor type (colorectal, lung, neuroendocrine, other) had no detectable effect on precision, implying that lesion location matters more than tumor origin.

Our exploratory case study attempted to explore the biological relevance of CT habitats, inspired by previous studies that highlighted the value of quantitative MRI-derived habitats for characterizing tumor heterogeneity [Divine et al., 2016, Jardim-Perassi et al., 2019]. In the 13 patients examined, CT habitats showed qualitative agreement with habitats derived from nine mpMRI maps and with tissue features visible on histology (e.g., necrotic cores, viable tumor rims). Though conclusions are difficult to draw in view of our limited sample size, CT and mpMRI habitats may be capturing biologically relevant imaging phenotypes, potentially serving as non-invasive markers of cancer aggressiveness. This underscores the potential clinical utility of our approach, still in an exploratory context.

To our knowledge, this is the first study to systematically evaluate the precision of 3D voxelwise radiomics features against both kernel radius and bin size in liver and lung lesions. Direct comparison with prior work is therefore limited. Previous studies reported higher proportions of repeatable features [Berenguer et al., 2018, Bernatowicz et al., 2021, Jha et al., 2021, Mottola et al., 2021], which could be attributed to differences in the number of RF, perturbation methods [Jha et al., 2021, Mottola et al., 2021] and the use of phantoms instead of real patients [Berenguer et al., 2018]

Our finding that texture features are more precise than first-order features contrasts with some earlier reports [Traverso et al., 2018]. This discrepancy may reflect our focus on voxelwise computation: histogram features, which summarize intensity distributions, are inherently less local and may be more stable when computed over large regions but less stable when computed voxel-by-voxel.

The high sensitivity of features to kernel radius has received little attention in the literature, yet it has important implications. Many radiomics studies do not report the kernel radius used, or report it inconsistently. If a feature's value changes substantially when  $R$  is varied from 1 mm to 3 mm—a seemingly minor adjustment—then comparing results across studies becomes difficult. Our results underscore the need for standardized reporting of all computational parameters, not just acquisition protocols.

The generalizability of our results is subject to several key limitations. First, we evaluated only original (unfiltered) features. Convolutional filters—such as wavelet and Laplacian of Gaussian filters—are commonly used in radiomics and have been shown to improve predictive performance in some contexts [Demircioğlu, 2022]. Whether filtered features are more or less precise than original features, and whether they improve habitat stability, remains unknown. Filter standardization is still under development, and extending our precision framework to filtered features is a natural next step.

Second, all segmentations were performed by a single radiologist. While this ensures consistency, it does not capture variability introduced by inter-observer differences in lesion delineation. Features depend on contours, and contours vary across readers. Assessing the impact of segmentation variability on habitat stability is important future work, but it was beyond the scope of this chapter. In addition, we did not evaluate reproducibility across different scanners. Scanner variability is a well-studied problem in radiomics and has been addressed extensively elsewhere. Our focus was on sources of variability that arise even when imaging is performed on the same scanner under identical protocols—namely, patient repositioning and computational parameter choices. These sources of variability are less explored and, we argue, equally important.

Fourth, our precision analysis was limited to CT. Whether the same features are precise in MRI or PET,

and whether habitats derived from different modalities overlap spatially, are open questions. The exploratory mpMRI analysis in this chapter suggests qualitative agreement, but quantitative validation requires dedicated studies with co-registered multimodal imaging and ground truth from spatially matched biopsies.

Finally, our habitat stability analysis relied on voxel-by-voxel comparison (Dice coefficient). This may underestimate global similarity: two habitat maps with slightly shifted boundaries but similar overall structure would receive a low DSC despite being qualitatively similar. Alternative metrics—such as mutual information or topology-based measures—might capture stability more holistically. The modest absolute DSC values we observed (median ~0.65 for precise features) likely reflect this limitation, along with genuine variability in how habitats are defined when clustering is applied to noisy features.

## 6.5 Summary

Our comprehensive repeatability and reproducibility analysis identified two subsets of precise RF for effectively computing stable CT tumor habitats in lung and liver lesions. By employing these precise RF and using unsupervised clustering models, we demonstrated the ability to identify distinct tumor phenotypes in an exploratory case study. CT tumor habitats correlated with biologically meaningful tumor aspects such as cellularity, vascularization, and necrosis, but further studies with larger sample sizes are needed to validate these findings. This approach to computing CT habitats holds great potential for studying intra-tumoral heterogeneity and monitoring cancer evolution throughout the course of the disease.

### Key Points:

- Voxelwise 3D radiomics features showed poor repeatability (median ICC LCL: 0.442) and reproducibility against kernel radius (ICC LCL: 0.440), but excellent reproducibility against bin size (ICC LCL: 0.929).
- Of 91 features, 26 were classified as precise for liver lesions. The set of precise features differed between liver and lung lesions.
- Habitats computed from precise features were 11% more stable (median DSC: 0.651 vs. 0.587,  $p < 0.001$ ) than those computed from all features.
- Exploratory validation with mpMRI and histopathology suggested that CT habitats capture biologically relevant tissue compartments (e.g., necrosis, viable tumor, vascularization differences).

# Chapter 7

## Development and Validation of an mpMRI-anchored CT Habitat Model

In Chapter 6, we identified a set of precise handcrafted radiomics features for stable CT habitat computation. An exploratory analysis suggested qualitative agreement between these habitats and histologic phenotypes such as necrosis and viable cancer. In this chapter, we address the central question arising from that observation: can CT habitats capture the tissue phenotypes that dominate colorectal liver metastases?

The key methodological strength is the use of the PREDICT dataset (see Section 5.1.). Each CT voxel is paired with aligned mpMRI maps reflecting cellularity and vascularity. This allows us to evaluate CT habitats against pseudo-biological ground truth during development—not just post-hoc.

### Contributions:

- We compare four CT data representations for habitat computation (raw HU, handcrafted radiomics features, deep learning features from a liver tumor segmentation model, and deep learning features from a foundation model) and identify the representation that yields the most biologically coherent habitats.
- We develop an mpMRI-anchored framework for CT habitat discovery, using co-registered multi-parametric MRI as a surrogate for tissue cellularity and vascularity.
- We validate CT habitats against whole-tumor histopathology annotations of necrosis, fibrosis, and viable cancer in resected samples.

The work presented in Chapters 7 and 8 forms the basis of a manuscript currently in preparation.

## 7.1 Rationale

Colorectal cancer is the third most common malignancy worldwide, and the liver is its most frequent site of metastasis [Arnold et al., 2017, Siegel et al., 2020]. Approximately half of patients with colorectal cancer will develop liver metastases during their disease course. While surgical resection offers the best chance for long-term survival, the majority of patients present with unresectable disease and receive systemic therapy (i.e. chemotherapy and/or targeted therapy). Response to these therapies is highly variable, and a significant proportion of patients develop resistance [Zeineddine et al., 2023].

Tumor heterogeneity is increasingly recognized as a driver of this resistance. Colorectal liver metastases are not uniform masses but histologically complex ecosystems containing variable proportions of viable tumor, necrosis, and fibrosis. These compartments carry prognostic significance: fibrosis often indicates favorable response to chemotherapy [Poultsides et al., 2012], while central necrosis may reflect treatment failure due to poor drug penetration [Wong and Neville, 2007]. If we could distinguish these tissue states non-invasively, we might stratify patients more accurately and guide treatment decisions earlier.

Habitat imaging offers a potential solution. By clustering voxels into spatially distinct subregions, habitats can identify tissue compartments that correspond to different biological states, potentially including necrosis, fibrosis and viable cancer. In Chapter 6, we demonstrated that stable CT habitats can be computed using a precise subset of handcrafted radiomics features, and we observed qualitative correspondence with histologic phenotypes in an exploratory analysis. Two questions remained open: Can CT habitats reliably capture these three dominant tissue phenotypes? And if so, are handcrafted features the optimal CT representation for this task?

This second question arises from the fact that deep learning has transformed medical image analysis over the past decade. Radiology foundation models have become recently available and encode rich hierarchical information after being trained on extremely large datasets. In conventional radiomics studies (i.e. bulk radiomics, not voxelwise studies) several studies have compared handcrafted and deep learning features for outcome prediction. For habitat imaging, however, no study has compared these representations. Thus, to date, we do not know whether the representational capacity of deep learning translated into more biologically coherent habitats, or if simpler handcrafted features are enough.

Answering these questions requires a reference standard that captures tissue biology at the voxel level. Histopathology is the gold standard for tissue characterization, but voxel-to-pixel alignment between radiology and pathology remains technically challenging. Tissue deforms during resection and fixation; the scale mismatch between radiology slices ( $\sim 1\text{--}2$  mm) and histology sections ( $\sim 4\text{--}5$   $\mu\text{m}$ ) complicates spatial correspondence; and registration errors accumulate.

Multiparametric MRI offers a middle ground between histology's cellular resolution and CT's accessibility. Diffusion-weighted imaging (DWI), the primary non-invasive MRI technique for probing tissue microstructure [Le Bihan et al., 2001], measures water diffusion at scales far below the millimeter-level image resolution. In tumors, ADC values inversely correlate with cellularity: dense cellular packing restricts water motion, yielding low ADC [Chen et al., 2013, Surov et al., 2017]. Advanced diffusion models can also estimate microvascular volume fraction ( $f_v$ ), a marker of tissue vascularity [Federau et al., 2014,

Togao et al., 2018]. Similarly, DCE MRI quantifies vascular properties through pharmacokinetic modeling. Parameters such as the transfer constant ( $K^{trans}$ ) have been validated as markers of tumor vascularity and perfusion [Jackson et al., 2007, Tofts and Kermode, 1991]. Together, these modalities provide voxel-level maps of biologically meaningful tissue states, cellularity and vascularity, that are impractical to obtain from histology alone. If CT habitats reflect true biology, they should naturally differentiate tissue according to these same underlying properties.

The PREDICT cohort provides a unique resource since patients underwent both contrast-enhanced CT and multiparametric MRI. This enables something beyond post-hoc validation; it allows us to select the CT representation whose habitats best align with mpMRI-derived biophysical metrics. Rather than computing CT habitats and hoping they reflect biology, we can ask directly: which CT features produce clusters that separate tissue with different cellularity and vascularity? The resulting model is not merely validated against mpMRI; it is anchored to it. mpMRI serves as both the selection criterion and the biological reference.

We followed a two-stage strategy. First, in the PREDICT cohort, we compared four CT representations and selected the one whose habitats connected measures (cellularity  $ADC$ ; vascularity  $f_v, K^{trans}$ ). Second, we applied the selected model to the POEM cohort, patients with resected colorectal liver metastases and whole-tumor histopathology, to confirm that habitats exhibit spatial organization consistent with necrotic cores, fibrotic regions, and viable tumor rims.

Our hypothesis was that CT habitats can capture the three dominant histological phenotypes of colorectal liver metastases: necrosis, fibrosis, and viable tumor. We tested this by identifying the optimal CT representation, characterizing the biological profile of each habitat using mpMRI, and validating the spatial correspondence against whole-tumor histopathology.

## 7.2 Methods

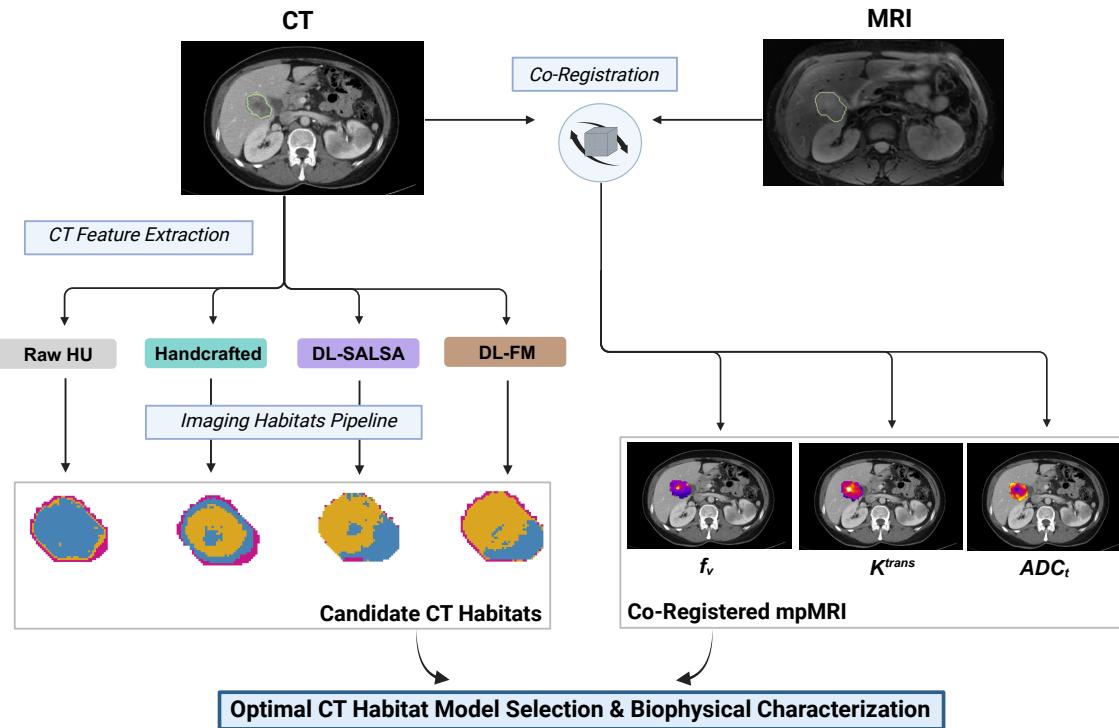
### 7.2.1 Study Design and Patient Cohorts

This study employed two cohorts to develop, characterize, and validate the CT habitat model (Figure 7.1).

**1. CT Habitat Model Development and Characterization (PREDICT).** The PREDICT cohort (Section 5.1.2) comprises patients with colorectal liver metastases who underwent both contrast-enhanced CT and multiparametric MRI before treatment. CT was acquired in the portal venous phase; mpMRI included T2-weighted, diffusion-weighted, and dynamic contrast-enhanced sequences. Quantitative mpMRI maps, including  $ADC_t$ ,  $K^{trans}$ , and  $f_v$ , were derived as described in Section 5.1.2. For the present analysis, we included 10 patients (42 tumors) with evaluable imaging on both modalities and successful CT–mpMRI co-registration.

**2. Histopathological Validation (POEM).** The POEM cohort (Section 5.1.3) was used for independent validation against biological ground truth. This cohort includes 6 patients with

colorectal liver metastases who underwent surgical resection. Whole-tumor specimens were processed for histology and annotated for necrosis, fibrosis, and viable tumor, providing a macroscopic spatial reference for validating habitat architecture.



**Figure 7.1: Schematic of the mpMRI-anchored CT habitat model development.** CT and MRI scans from the PREDICT cohort were co-registered to T2w space. Four voxelwise CT feature representations (Raw HU, Handcrafted Radiomics, DL-SALSA, and DL-FM) were extracted and processed through the imaging habitats pipeline to generate candidate habitat maps. Co-registered mpMRI maps (including vascular fraction,  $f_v$ ; capillary permeability,  $K^{trans}$ ; and tissue ADC,  $ADC_t$ ) provided biophysical reference for model selection. The optimal CT habitat model was selected based on its ability to spatially separate tissue properties reflecting cellularity and vascularity (with the three mpMRI maps shown:  $ADC_t$ ,  $K^{trans}$  and  $f_v$ ), and further characterized using additional mpMRI maps. Created with BioRender.com.

## 7.2.2 CT-mpMRI Co-Registration

To enable voxelwise comparison between CT and mpMRI, all images were spatially co-registered to the T2-weighted image, which served as the fixed reference. This required three registration pipelines: CT→T2w, DWI→T2w, and GRE→T2w. We followed

Images were cropped to a tumor-centered bounding box (5–7 mm margin) and resampled to 2×2×2 mm isotropic resolution prior to registration. Registration was performed using NiftyReg in sequential stages: rigid, affine, then deformable B-spline if alignment improved. Quality was assessed by DSC between the T2w tumor mask and the warped mask; the transformation yielding the highest DSC was retained. Full methodological details are provided in Appendix C.

### 7.2.3 CT Feature Extraction

Four CT representations were evaluated:

- **Raw HU.** The original portal-venous phase CT images, serving as baseline.
- **Handcrafted radiomics features.** Voxelwise texture features computed using PyRadiomics, based on the 26 precise features identified in Chapter 6 (see Section 6.3.2). To avoid clustering on redundant information, we removed highly correlated features (Spearman  $|\rho| \geq 0.80$ ), retaining 6 non-redundant features: 10th percentile intensity, GLDM dependence entropy, GLDM small dependence high gray level emphasis, GLRLM gray level non-uniformity, GLRLM run length non-uniformity, and NGTDM coarseness. Correlation analysis is detailed in Appendix C.
- **Deep learning features from a liver tumor segmentation model (DL-SALSA).** Activations from the penultimate decoder layer of SALSA, a 3D nnU-Net trained for liver tumor segmentation [Balaguer-Montero et al., 2025].
- **Deep learning features from a foundation model (DL-FM).** Activations from a SegResEncoder foundation model pretrained on diverse CT datasets using self-supervised learning [Pai et al., 2025].

All CT images were resampled to  $1 \times 1 \times 1$  mm isotropic resolution prior to feature extraction. For biological evaluation, the resulting habitat maps were warped to the T2w reference space ( $2 \times 2 \times 2$  mm) using nearest-neighbor interpolation.

### 7.2.4 CT Habitat Model Development

#### Clustering Configuration

Habitats were computed using the imaging habitats pipeline described in Section 5.4, with standard pre-processing: features with high right-skewness ( $>1.0$ ) were log-transformed; all features were then standardized to zero mean and unit variance. Clustering followed the two-stage hybrid approach: local GMM clustering within each tumor, followed by meta-clustering of local centroids to define global prototypes. The number of habitats was set to K=3, based on the hypothesis that CT-derived clusters could correspond to the dominant histological compartments of colorectal liver metastases: viable tumor, necrosis, and fibrosis. The fitted scaler and global prototypes were saved, enabling application to new data without retraining.

#### CT Habitat Model Selection

To select the optimal CT representation, we evaluated how well each feature set produced habitats that separate *biologically* distinct tissue, using the co-registered mpMRI maps as reference. Specifically, we selected the following three:

- $ADC_t$  (tissue apparent diffusion coefficient): a proxy for cellularity

- $f_v$  (vascular signal fraction): a measure of blood volume
- $K^{trans}$  (volume transfer constant): a measure of capillary permeability

Both  $f_v$  and  $K^{trans}$  were included to capture different aspects of tumor vascularity—structural vascular density  $f_v$  versus functional leakiness and perfusion  $K^{trans}$ .

For each representation, we computed habitats and extracted the median value of each mpMRI metric within each habitat for each patient. We aggregated to patient level (N=10) to avoid pseudo-replication. Habitat separation was assessed using the Friedman test, with Kendall's W as effect size [Kendall and Smith, 1939, Tomczak and Tomczak, 2014]. To summarize performance across the three metrics, we computed a biophysical separation score defined as the mean Kendall's W across  $ADC_t$ ,  $f_v$ , and  $K^{trans}$ . The representation with the highest biophysical separation score was selected for further validation.

## Technical Validation

Strong biophysical separation could be spurious if the clustering is unstable. We therefore evaluated:

- **Initialization stability.** Mean pairwise Adjusted Rand Index across 5 random seeds. High ARI indicates the solution is reproducible regardless of initialization.
- **Data stability.** Mean ARI across 30 bootstrap iterations, comparing each resampled model to the full-cohort reference. High ARI indicates that habitat definitions are not driven by a few patients.
- **Spatial coherence.** Moran's I (spatial autocorrelation) and surface-area-to-volume ratio. Biologically plausible habitats should form contiguous regions, not scattered noise.

## Habitat Characterization with mpMRI

Having selected the optimal CT representation based on biophysical separation, we characterized the resulting habitats using the full panel of mpMRI metrics available in the PREDICT cohort.

Our initial hypothesis was that CT habitats would correspond to histologically defined compartments: necrosis, fibrosis, and viable tumor. The mpMRI characterization served to test this hypothesis and, if the correspondence proved incomplete, to build an alternative interpretive framework based on the tissue properties the habitats actually captured.

For each of the 13 mpMRI maps (Table 5.2), we extracted the median value within each habitat for each patient, following the same aggregation approach used for model selection. Differences across the three habitats were assessed using the Friedman test, with Kendall's W as effect size. Post-hoc pairwise comparisons used Wilcoxon signed-rank tests with Benjamini-Hochberg [Benjamini and Hochberg, 1995] correction for multiple comparisons within each metric.

### 7.2.5 Histopathological Validation

To assess whether CT habitats correspond to histologically defined tissue compartments, we applied the trained model to the POEM cohort (Section 5.1.3). Whole-tumor H&E sections were digitized and an-

notated for viable tumor, necrosis, and fibrosis using a supervised pixel classifier in QuPath [Bankhead et al., 2017], trained on regions delineated by the author of this thesis, based on an experienced pathologist’s guidance. Automated annotations were reviewed by the pathologist for quality assurance. We then performed two experiments:

- **Qualitative spatial correspondence.** Direct voxel-to-voxel co-registration between CT and histology was not feasible. We therefore evaluated correspondence qualitatively, comparing the spatial distribution of CT habitats with the architecture visible on annotated histology sections.
- **Correlation with histological tissue percentages.** As an exploratory analysis, we computed Spearman correlations between whole-tumor habitat proportions and whole-tumor tissue percentages.

The primary goal was not to establish quantitative agreement, which would require spatial co-registration and larger samples, but to confirm that CT habitats exhibit spatial organization consistent with known histological architecture: viable tumor rims surrounding necrotic or fibrotic cores.

## 7.3 Results

### 7.3.1 Handcrafted Radiomics Yields the Most Biologically Coherent CT Habitats

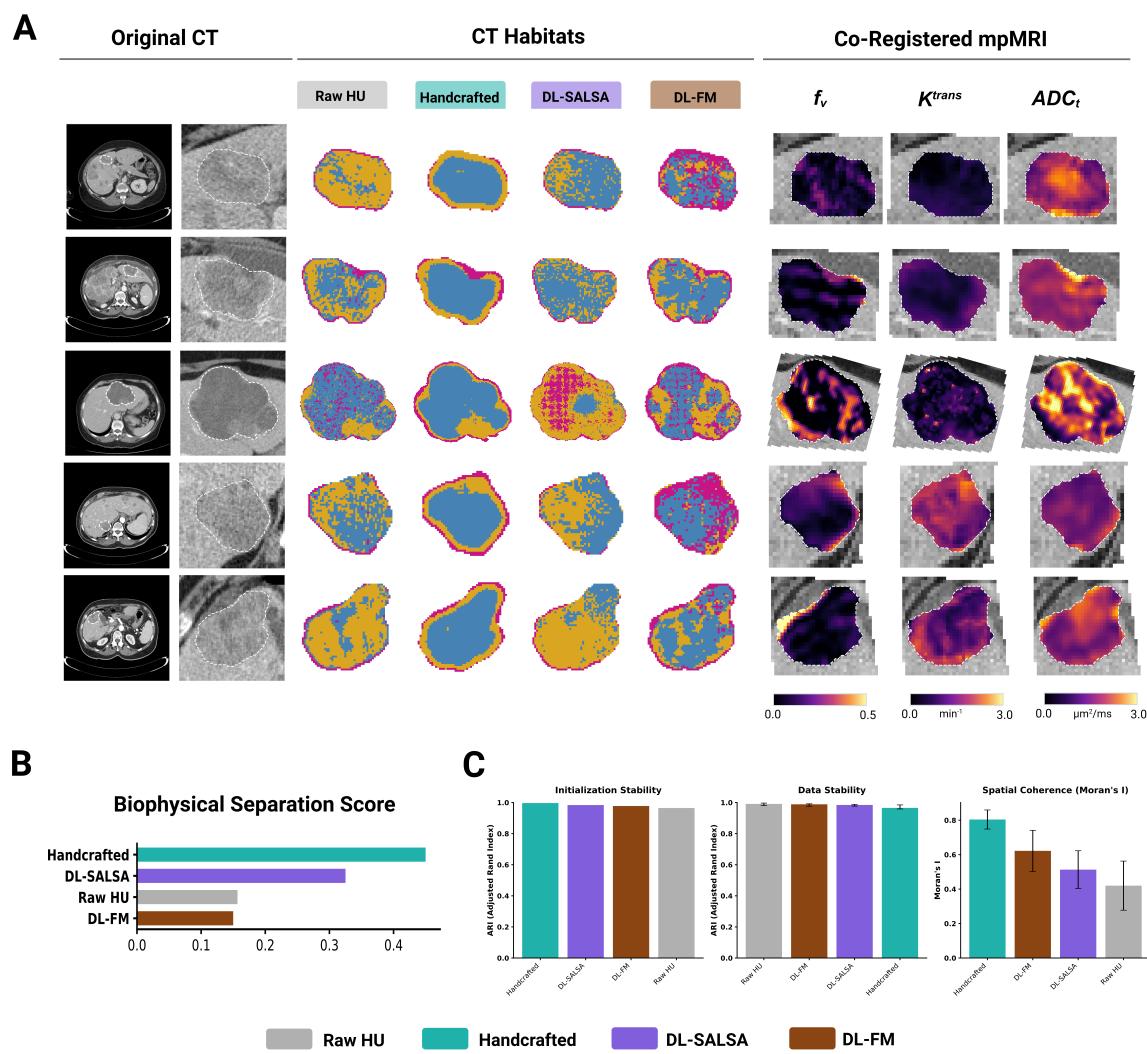
**Co-registration quality.** We first assessed the quality of the three registration pipelines required for voxelwise CT–mpMRI comparison. The majority of registrations used rigid transformations; no deformable registration was required. Across 42 tumors, the DWI-to-T2w registration achieved the highest alignment (median DSC = 0.79, IQR: 0.71–0.84), followed by GRE-to-T2w (median DSC = 0.76, IQR: 0.66–0.85) and CT-to-T2w (median DSC = 0.70, IQR: 0.65–0.78).

**CT Habitat Model Selection.** We compared the four CT representations by their ability to produce habitats that separate the three pre-specified mpMRI metrics: tissue ADC ( $ADC_t$ ), vascular fraction ( $f_v$ ), and perfusion ( $K^{trans}$ ). Figure 7.2 shows examples of CT habitats with the four feature sets. Table 7.1 summarizes the results. Handcrafted radiomics achieved the highest biophysical separation score (mean  $W = 0.45$ ), driven by strong effects for vascular fraction ( $W = 0.67$ ,  $p = 0.005$ ) and perfusion ( $W = 0.52$ ,  $p = 0.018$ ). Raw HU showed moderate separation for cellularity ( $W = 0.31$ ) but failed to distinguish vascular phenotypes. DL-SALSA performed moderately on vascular metrics; DL-FM showed no significant separation on any metric. None of the representations achieved strong separation on  $ADC_t$ , our primary cellularity proxy.

To assess whether the three-habitat solution was optimal for the selected representation, we performed post-hoc sensitivity analyses with K=2 and K=4. The K=3 solution provided the best balance between biological interpretability and cluster stability; K=2 merged biologically distinct vascular phenotypes, while K=4 introduced a habitat with unstable membership across bootstrap iterations. Results are detailed in Appendix C.

**Table 7.1: Comparison of candidate CT feature representations.** \* $p < 0.05$  after Bonferroni correction. W = Kendall's W effect size.

	$ADC_t$		Vascular fraction ( $f_v$ )		$K^{trans}$		Biophysical Score
	W	P-value	W	P-value	W	P-value	
Raw HU	0.31	0.2928	0.12	0.4894	0.04	0.6703	0.1567
Handcrafted	0.16	0.3281	0.67	0.0053*	0.52	0.0055*	0.4500
DL-SALSA	0.1481	0.3808	0.4815	0.0853	0.3457	0.0446*	0.3251
DL-FM	0.13	0.4429	0.19	0.3889	0.13	0.2725	0.1500



**Figure 7.2: Selection of the optimal CT representation.** (A) Representative habitat maps for a single tumor generated by the four candidate representations (Raw HU, Handcrafted Radiomics, DL-SALSA, DL-FM), alongside reference mpMRI maps ( $ADC_t$ ,  $f_v$ ,  $K^{trans}$ ). (B) Biophysical separation performance (Kendall's W) for each representation across the three pre-specified mpMRI metrics. Handcrafted radiomics achieved the highest separation on vascular metrics ( $f_v$ ,  $K^{trans}$ ) but not on cellularity ( $ADC_t$ ). (C) Technical stability metrics: initialization stability (ARI across seeds), data stability (bootstrap ARI), and spatial coherence (Moran's I).

**Technical validation.** All representations demonstrated high initialization stability ( $ARI > 0.96$ ), indicating that the clustering solution was reproducible regardless of random seed. However, spatial coherence differed substantially. Handcrafted features produced spatially contiguous habitats (Moran's  $I \approx 0.80$ ), while DL-SALSA and DL-FM produced fragmented, "salt-and-pepper" patterns with low spatial auto-correlation. Full technical validation results are reported in Appendix C.

Based on these findings, handcrafted radiomics was selected as the final CT habitat model.

### 7.3.2 CT Habitats Capture Distinct Vascular and Cellular Phenotypes

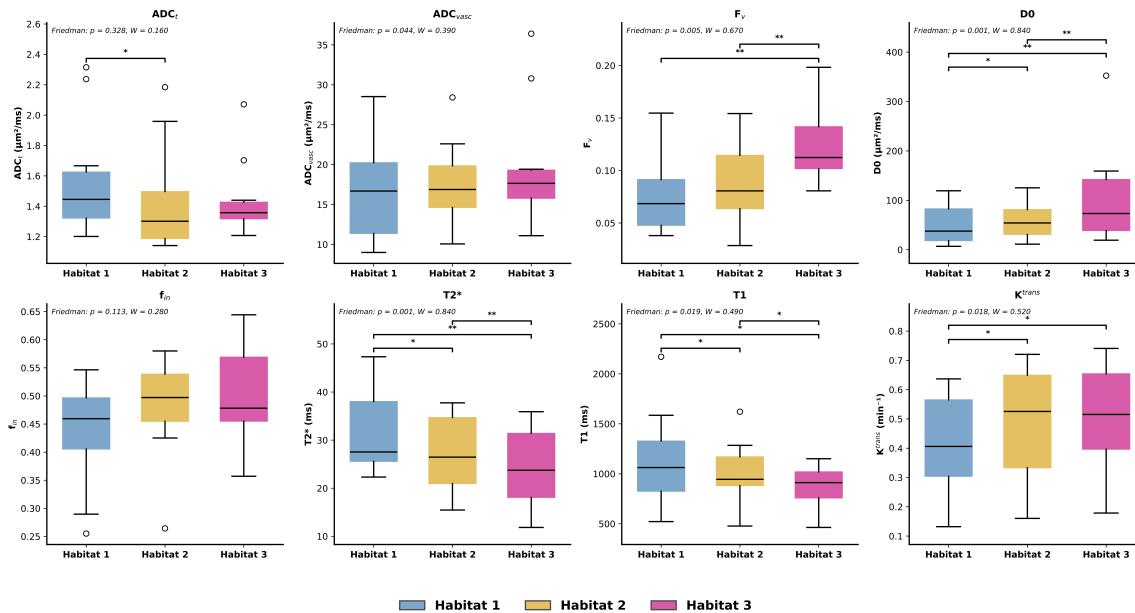
We characterized the three habitats using 13 mpMRI metrics to test whether they corresponded to the hypothesized histological compartments (necrosis, fibrosis, and viable tumor) (Table 7.2). Figure 7.3 shows boxplots for the eight metrics with significant differences across habitats, plus two cellularity-related metrics ( $f_{in}$ ,  $ADC_t$ ). Full results with pairwise significance tests are available in Appendix C.

Vascular parameters showed consistent gradients across habitats. Vascular fraction ( $f_v$ ) increased monotonically from H1 to H3 ( $0.068 \rightarrow 0.112$ ;  $W = 0.67$ ,  $p = 0.005$ ), as did vascular ADC. Relaxation times ( $T_2^*$ ,  $T_1$ ) decreased in parallel, consistent with increased blood content and contrast uptake. However, capillary permeability ( $K^{trans}$ ) did not follow this gradient; it was highest in H2 ( $0.52 \text{ min}^{-1}$ ), not H3 ( $0.51 \text{ min}^{-1}$ ), with both elevated relative to H1 ( $W = 0.52$ ,  $p = 0.018$ ). If H3 represented angiogenic tumor tissue, we would expect it to show the highest  $K^{trans}$ , since tumor neovessels are typically leaky. The combination of high  $f_v$  but moderate  $K^{trans}$  in H3 suggests partial volume effects at the tumor-liver interface, where voxels mix tumor tissue with adjacent normal liver parenchyma, which has mature, less permeable vasculature.

Cellular parameters showed a different pattern. Metrics related to cellularity did not follow a simple gradient; instead, H2 emerged as the most cellular habitat. Tissue ADC was lowest in H2 ( $1.30 \mu\text{m}^2/\text{ms}$ ; pairwise  $p = 0.029$  vs H1), indicating more restricted diffusion. Intracellular fraction ( $f_{in}$ ), cell density (CD), and cell size (vCS) pointed in the same direction: H2 contains the most densely packed cells. H1, by contrast, showed low-moderate cellularity consistent with necrotic or fibrotic tissue; H3 showed intermediate values.

**Table 7.2: CT Habitat Characterization with mpMRI metrics.** Patient-level median values for 13 mpMRI-derived biophysical parameters across the three CT habitats. W = Kendall's W. \* $p < 0.05$ , \*\* $p < 0.01$ .

mpMRI metric	Units	Habitats (medians)			p-value (BH)	Effect Size (W)	
		H1	H2	H3			
$ADC_t$	Tissue ADC	$\mu\text{m}^2/\text{ms}$	1.45	1.30	1.36	0.328	0.16
$ADC_v$	Vascular ADC	$\mu\text{m}^2/\text{ms}$	16.6	16.8	17.6	0.044*	0.39
$K_t$	Tissue kurtosis excess	Dim.less	0.75	0.69	0.70	0.587	0.07
$f_v$	Vasc. signal fraction	Norm.	0.068	0.081	0.112	0.005**	0.67
$T_{2t}$	Tissue $T_2$	ms	81.4	100.8	82.9	0.726	0.04
$D_0$	Intrinsic diffusivity	$\mu\text{m}^2/\text{ms}$	37.5	54.1	71.6	0.001**	0.84
$vCS$	Vol-weighted cell size	$\mu\text{m}$	24.0	23.6	23.9	0.529	0.09
$f_{in}$	Intracellular fraction	Norm.	0.46	0.50	0.48	0.113	0.28
$CD$	Cell density	$\times 10^3 \text{ Cells/mm}^3$	51.2	56.8	53.6	0.529	0.09
$T_1$	$T_1$	ms	1063	945	911	0.019*	0.49
$T_2^*$	$T_2^*$	ms	27.5	26.5	23.7	0.001**	0.84
$K^{trans}$	Capillary permeability	$\text{min}^{-1}$	0.41	0.52	0.51	0.018*	0.52
$v_e$	EES volume	Norm.	0.66	0.73	0.74	0.741	0.03



**Figure 7.3: Biophysical characterization of CT habitats using mpMRI.** Distribution of eight mpMRI metrics across the three CT habitats (H1: blue, H2: yellow, H3: pink). Vascular parameters ( $f_v$ ,  $K^{trans}$ ,  $ADC_v$ ,  $D_0$ ) show gradients from H1 to H3. Cellular parameters ( $ADC_t$ ,  $f_{in}$ ) show H2 as the most cellular habitat. Boxplots show patient-level medians (N = 10). Brackets indicate significant pairwise comparisons (Wilcoxon, BH-corrected): \* $p < 0.05$ , \*\* $p < 0.01$ .

Based on these profiles, we propose the following biological interpretation:

- **Habitat 1: Avascular.** Low vascular fraction, low permeability, low-moderate cellularity. Likely contains necrosis, fibrosis, or both; CT cannot distinguish between them.
- **Habitat 2: Cellular-Perfused.** Intermediate vascular fraction, highest capillary permeability, highest cellularity. Consistent with actively proliferating tumor tissue: densely packed viable cells with leaky neovessels.
- **Habitat 3: Vascular.** Highest vascular fraction, moderate permeability, moderate cellularity. May capture the tumor-liver interface with partial volume from normal liver parenchyma, or a vascular-dominant tumor compartment.

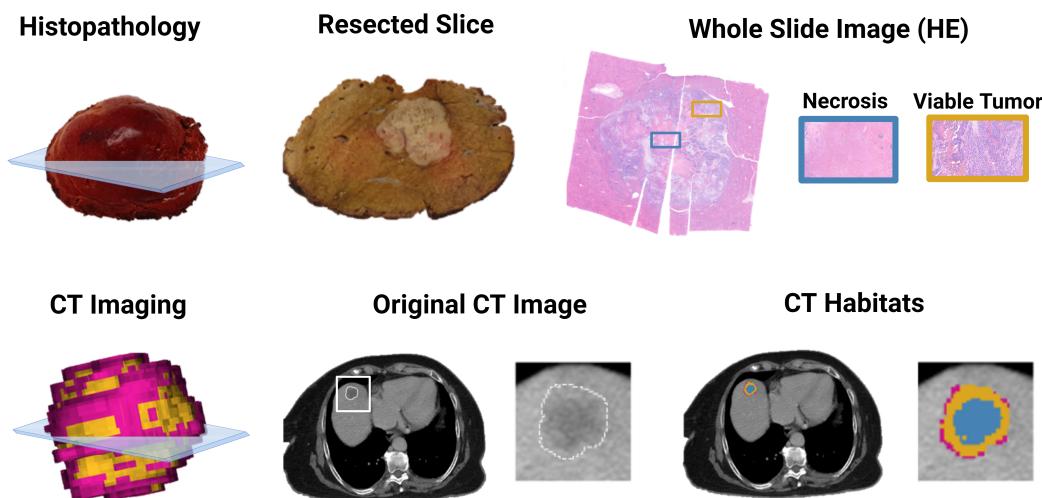
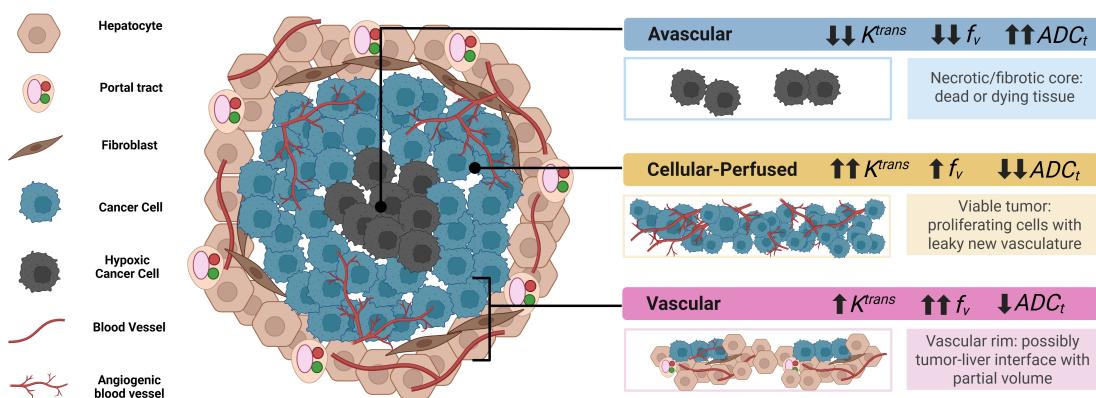
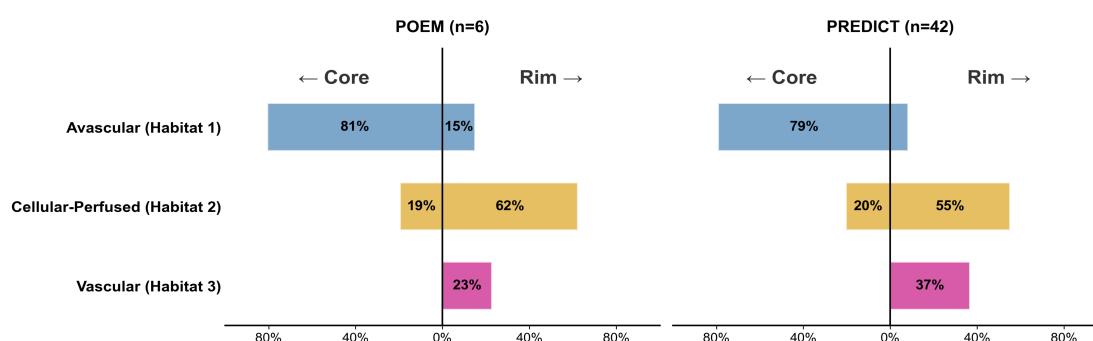
CT habitats thus reflect two aspects of tumor heterogeneity: a vascular gradient ( $H1 \rightarrow H3$ ) and a cellularity peak ( $H2$ ). The model distinguishes viable, proliferative tumor ( $H2$ ) from both the avascular core ( $H1$ ) and the vascular interface zone ( $H3$ ).

### 7.3.3 Spatial Architecture and Histopathological Validation

**Qualitative spatial correspondence.** We applied the trained CT habitat model to the POEM cohort ( $n = 6$ ) to assess correspondence with whole-tumor histology. Resected specimens displayed the architecture characteristic of colorectal liver metastases: central regions of necrosis and fibrosis surrounded by viable tumor at the periphery (Figure 7.4A). CT habitat maps captured this organization;  $H1$  (Avascular) localized to tumor cores, coinciding mostly with histologically annotated necrosis, while  $H2$  (Cellular-Perfused) and  $H3$  (Vascular) predominated at the periphery, corresponding to viable tumor margins (Figure 7.4B).

To quantify this spatial organization, we computed habitat proportions separately for the tumor rim (2mm outer shell) and core (interior) across both cohorts ( $N = 48$  tumors: 6 POEM, 42 PREDICT; Figure 7.4C).  $H1$  dominated the core ( $79.4\% \pm 2.9\%$ ) but was largely absent from the rim ( $9.0\% \pm 2.3\%$ ).  $H3$  was concentrated in the rim ( $35.0\% \pm 1.5\%$ ) and virtually absent from the core ( $0.4\% \pm 0.2\%$ ).  $H2$  showed intermediate localization, enriched in the rim ( $56.0\% \pm 2.5\%$ ) compared to the core ( $20.2\% \pm 2.9\%$ ). This pattern was consistent across cohorts: in POEM, cores were 80.5%  $H1$ ; in PREDICT, 79.2%  $H1$ .  $H3$  was absent from POEM cores entirely and near-absent from PREDICT cores (0.4%).

**Correlation with histological tissue percentages.** As an exploratory analysis, we computed Spearman correlations between whole-tumor habitat proportions and histological tissue percentages. Correlations were weak and non-significant (Appendix C), reflecting the scale mismatch between voxel-level imaging and microscopic histology, as well as CT's inability to distinguish necrosis from fibrosis within the avascular compartment.

**A****B****C**

**Figure 7.4: Spatial architecture and histopathological validation of CT habitats.** (A) Schematic representation of habitat organization in colorectal liver metastases. H1 (Avascular, blue) localizes to the tumor core; H2 (Cellular-Perfused, yellow) represents viable proliferating tumor; H3 (Vascular, pink) localizes to the periphery at the tumor-liver interface. (B) Habitat proportions in whole tumor, rim (2mm outer shell), and core (interior) across both cohorts (POEM n=6, PREDICT n=42). H1 dominates the core (~80%); H2 and H3 dominate the rim. (C) Representative case from the POEM cohort showing correspondence between CT habitats and whole-tumor histology. Left: CT slice with habitat overlay. Right: H&E section. H1 coincides with necrotic/fibrotic regions (blue); H2 corresponds to viable tumor at the periphery (yellow).

## 7.4 Discussion

Habitat imaging offers a framework for quantifying intratumor heterogeneity, but most studies cluster voxels without establishing what the resulting regions represent biologically. In this chapter, we addressed this gap by developing an mpMRI-anchored framework for CT habitat discovery. We also addressed a question that has not yet been studied: which CT representation produces the most biologically coherent habitats? Using co-registered mpMRI as a proxy for tissue cellularity and vascularity, we compared four CT representations and selected the one whose habitats best separated biologically distinct tissue phenotypes. We then validated the spatial organization of these habitats against whole-tumor histopathology.

Handcrafted features outperformed deep learning embeddings for habitat computation. This was unexpected: foundation models trained on large CT datasets encode rich information, and one might assume they would produce superior results. However, for unsupervised voxelwise clustering, stability and spatial coherence matter more than representational capacity. Deep learning features produced fragmented patterns while handcrafted features produced contiguous regions aligned with biological gradients. We note that we did not fine-tune the foundation model or task-specific embeddings for this application; with domain-specific optimization, learned representations might perform better. In Section 9.3. we discuss further the use of handcrafted features in the era of deep learning.

The mpMRI characterization showed that CT habitats mostly reflect vascular architecture. This is not surprising given that we used contrast-enhanced portal-venous phase CT, where signal intensity directly reflects contrast uptake and thus tissue perfusion. Our interpretation of the three habitats was: avascular tissue (H1) with low  $f_v/K^{trans}$ ; a cellular-perfused habitat (H2) towards the rim with highest  $K^{trans}$  and cellularity indicating possible viable cancer, and finally a vascular rim (H3) with highest  $f_v$  and moderate-high  $K^{trans}$ . The  $f_v/K^{trans}$  dissociation in H3 has a plausible explanation: high vascular fraction but moderate permeability is consistent with the tumor-liver interface, where voxels mix tumor with normal liver parenchyma that has mature, less leaky vessels.

The spatial analysis confirmed biological plausibility. Across 48 tumors from two independent cohorts, H1 dominated tumor cores (~80%) while H2 and H3 were enriched at the rim (56% and 35%, respectively). This architecture (avascular interior, vascularized periphery) matches the known histology of colorectal liver metastases, where viable tumor surrounds a necrotic or fibrotic core [Poulsides et al., 2012, Van den Eynden et al., 2013]. Our initial hypothesis (that CT habitats would map onto necrosis, fibrosis, and viable tumor as discrete categories) was not fully supported. CT reliably separates vascularized from avascular tissue, but within the avascular core it cannot distinguish necrosis from fibrosis. Both are hypovascular, lack contrast enhancement, and exhibit similar texture on portal-venous imaging.

This study presented several limitations. The first one and most important: habitat labels represent the dominant phenotype for each voxel. In practice, each voxel (which has an average size of one cubic millimeter) contains several tissue phenotypes at once. The GMM clustering assigns each voxel a probability distribution across all three habitats and we report the maximum-probability assignment. This discretization loses information; a voxel with 40% probability for H1 and 35% for H2 is assigned to H1. Despite this inherent uncertainty, the spatial segregation is clear: avascular tissue dominates the core, vascular tissue dominates the rim.

Moreover, sample sizes were small ( $N = 10$  PREDICT,  $N = 6$  POEM), reflecting the difficulty of acquiring co-registered multimodal imaging and whole-tumor histology. CT-mpMRI registration introduces uncertainty that propagates into habitat analyses although our median accuracy (DSC 0.70–0.79) compares favorably with recent benchmarks [Demir et al., 2025]. The biological interpretation of H3 as tumor-liver interface is plausible but not proven; alternative explanations such as a stromal-rich compartment cannot be excluded. Finally, histopathological validation was qualitative since voxel-wise co-registration between CT and histology was not feasible.

Despite these limitations, the findings have practical relevance. CT is widely available, reproducible, and already part of standard oncologic care. If CT habitats can stratify patients by tumor vascular architecture, they could inform treatment decisions without additional imaging. The distinction between vascularized and avascular compartments is clinically meaningful: vascularized tissue is more likely to respond to systemic therapy, while avascular cores may indicate resistance. Whether habitat-derived metrics predict treatment response or survival is tested in Chapter 8.

## 7.5 Summary

We developed and validated an mpMRI-anchored CT habitat model for colorectal liver metastases. The central question was whether CT habitats can capture the dominant tissue phenotypes of these tumors. Using co-registered multiparametric MRI as biological reference and whole-tumor histopathology for validation, we found that CT habitats reflect vascular architecture, but not discrete histological categories.

### Key Points:

- Handcrafted radiomics features yield the most biologically coherent habitats. Comparing four CT representations (raw HU, handcrafted radiomics, DL-SALSA, DL-FM), handcrafted texture features produced habitats with the strongest separation of mpMRI-derived vascular parameters and the highest spatial coherence.
- CT habitats capture vascular heterogeneity, not discrete histological compartments. The three habitats represent an avascular core (H1), a cellular-perfused zone with leaky neovessels (H2), and a vascular rim likely reflecting the tumor-liver interface (H3). CT cannot distinguish necrosis from fibrosis within the avascular compartment.
- Spatial architecture is biologically plausible and consistent across cohorts. H1 dominates tumor cores (~80%); H2 and H3 dominate the rim. This organization matches the known histology of colorectal liver metastases and was validated against whole-tumor histopathology in resected specimens.

# Chapter 8

## Clinical Relevance of CT Habitats

In Chapter 7, we developed an mpMRI-anchored CT habitat model that segments colorectal liver metastases into three biologically meaningful compartments: avascular core, cellular-perfused viable tumor, and vascular rim. Histopathological validation confirmed that this spatial architecture reflects genuine tumor biology. In this chapter, we address the central follow-up question: Do these biologically-grounded habitats provide clinical value beyond standard imaging biomarkers?

### Contributions:

- We assess whether habitat-derived metrics predict survival independently of tumor volume across two clinical contexts: resectable disease treated with curative-intent surgery (TCIA, n=189) and unresectable disease treated with palliative systemic therapy (VHIO, n=344).
- We evaluate how treatment context modifies prognostic value, testing whether habitat metrics are informative after neoadjuvant chemotherapy (TCIA) and with anti-angiogenic therapy (VHIO).
- We test whether prognostic signal concentrates at specific spatial locations, comparing rim, core, and whole-tumor metrics.
- We explore whether longitudinal habitat changes during treatment capture response biology that RECIST classification misses, focusing on the ambiguous stable disease category.

The work presented in Chapters 7 and 8 forms the basis of a manuscript currently in preparation.

### 8.1 Rationale

Tumor volume is the dominant imaging biomarker in oncology. Together with anatomical location, it determines resectability and guides the choice between curative surgery and palliative systemic therapy. RECIST criteria [Eisenhauer et al., 2009], which inform treatment decisions across solid tumors, define response and progression primarily through size change. This framework works reasonably well when

## CLINICAL RELEVANCE OF CT HABITATS

tumors shrink or grow substantially, but there are moments during treatment when changes in size are not informative.

The limitation is not volume itself but what volume ignores: the internal composition of the tumor. A  $30\text{ cm}^3$  metastasis that is 80% necrotic differs fundamentally from one that is 80% viable tumor, yet both receive identical volumetric assessment. This matters for two clinical problems. First, response assessment: a tumor that neither shrinks nor grows might be genuinely controlled by treatment, or it might be slowly progressing. Survival outcomes among volumetrically similar patients vary widely, yet current imaging cannot distinguish these scenarios. Second, treatment selection: many patients do not respond as expected to first-line therapy, and we lack imaging tools to predict who will benefit from which regimen. Both problems (assessing response and selecting treatment) could benefit from imaging biomarkers that characterize tumor heterogeneity beyond size.

Habitat imaging offers a potential solution. By segmenting tumors into spatially distinct subregions with different imaging phenotypes, habitats capture the heterogeneity information that volume discards. In Chapter 7, we established that CT habitats in colorectal liver metastases reflect vascular architecture: an avascular compartment representing necrotic or fibrotic tissue, a cellular-perfused compartment representing viable tumor, and a vascular rim at the tumor-liver interface. If these phenotypes have prognostic relevance, habitat-derived metrics should predict survival independently of volume and potentially discriminate outcomes among volumetrically similar patients.

This chapter tests these hypotheses across two cohorts with complementary clinical contexts. First, the TCIA cohort ( $n=189$ , resectable disease): Patients underwent hepatic resection with curative intent; 60% received neoadjuvant chemotherapy. We ask: Does neoadjuvant treatment alter habitat composition? Do post-treatment habitat patterns predict survival? We hypothesize that treatment remodels tumor composition in ways detectable by CT, and that spatial heterogeneity at the invasive rim will carry prognostic information beyond what volume provides.

Second, VHIO cohort ( $n=344$ , unresectable disease): Patients received first-line systemic therapy (chemotherapy alone, chemotherapy plus bevacizumab, or chemotherapy plus targeted agents). We ask: Do baseline habitat metrics predict survival? Does prognostic value depend on treatment type and molecular context? We hypothesize that for cytotoxic chemotherapy, where response manifests as tumor shrinkage, volume will dominate. For anti-angiogenic therapy, where treatment slows growth without necessarily shrinking tumors, spatial heterogeneity at the rim may capture treatment-relevant biology that volume cannot.

A subset of VHIO patients ( $n=38$ ) had paired baseline and follow-up imaging, allowing us to ask a third question: Does the direction of habitat change during treatment predict outcomes beyond what RECIST categories provide? If rim heterogeneity captures treatment response, then the direction of rim entropy change may differ systematically between responders and progressors.

## 8.2 Methods

### 8.2.1 Patient Cohorts

Two independent cohorts were analyzed (Figure 8.1A, Table 8.1). The TCIA cohort included 189 patients with resectable colorectal liver metastases from a publicly available dataset. All patients underwent hepatic resection with curative intent; 115 (60.8%) received neoadjuvant chemotherapy prior to surgery. CT imaging was acquired pre-surgery, representing post-neoadjuvant status in treated patients and treatment-naïve status in untreated patients.

The VHIO cohort included 344 patients with unresectable metastatic colorectal cancer treated at Vall d'Hebron Institute of Oncology. All patients received first-line systemic therapy: chemotherapy alone (n=122), chemotherapy plus bevacizumab (n=133), or chemotherapy plus targeted therapy (n=69). CT imaging was acquired at baseline (pre-treatment) for all patients; a subset of 38 patients had paired follow-up imaging for longitudinal analysis. RAS mutation status was available for 331 patients (96%): 136 RAS wild-type, 195 RAS mutant. Inclusion criteria and image preprocessing for both cohorts are detailed in Section 5.1.

### 8.2.2 Habitat Computation

CT habitats were computed using the biologically-anchored model developed in Chapter 7 (Section 7.3.1). Briefly, six handcrafted features were computed from contrast-enhanced CT (portal phase) and clustered using a Gaussian mixture model (K=3) in two steps (hybrid mode in the imaging habitats pipeline, see Section 5.2). The handcrafted representation was selected over alternatives (raw HU, deep learning features) based on its superior ability to separate biologically distinct tissue compartments as measured by co-registered multiparametric MRI.

This produced three habitats (Figure 8.1B): the avascular habitat (corresponding to necrotic-like or poorly perfused tissue with low cellularity), the cellular-perfused habitat (highest cellularity and perfusion, corresponding to viable tumor), and the vascular habitat (moderate cellularity and high vascularity). The vascular habitat may partly reflect partial volume effects at the tumor-liver interface. The biologically-anchored approach ensured that habitats captured cellular and vascular heterogeneity validated against ground truth rather than arbitrary clusters.

### 8.2.3 Habitat-Derived Quantitative Metrics

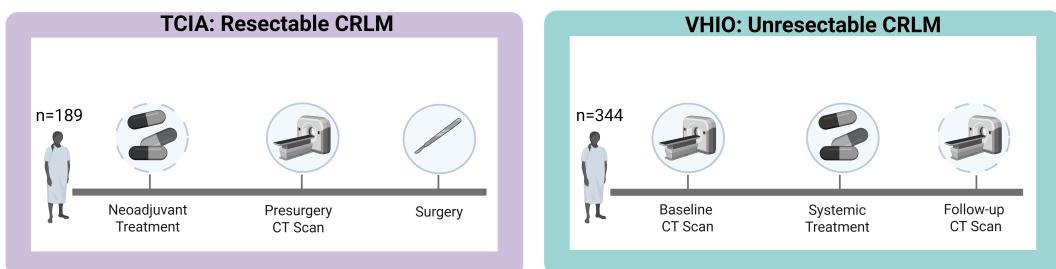
Quantitative metrics were derived at three spatial scales: whole tumor, rim (outer 2mm), and core (tumor interior) (Figure 8.1C).

For each scale, we computed habitat proportions (fraction of voxels assigned to each habitat) and Shannon entropy (diversity of habitat composition). Higher entropy indicates a more heterogeneous mixture of

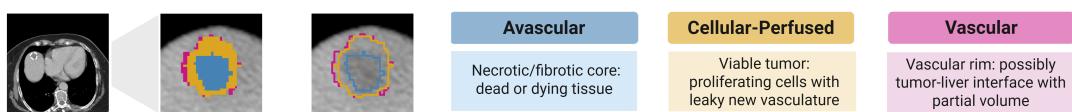
## CLINICAL RELEVANCE OF CT HABITATS

habitats; lower entropy indicates dominance by a single habitat. Tumor volume was computed as the sum of tumor voxels multiplied by voxel spacing. These metrics follow the definitions in Section 5.2.3. For patients with multiple liver metastases, metrics were aggregated across lesions using volume-weighted averaging, giving greater weight to larger lesions.

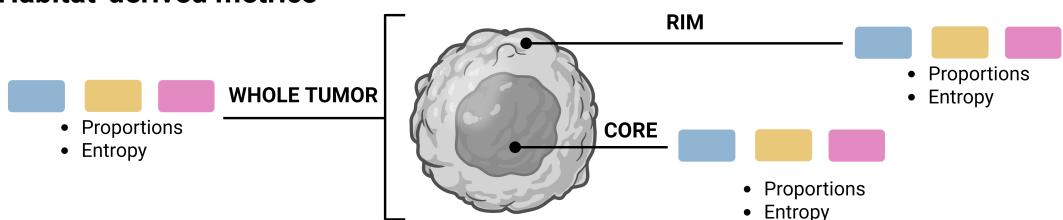
### A Cohorts



### B Habitats



### C Habitat-derived metrics



**Figure 8.1: Clinical validation study design.** (A) Two independent cohorts with distinct clinical scenarios. TCIA (n=189): resectable colorectal liver metastases, CT acquired pre-surgery (60.8% received neoadjuvant chemotherapy). VHIO (n=344): unresectable metastatic colorectal cancer receiving first-line systemic therapy, CT acquired at baseline; a subset (n=38) had paired follow-up imaging for longitudinal analysis. (B) CT habitats derived from the biologically-anchored model (Chapter 7). The avascular habitat (blue) dominates tumor cores. The cellular-vascular habitat (yellow) and vascular habitat (red) enrich at the invasive rim. The vascular habitat may partly reflect partial volume effects at the tumor-liver interface. (C) Habitat-derived quantitative metrics computed at three spatial scales (whole tumor, rim, core): habitat proportions and Shannon entropy. Tumor volume computed as sum of tumor voxels  $\times$  voxel spacing. For patients with multiple lesions, metrics were aggregated using volume-weighted averaging. Created with BioRender.com.

## 8.2.4 Statistical Analyses

Group comparisons used Mann-Whitney U tests [Mann and Whitney, 1947] with false discovery rate (FDR) correction for multiple comparisons [Benjamini and Hochberg, 1995]. Survival analyses used Cox proportional hazards regression [Cox, 1972] in a two-stage approach. Stage 1 screened all variables in univariable models, retaining those with  $p < 0.10$ . Stage 2 fitted a multivariable model with retained

variables. Hazard ratios (HR) are reported per standard deviation increase. Model discrimination was assessed using Harrell's C-index [Harrell et al., 1996]. The proportional hazards assumption was tested using Schoenfeld residuals [Schoenfeld, 1982]. Kaplan-Meier curves [Kaplan and Meier, 1958] were generated using median splits, with differences assessed by log-rank test [Peto et al., 1976]. Statistical analyses used Python 3.10 with lifelines 0.27 [Davidson-Pilon, 2019] and scipy 1.11. Significance was set at  $p<0.05$ .

## 8.3 Results

### 8.3.1 Patient characteristics

Clinical characteristics of both cohorts are summarized in Table 8.1. The TCIA cohort included 189 patients with resectable colorectal liver metastases who underwent hepatic resection with curative intent; 115 (60.8%) received neoadjuvant chemotherapy prior to surgery. The VHIO cohort included 344 patients with unresectable disease receiving first-line systemic therapy. VHIO patients had higher liver tumor burden (median  $32.4 \text{ cm}^3$  vs  $10.7 \text{ cm}^3$ ) and more metastases per patient (median 3 vs 2), consistent with their more advanced disease stage. Overall survival was longer in TCIA (median 67.1 months) than VHIO (median 19.1 months), reflecting the curative versus palliative treatment intent.

### 8.3.2 Habitat Characteristics Across Cohorts

**Spatial architecture.** Despite differences in tumor burden and clinical setting, both cohorts showed the same spatial pattern observed in POEM and PREDICT (Section 7.3.3): cores dominated by the avascular habitat (~75%), rims enriched in cellular-perfused (~50%) and vascular (~35%) habitats (Figure 8.2A). This consistency confirms that the CT habitat model captures generalizable tumor biology rather than cohort-specific artifacts.

**Correlation with volume.** Habitat metrics were correlated with tumor volume (Figure 8.2B). The strongest relationship was between rim entropy and volume ( $\rho = -0.77$  TCIA,  $\rho = -0.71$  VHIO, both  $p<0.001$ ): larger tumors have more homogeneous rims, while smaller tumors maintain greater spatial diversity at the invasive margin. Volume also correlated positively with rim cellular-perfused proportion ( $\rho = 0.49$  TCIA,  $\rho = 0.45$  VHIO) and whole avascular proportion ( $\rho = 0.44$  TCIA,  $\rho = 0.70$  VHIO), and negatively with whole entropy ( $\rho = -0.44$  TCIA,  $\rho = -0.62$  VHIO), all  $p<0.001$ .

Our interpretation is as follows: as tumors grow, their cores outpace blood supply, leading to central necrosis and expansion of the avascular core. Larger tumors therefore have proportionally more avascular core, reducing overall habitat diversity. At the rim, larger tumors show higher cellular-perfused proportion—more viable tumor at the invasive front—and lower entropy. The result is that large tumors appear more homogeneous on CT: dominated by avascular core with a uniform, viable rim. Smaller tumors retain more balanced composition and greater spatial diversity.

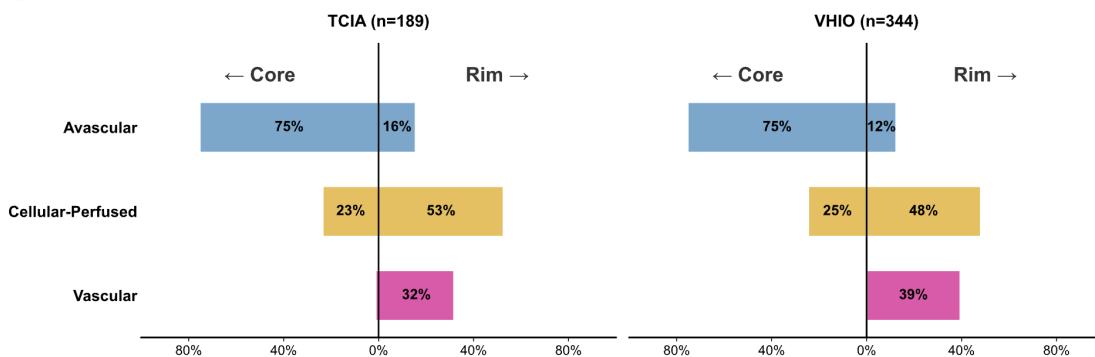
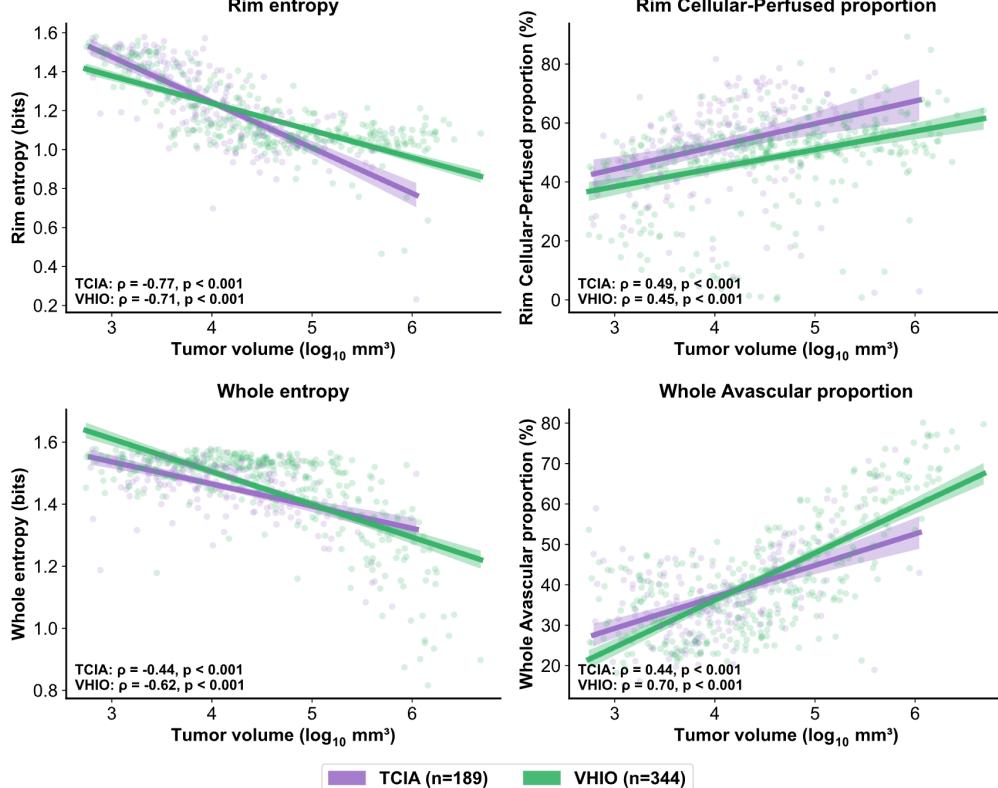
## CLINICAL RELEVANCE OF CT HABITATS

**Table 8.1: Clinical characteristics of the TCIA and VHIO cohorts.** TCIA patients had resectable disease treated with curative-intent surgery; VHIO patients had unresectable disease treated with palliative systemic therapy. IQR = interquartile range; OS = overall survival; DFS = disease-free survival; PFS = progression-free survival.

Variables	TCIA (n=189)	VHIO (n=344)
<b>Age [years, Median(range)]</b>	61 (30 – 88)	69 (32 – 88)
<b>Sex [n (%)]</b>		
Male	111 (58.7)	205 (59.6)
Female	78 (41.3)	139 (40.4)
<b>Primary Tumor Location [n (%)]</b>		
Right	–	139 (40.4)
Left	–	183 (53.2)
Rectum	–	17 (4.9)
Unknown	189 (100)	5 (1.5)
<b>RAS Status [n (%)]</b>		
Wild-type	–	136 (39.5%)
Mutant	–	195 (56.7%)
Unknown	–	13 (3.8%)
<b>Synchronous CRLM [n(%)]</b>	104 (55.0)	277 (80.5)
<b>Extrahepatic Disease [n(%)]</b>	15 (7.9)	203 (59.0)
<b>No. of liver metastases per patient</b> [Median (IQR)]	2.0 (1.0, 3.0)	3.0 (2.0, 7.0)
<b>Median liver metastasis size</b> [cm <sup>3</sup> , Median (IQR)]	3.2 (1.2, 10.9)	5.1 (2.0, 15.0)
<b>Liver Disease Volume</b> [cm <sup>3</sup> , Median (IQR)]	10.7 (4.1, 32.8)	32.4 (6.8, 180.3)
<b>First-line Treatment Type [n (%)]</b>	–	
Chemotherapy Only	–	122 (35.5)
Chemotherapy + Antiangiogenic	–	133 (38.7)
Chemotherapy + Targeted	–	69 (20.1)
Other	–	20 (5.8%)
<b>Neoadjuvant chemotherapy [n (%)]</b>	115 (60.8)	–
<b>Progression Free Survival</b> [months, Median (IQR)]	–	8.9 (5.0, 15.3)
<b>Overall Survival</b> [months, Median (IQR)]	67.1 (34.4, 97.5)	19.1 (11.1, 32.8)
<b>Disease Free Survival</b> [months, Median (IQR)]	22.3 (9.4, 69.3)	8.9 (5.0, 15.3)

### 8.3.3 TCIA: Neoadjuvant Chemotherapy Remodels Tumor Composition

We examined whether neoadjuvant chemotherapy alters habitat composition. Figure 8.3A compares habitat metrics between treatment-naive (n=74) and neoadjuvant-treated (n=115) patients.

**A****B**

**Figure 8.2: Spatial architecture and habitat-volume correlations across cohorts.** (A) Diverging barplot showing habitat composition in tumor cores (left) and rims (right) for TCIA and VHIO. Both cohorts show the same pattern: cores dominated by avascular habitat, rims enriched in cellular-perfused and vascular habitats. (B) Scatter plots showing key correlations between habitat metrics and tumor volume ( $\log_{10}$  scale). Larger tumors have more homogeneous rims (rim entropy:  $\rho = -0.77$  TCIA,  $\rho = -0.71$  VHIO), more viable tumor at the rim (rim cellular-perfused:  $\rho = 0.49$  TCIA,  $\rho = 0.45$  VHIO), higher avascular proportion (whole avascular:  $\rho = 0.44$  TCIA,  $\rho = 0.70$  VHIO), and lower diversity (whole entropy:  $\rho = -0.44$  TCIA,  $\rho = -0.62$  VHIO). All  $p < 0.001$ . Full correlation matrix in Appendix D.

Several metrics differed between groups after FDR correction. Neoadjuvant-treated tumors showed higher whole entropy (median 1.51 vs 1.45, adjusted  $p=0.001$ ), higher vascular habitat proportion (21.4% vs 16.8%, adjusted  $p=0.001$ ), and higher rim entropy (1.25 vs 1.13, adjusted  $p=0.001$ ). Avascular proportion was lower in treated tumors (35.4% vs 39.8%, adjusted  $p=0.057$ ) and tumor volume was smaller (median

## CLINICAL RELEVANCE OF CT HABITATS

8,620 vs 14,639 mm<sup>3</sup>, adjusted p=0.057), though neither reached the corrected significance threshold. Both whole entropy and rim entropy showed stronger evidence of treatment-associated difference than tumor volume (adjusted p=0.001 vs p=0.057), suggesting that compositional remodeling may be a more sensitive marker of treatment effect than size reduction alone.

Surprisingly, cellular-perfused habitat proportion did not differ between groups at any spatial scale (whole: p=0.75; rim: p=0.057; core: p=0.23). If this habitat represents viable tumor, one might expect treatment to reduce it. The lack of difference suggests that neoadjuvant chemotherapy may not selectively eliminate the cellular-perfused compartment—or that treated and untreated tumors in this cohort had similar proportions of viable tissue at the time of imaging, perhaps because responding tumors were selected for surgery while non-responders were excluded. The treatment effect appears primarily as increased heterogeneity (entropy) rather than selective habitat depletion. Extended results are available in Appendix D.

### 8.3.4 TCIA: Prognostic Value Depends on Treatment Context and Spatial Location

We next assessed whether habitat metrics predict survival. Rim metrics were the most prognostic and only in neoadjuvant-treated patients, not in treatment-naive tumors (Table 8.2).

In treatment-naive patients (n=74, 36 events), no habitat metric reached significance (all p>0.15), and tumor volume showed no association with survival (HR=0.99, p=0.97). Extrahepatic disease was the only prognostic factor, though it violated the proportional hazards assumption. The model achieved only marginal discrimination (C-index 0.611).

In neoadjuvant-treated patients (n=115, 66 events), a spatial gradient emerged. Rim metrics were strongly prognostic in univariable analysis: rim entropy (HR=0.59, p<0.001), rim avascular proportion (HR=0.60, p<0.001), and rim cellular-perfused proportion (HR=1.65, p=0.001). Whole-tumor metrics showed intermediate effects: whole entropy (HR=0.62, p=0.001), whole cellular-perfused proportion (HR=1.49, p=0.010). Core metrics showed no associations (all p>0.12). Tumor volume was also prognostic (HR=2.01, p=0.001). Kaplan-Meier curves confirmed these patterns (Figure 8.3B): high rim cellular-perfused proportion predicted worse survival (median 48 vs 92 months, p=0.002); high rim entropy was protective (median 92 vs 51 months, p=0.002); core entropy showed no separation (p=0.087).

Across metrics, higher entropy and higher avascular proportion were protective, while higher cellular-perfused proportion predicted worse outcomes. This is consistent with the interpretation that residual cellular-perfused habitat after treatment indicates viable, treatment-resistant tumor, while high entropy or avascular dominance (perhaps indicating treatment-induced necrosis or fibrosis) indicates treatment response.

The pattern was consistent across metrics: higher entropy and higher avascular proportion were protective, while higher cellular-perfused proportion predicted worse outcomes. Patients with high rim cellular-perfused proportion had markedly worse survival than those with low rim cellular-perfused (median 48 vs 92 months, p=0.002). High rim entropy (median 92 vs 51 months, p=0.002) and high rim avascular proportion (median 109 vs 52 months, p=0.002) were protective. Core entropy showed no significant

separation ( $p=0.087$ ), confirming that prognostic information concentrates at the tumor rim rather than the interior (Figure 8.3B).

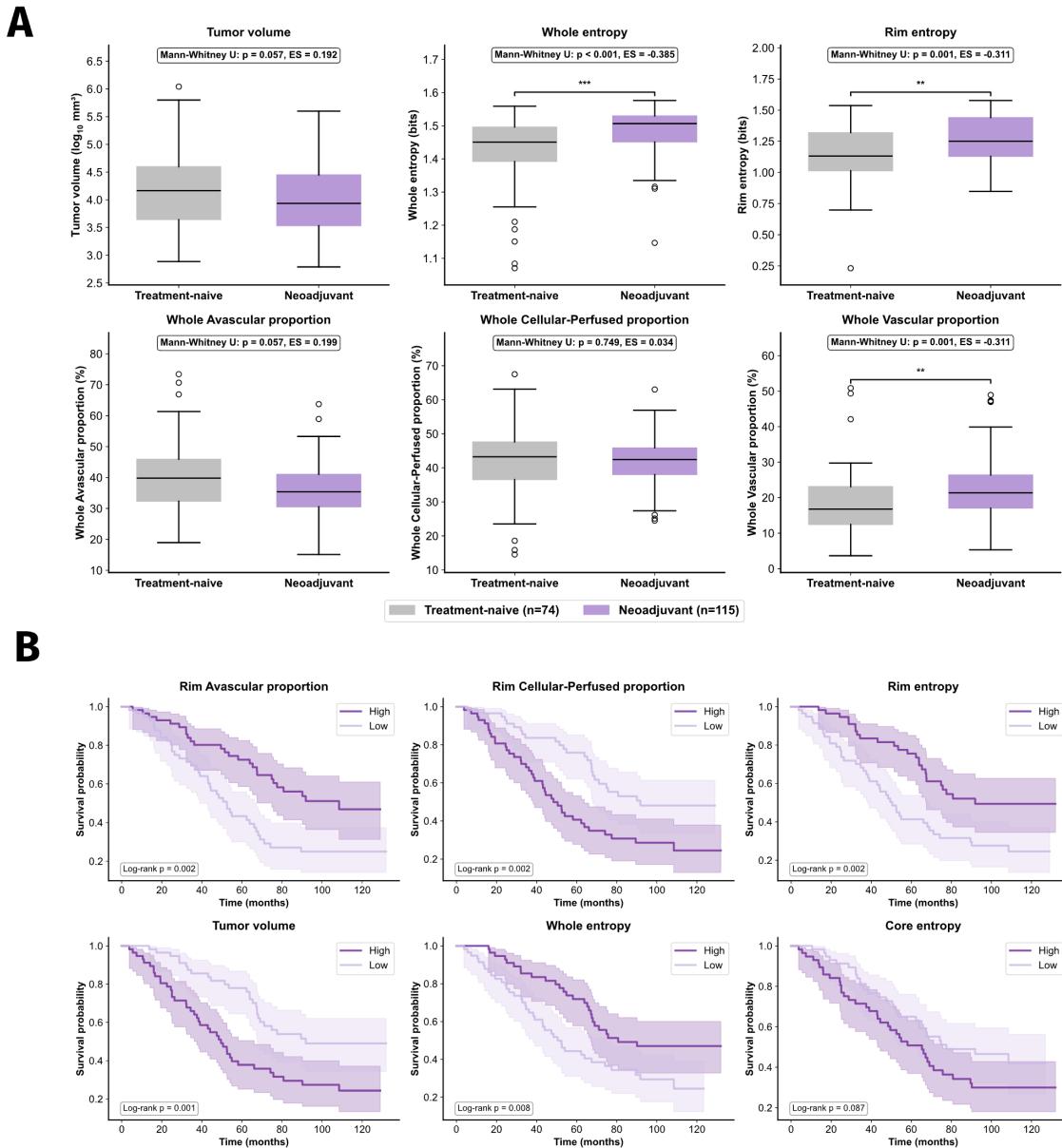
In multivariable analysis, most rim metrics lost significance despite strong univariable associations. This attenuation reflects the collinearity documented in Section 8.3.2: rim entropy correlates strongly with tumor volume ( $\rho = -0.77$ ), and rim metrics correlate with each other. When modeled jointly, they compete for the same variance. Whole cellular-perfused proportion was the only habitat metric to retain independent significance ( $HR=1.66$ , 95% CI 1.04–2.65,  $p=0.033$ ), likely because it shares less variance with volume than the rim metrics do. The multivariable model showed improved discrimination over the treatment-naive model (C-index 0.699 vs 0.611), confirming that habitat information adds prognostic value in the post-treatment context.

**Table 8.2: Cox regression for overall survival in neoadjuvant-treated TCIA patients.** Multivariable model includes variables with univariable  $p<0.10$ . Dashes indicate variables not entered. Bold indicates  $p<0.05$ .

Variable	Univariable		Multivariable	
	HR [95% CI]	P-value	HR [95% CI]	P-value
<b>Clinical</b>				
Extrahepatic disease	2.61 [1.28–5.30]	<b>0.008</b>	2.64 [0.96–7.27]	0.059
Synchronous metastases	0.76 [0.46–1.25]	0.280	—	—
<b>Tumor volume</b>				
Tumor volume	2.01 [1.31–3.07]	<b>0.001</b>	1.90 [0.75–4.80]	0.173
<b>Habitats-Whole Tumor</b>				
Whole entropy	0.62 [0.47–0.83]	<b>0.001</b>	0.73 [0.32–1.66]	0.455
Whole Avasc. prop.	0.92 [0.69–1.23]	0.585	—	—
Whole Cell.-Perf. prop.	1.49 [1.10–2.01]	<b>0.010</b>	1.66 [1.04–2.65]	<b>0.033</b>
<b>Habitats-Rim</b>				
Rim entropy	0.59 [0.45–0.78]	< <b>0.001</b>	0.90 [0.60–1.35]	0.622
Rim Avasc. prop.	0.60 [0.45–0.80]	< <b>0.001</b>	0.60 [0.29–1.24]	0.168
Rim Cell.-Perf. prop.	1.65 [1.21–2.23]	<b>0.001</b>	0.59 [0.26–1.33]	0.201
<b>Habitats-Core</b>				
Core entropy	1.22 [0.95–1.57]	0.121	—	—
Core Avasc. prop.	1.11 [0.88–1.41]	0.370	—	—
Core Cell.-Perf. prop.	0.87 [0.69–1.09]	0.224	—	—
<b>Model C-index</b>				<b>0.699</b>

### 8.3.5 VHIO: Prognostic Value Depends on Treatment and Molecular Context

We then studied whether baseline habitat metrics predict survival in unresectable disease and we found that it's context-dependent, similar to the resectable clinical scenario.



**Figure 8.3: Neoadjuvant chemotherapy remodels tumor composition.** (A) Boxplots comparing habitat metrics between treatment-naive (n=74, grey) and neoadjuvant-treated (n=115, purple) TCIA patients. Treated tumors show higher whole entropy, vascular proportion, and rim entropy (FDR  $p < 0.05$ ). Avascular proportion and volume show trends ( $p = 0.057$ ). Mann-Whitney U test; \*\*\* $p < 0.001$ , \*\* $p < 0.01$ . Extended results are available in Appendix D. (B) Kaplan-Meier curves for overall survival in neoadjuvant-treated patients, stratified by median split. Top row: rim metrics (avascular, cellular-perfused, entropy)—all prognostic ( $p = 0.002$ ). Bottom row: tumor volume ( $p = 0.001$ ), whole entropy ( $p = 0.008$ ), core entropy (not significant,  $p = 0.087$ ). Prognostic signal concentrates at the rim.

## Overall Cohort

In the full VHIO cohort (n=344, 246 events), tumor volume was prognostic in multivariable analysis (HR=1.29, 95% CI 1.08–1.55,  $p=0.006$ ), along with clinical factors that were not surprising: extrahepatic disease (HR=1.50,  $p=0.003$ ), left-sided primary (HR=0.61,  $p < 0.001$  vs right-sided), rectal primary (HR=0.32,  $p=0.001$ ), and older age (HR=1.19,  $p=0.038$ ). Rim entropy showed a trend toward protection

(HR=0.88, p=0.10) but did not survive adjustment. No other habitat metric reached significance.

This partially parallels TCIA, where baseline habitat metrics also failed to predict survival in treatment-naïve patients. The difference is that in VHIO, volume was prognostic—probably because unresectable disease spans a wider range of tumor burden where size meaningfully stratifies risk. In resectable TCIA, where patients are selected for surgery based on resectability criteria, volume varies less and other factors (like lesion location and surgical margins) likely matter more.

## Treatment-Stratified Analysis

We hypothesized that habitat metrics would matter more for cytostatic therapies, where treatment slows growth without necessarily shrinking tumors, than for cytotoxic chemotherapy, where response manifests as cell death and volume reduction.

The data supported this hypothesis (Figure 8.4A, Table 8.3). In patients receiving chemotherapy alone (n=122), tumor volume dominated with a massive effect size (HR=5.46, 95% CI 3.19–9.34, p<0.001). No habitat metric added value. In this cytotoxic context, tumor burden determines outcomes.

In patients receiving chemotherapy plus bevacizumab (n=133), a different pattern emerged. Volume remained prognostic (HR=1.69, p<0.001), but rim entropy emerged as an independent predictor (HR=0.68, 95% CI 0.52–0.88, p=0.004). Patients with more heterogeneous rims lived longer. This finding has a plausible biological explanation: bevacizumab targets VEGF, and a heterogeneous vascular rim may indicate diverse vessel populations—some mature, some immature—that respond differentially to anti-angiogenic attack. Alternatively, baseline rim heterogeneity may reflect the desmoplastic reaction, which is associated with both better prognosis and potentially different vascular biology. In patients receiving chemotherapy plus targeted therapy (n=69), volume showed a modest effect (HR=1.21, p=0.043) but no habitat metric reached significance. These patients receive anti-EGFR agents, which target a different pathway; rim heterogeneity may be less relevant to this treatment mechanism.

## RAS-Stratified Analysis

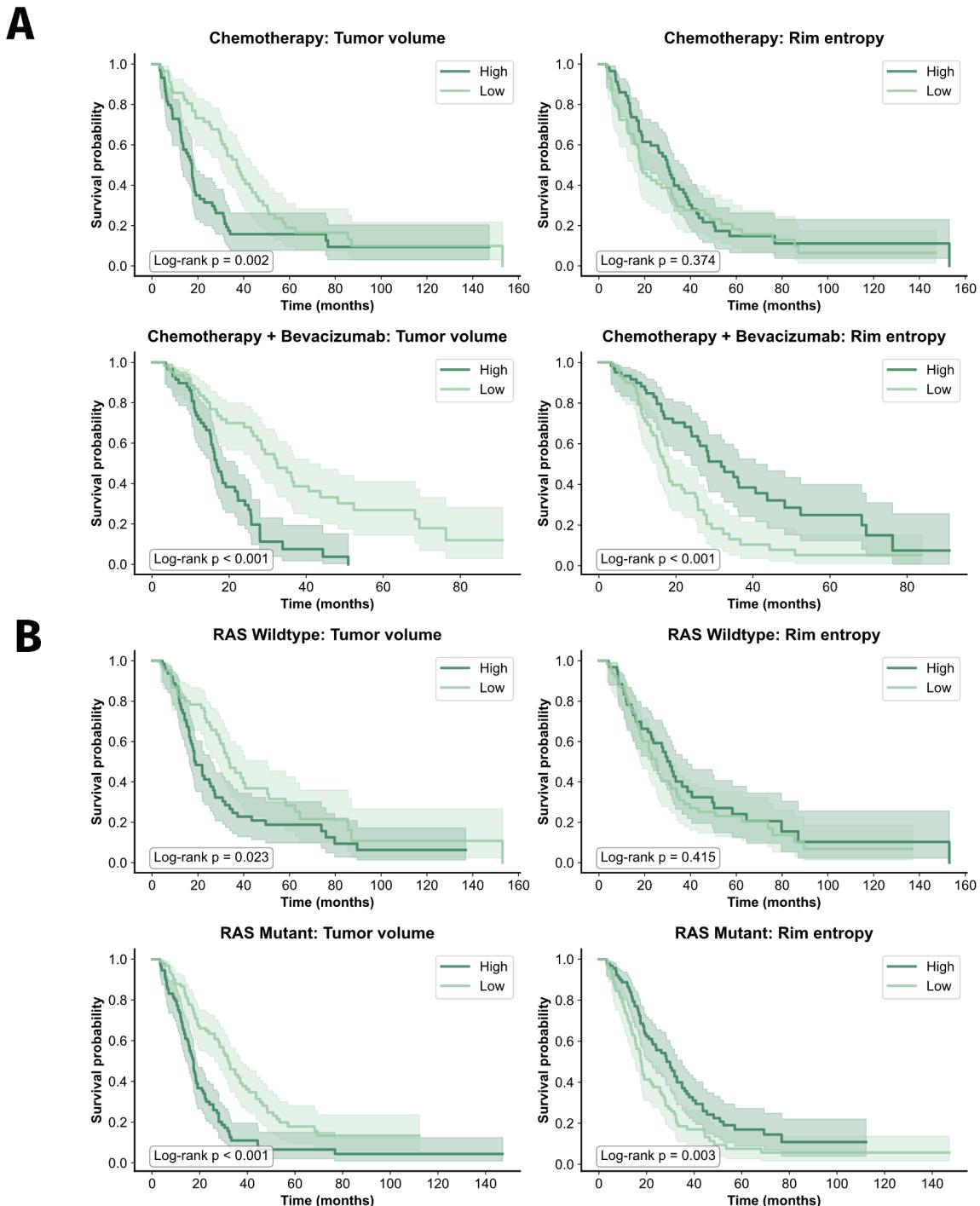
RAS-mutant tumors cannot receive anti-EGFR therapy and are typically treated with bevacizumab when combination therapy is indicated. If rim entropy captures biology relevant to anti-angiogenic response, it should also be prognostic in RAS-mutant patients regardless of the specific treatment they received.

As predicted, in RAS wild-type patients (n=136), only volume was prognostic (HR=1.23, p=0.028); rim entropy showed no association (HR=0.96, p=0.76). In RAS-mutant patients (n=195), both volume (HR=1.89, p<0.001) and rim entropy (HR=0.80, 95% CI 0.65–1.00, p=0.047) were independently prognostic (Figure 8.4B).

The bevacizumab and RAS-mutant findings converge—unsurprising given that these groups overlap substantially—RAS-mutant patients often receive bevacizumab—and both show the same rim entropy signal. This suggests we are capturing real biology, not statistical noise. Extended results are available in Appendix D.

Context-dependent prognostic value of habitat metrics in VHIO is summarized in Table 8.3.

## CLINICAL RELEVANCE OF CT HABITATS



**Figure 8.4: Treatment and molecular context modify prognostic value in VHIO.** Kaplan-Meier curves for overall survival stratified by tumor volume (left column) and rim entropy (right column), using median splits. (A) Treatment-stratified analysis. Top row: Chemotherapy alone (n=122)—volume separates survival ( $p=0.002$ ) but rim entropy does not ( $p=0.374$ ). Bottom row: Chemotherapy plus bevacizumab (n=133). Both volume ( $p<0.001$ ) and rim entropy ( $p<0.001$ ) separate survival. (B) RAS-stratified analysis. Top row: RAS wild-type (n=136)—volume separates survival ( $p=0.023$ ) but rim entropy does not ( $p=0.415$ ). Bottom row: RAS-mutant (n=195). Both volume ( $p<0.001$ ) and rim entropy ( $p=0.003$ ) separate survival. The rim entropy signal appears specifically in contexts where anti-angiogenic therapy is relevant: bevacizumab-treated and RAS-mutant patients (who typically receive bevacizumab rather than anti-EGFR therapy).

**Table 8.3: Context-dependent prognostic value of habitat metrics in VHIO.** Multivariable hazard ratios adjusted for clinical covariates (extrahepatic disease, primary site, age, synchronous metastases where  $p < 0.10$ ). Bold indicates  $p < 0.05$ . Rim entropy is independently prognostic in bevacizumab-treated and RAS-mutant patients—contexts where anti-angiogenic therapy is relevant.

	n	Events	Volume		Rim Entropy			C-index
			HR [95% CI]	p	HR [95% CI]	p	C-index	
<b>All patients</b>	344	246	1.29 [1.08–1.55]	<b>0.006</b>	0.88 [0.76–1.02]	0.10	0.665	
<b>By Treatment</b>								
Chemo alone	122	–	5.46 [3.19–9.34]	<b>&lt;0.001</b>	0.88 [0.54–1.43]	0.61	0.703	
Chemo + Bevacizumab	133	–	1.69 [1.31–2.17]	<b>&lt;0.001</b>	0.68 [0.52–0.88]	<b>0.004</b>	0.674	
Chemo + Targeted	69	–	1.21 [1.01–1.47]	<b>0.043</b>	1.14 [0.81–1.60]	0.46	0.723	
<b>By RAS Status</b>								
RAS Wild-Type	136	–	1.23 [1.02–1.48]	<b>0.028</b>	0.96 [0.77–1.21]	0.76	0.705	
RAS Mutant	195	–	1.89 [1.41–2.53]	<b>&lt;0.001</b>	0.80 [0.65–1.00]	<b>0.047</b>	0.700	

### Biological Interpretation: Rim Entropy and Growth Patterns

We do not know what rim entropy is measuring at the tissue level. But one hypothesis emerged from the POEM cohort, where an experienced pathologist annotated histopathological growth patterns. Colorectal liver metastases exhibit either desmoplastic growth (tumor surrounded by fibrous stroma, associated with better prognosis) or replacement growth (tumor directly replacing hepatocytes, associated with worse prognosis). Growth pattern is currently assessed only through histopathology.

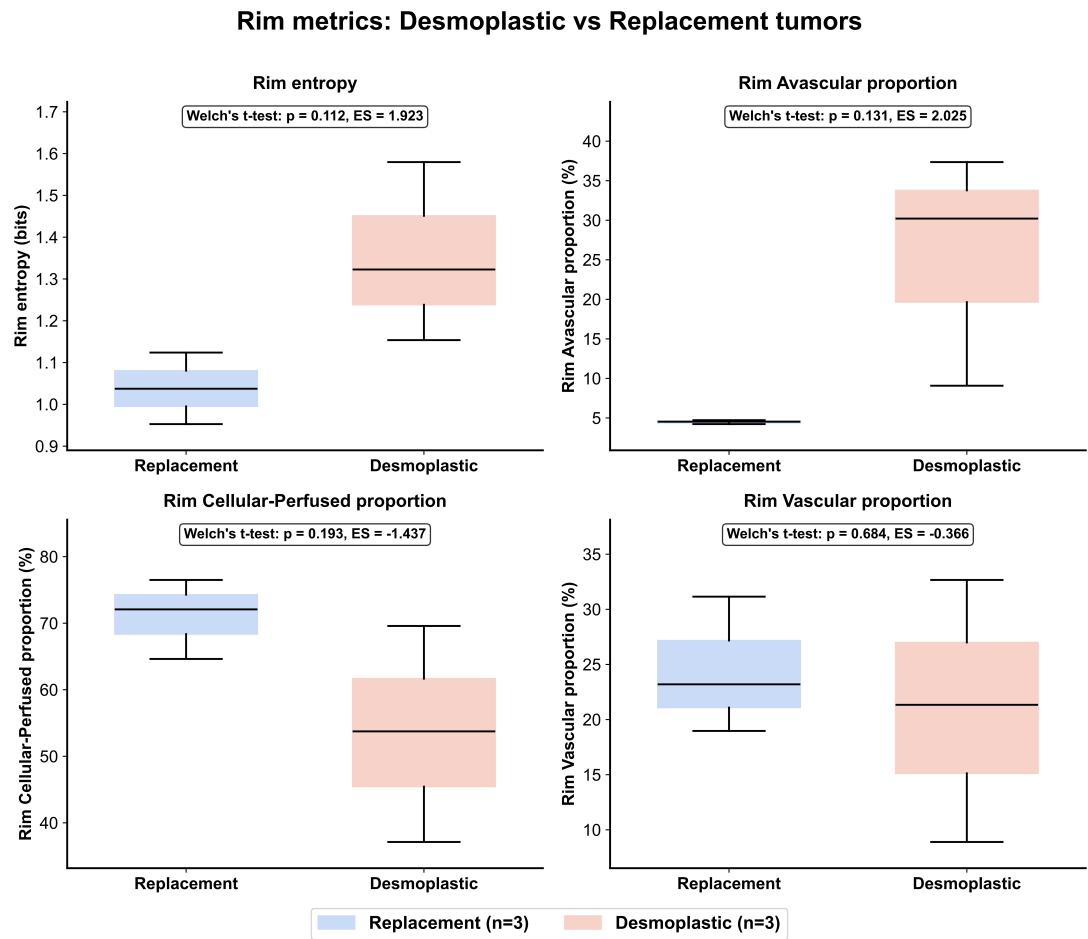
In the six POEM tumors with annotated growth patterns, desmoplastic tumors ( $n=3$ ) showed higher rim entropy than replacement tumors ( $n=3$ ), with large effect sizes despite the small sample (rim entropy: Cohen's  $d = 1.92$ ; rim avascular proportion:  $d = 2.03$ ; (Figure 8.5). Desmoplastic tumors also showed lower rim cellular-perfused proportion ( $d = -1.44$ ), consistent with the fibrous stroma displacing viable tumor at the interface.

This is exploratory—six tumors cannot establish a robust correlation. But the pattern suggests that rim entropy may capture the desmoplastic reaction: the fibrous, heterogeneous interface that characterizes less aggressive disease. If confirmed in larger cohorts, CT habitat analysis could provide a non-invasive surrogate for growth pattern assessment, currently requiring resection or biopsy.

### 8.3.6 VHIO: Rim Entropy Change During Treatment

RECIST classifies patients by tumor size change, but PR and SD categories have overlapping survival curves (Figure 8.6A, left)—size alone does not cleanly separate outcomes. We asked whether rim entropy change could provide additional information.

In 38 patients with paired imaging, rim entropy direction did not significantly predict OS (HR=0.90,  $p=0.76$ ). However, the pattern was biologically consistent: partial responders showed increasing entropy (median  $\Delta = +0.053$ , median OS = 32.8 months), while progressors showed decreasing entropy (median

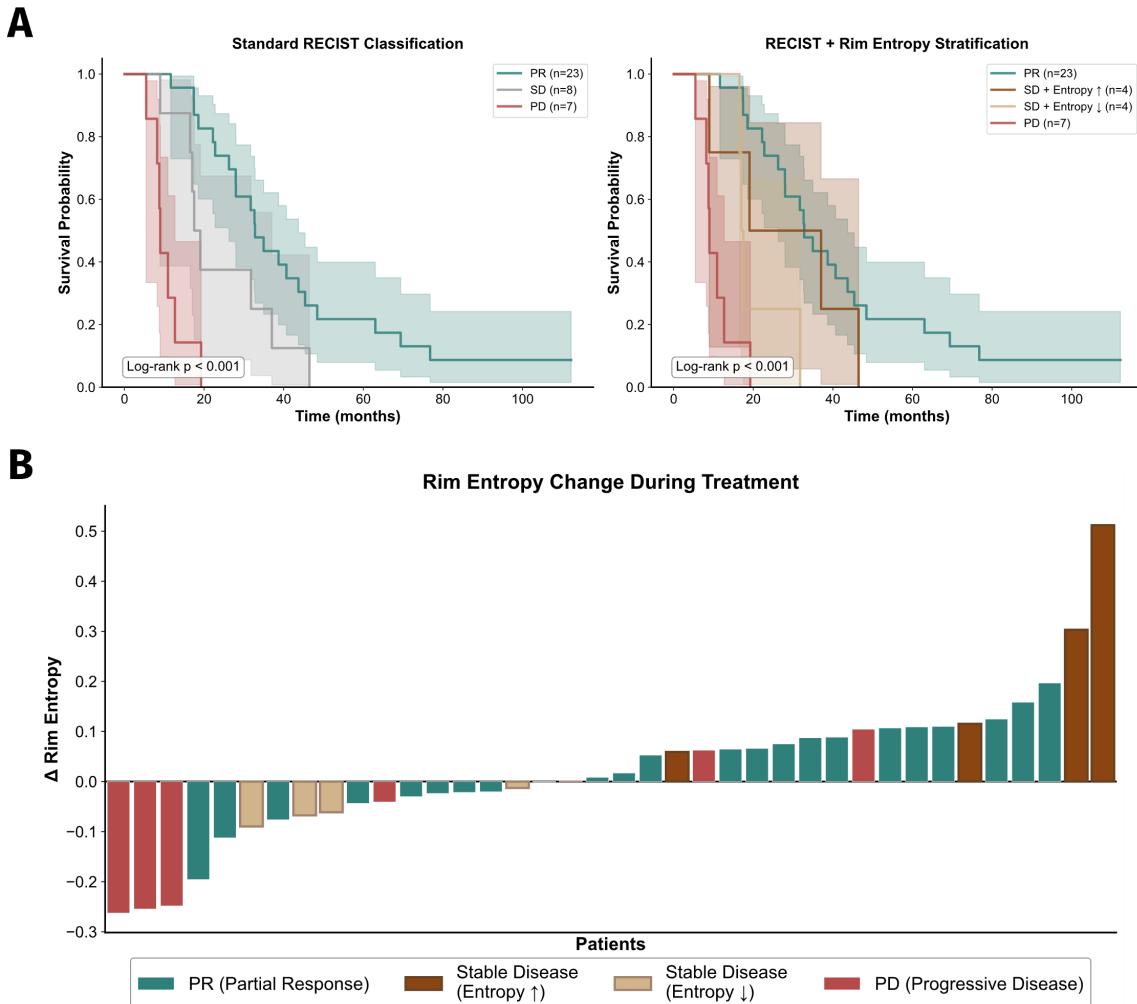


**Figure 8.5: Rim metrics differ between histopathological growth patterns.** Boxplots comparing rim habitat metrics between replacement (n=3, blue) and desmoplastic (n=3, salmon) tumors in the POEM cohort, annotated by an experienced pathologist. Desmoplastic tumors show higher rim entropy (Cohen's  $d = 1.92$ ) and higher rim avascular proportion ( $d = 2.03$ ), with lower rim cellular-perfused proportion ( $d = -1.44$ ). Rim vascular proportion did not differ ( $d = -0.37$ ). Despite the small sample, effect sizes are large, suggesting rim entropy may capture the fibrous stromal reaction characteristic of desmoplastic growth. Welch's t-test p-values shown; statistical significance limited by sample size.

$\Delta = -0.041$ , median OS = 9.0 months) (Appendix D).

The waterfall plot (Figure 8.6B) reveals the key finding: SD patients span the full range of entropy change. Four showed increasing entropy (among the highest in the cohort), four showed decreasing entropy. When substratified by entropy direction (Figure 8.6A, right), SD patients with increasing entropy tracked with PR, while those with decreasing entropy tracked closer to PD.

This analysis is exploratory and limited by sample size. With only 38 patients and 8 in the SD category, we lacked power to detect modest effects. The non-significant p-value does not exclude a true association; it reflects uncertainty. What the data do show is that rim entropy dynamics vary systematically with RECIST category and that the SD category contains biologically divergent subgroups. Whether rim entropy change can improve response classification requires validation in larger cohorts.



**Figure 8.6: Rim entropy dynamics during treatment.** (A) Standard RECIST classification (left) versus RECIST with rim entropy substratification (right). Left: PR, SD, and PD categories show separation ( $p<0.001$ ). Right: SD patients split by entropy direction—those with increasing entropy ( $n=4$ ) track with PR, while those with decreasing entropy ( $n=4$ ) track closer to PD. (B) Waterfall plot showing rim entropy change for each patient, sorted by magnitude and colored by RECIST category. Partial responders (dark green) cluster toward increasing entropy; progressive disease (red) clusters toward decreasing entropy. Stable disease patients span the full range: four show increasing entropy (dark brown, “SD + Entropy ↑”) representing potential hidden responders, while four show decreasing entropy (light brown, “SD + Entropy ↓”) with trajectories resembling progressors. This heterogeneity within the SD category—invisible to size-based assessment—suggests rim entropy captures treatment-tumor interactions that RECIST classification misses.

## 8.4 Discussion

In this Chapter we study whether CT habitats provide clinical value beyond tumor volume in colorectal liver metastases. We consistently found that habitat-derived heterogeneity metrics are prognostic in specific treatment contexts but not universally. In TCIA, post-treatment rim entropy predicted survival only after neoadjuvant chemotherapy ( $HR=0.59$ ) and not in treatment-naïve patients. In VHIO, baseline rim entropy was prognostic specifically with bevacizumab ( $HR=0.68$ ,  $p=0.004$ ) and in RAS-mutant patients ( $HR=0.80$ ,  $p=0.047$ ), who typically receive anti-angiogenic rather than anti-EGFR therapy. For chemotherapy alone, volume dominated with a large effect size ( $HR=5.46$ ) and habitat metrics added

## CLINICAL RELEVANCE OF CT HABITATS

nothing.

This pattern has a coherent interpretation. Cytotoxic chemotherapy produces measurable tumor shrinkage; response assessment based on size works. Anti-angiogenic therapy works differently: it stops blood supply, slows growth, and may induce compositional changes without substantial size reduction. In this context, spatial heterogeneity at the invasive margin captures treatment-relevant biology that volume cannot. The implication is that habitat analysis is not universally prognostic but specifically informative where size-based assessment is limited.

Across both cohorts, rim metrics consistently outperformed whole-tumor and core metrics. In TCIA, univariable associations were strongest for rim entropy ( $HR=0.59$ ), rim avascular proportion ( $HR=0.60$ ), and rim cellular-perfused proportion ( $HR=1.65$ ); core metrics showed no signal. In VHIO, rim entropy was the only habitat metric to survive multivariable adjustment in the bevacizumab subgroup. This gradient supports the biological premise that the invasive margin, where tumor meets liver, where angiogenesis and immune interactions concentrate, is the most clinically relevant compartment. This finding aligns with growing recognition in the literature that the tumor-liver interface carries distinct prognostic information [Nielsen et al., 2014]. Our results suggest that CT-derived rim heterogeneity may capture similar biology non-invasively.

Higher rim entropy predicted better survival across multiple contexts, which seems counterintuitive: shouldn't heterogeneous tumors do worse? One explanation is that rim entropy captures biological features associated with better prognosis. The POEM analysis suggests a specific hypothesis: rim entropy may reflect desmoplastic growth pattern. Desmoplastic tumors develop a fibrous capsule at the tumor-liver interface, creating a heterogeneous rim on CT; replacement tumors lack this capsule and show uniform rims. If rim entropy is indeed capturing desmoplasia the prognostic association would be explained [Fernández Moro et al., 2023]. In the post-treatment context, a different interpretation is possible: low entropy may suggest a tumor that is uniformly resistant, whereas high entropy can suggest response heterogeneity—a mixture of necrotic, fibrotic, and viable tissue.

The exploratory longitudinal analysis in 38 patients with paired imaging did not demonstrate statistically significant prediction of OS by rim entropy direction ( $p=0.76$ ). However, the pattern was consistent with other findings: partial responders showed increasing entropy (median OS 32.8 months), progressors showed decreasing entropy (median OS 9.0 months). Among patients classified as stable disease by RECIST, where clinical decisions are most uncertain, rim entropy trajectories varied substantially. Four showed the highest entropy increases in the cohort; four showed decreasing entropy. This heterogeneity suggests that rim entropy dynamics may capture treatment-tumor interactions invisible to size measurement, though larger studies are needed.

Regarding limitations, we should note that both cohorts are retrospective, and treatment was not randomized. The treatment-stratified and RAS-stratified analyses involve subgroups with reduced power. Both the desmoplastic growth pattern analysis ( $n=6$ ) and the longitudinal analysis ( $n=38$  with only 8 stable disease patients) are exploratory and require proper validation.

In conclusion, CT habitat analysis provides prognostic information beyond tumor volume, but this value depends on context. Rim heterogeneity matters in anti-angiogenic treatment contexts—post-neoadjuvant

in resectable disease, with bevacizumab or RAS-mutant status in unresectable disease. For cytotoxic chemotherapy alone, volume is enough.

## 8.5 Summary

In this Chapter we study whether CT habitats provide prognostic information beyond tumor volume in colorectal liver metastases. Using two independent cohorts representing resectable and unresectable disease, we found that the biologically-anchored CT habitat model developed in Chapter 7 translates into clinical signal where spatial heterogeneity captures treatment response that size-based metrics can't.

### Key Points:

- Prognostic value depends on context. Rim entropy predicted survival only in specific treatment contexts: after neoadjuvant chemotherapy in resectable disease ( $HR=0.59$ ,  $p<0.001$ ) and with bevacizumab or in RAS-mutant patients in unresectable disease ( $HR=0.68\text{--}0.80$ ,  $p<0.05$ ). For cytotoxic chemotherapy alone, tumor volume dominated ( $HR=5.46$ ) and habitat metrics added nothing.
- Prognostic signal concentrates at the invasive rim. Across both cohorts, rim metrics consistently outperformed whole-tumor and core metrics. The 2mm vascular rim—where tumor meets liver, where angiogenesis and treatment resistance concentrate—encodes clinically relevant information that whole-tumor averaging discards.
- Rim entropy may reflect histopathological growth pattern. Exploratory analysis in tumors with annotated growth patterns showed that desmoplastic tumors (better prognosis) had higher rim entropy than replacement tumors. If confirmed, CT habitat analysis could provide a non-invasive surrogate for growth pattern assessment.
- Longitudinal rim entropy change shows patterns consistent with treatment response. In patients with follow-up imaging available, the direction of rim entropy change did not significantly predict OS ( $p=0.76$ ), but the pattern was biologically coherent: responders showed increasing entropy (median OS 32.8 months), progressors showed decreasing entropy (median OS 9.0 months). Half of the stable disease (SD) showed increasing entropy and the other half decreasing. This heterogeneity within SD suggests rim entropy may capture treatment-tumor interactions that size-based assessment misses, though validation in larger cohorts is needed.





## Synthesis and Conclusions

*For now, at least, we need to wrestle with the grim realities of drug development, the inadequate animal models, our ignorance of the behavior of cellular regulatory circuitry, and the confounding biological complexities of human cancer. And most importantly, we must never give up. If our ancestors had, we would still be living in the Stone Age.*

— Robert Weinberg, *The Biology of Cancer*



# Chapter 9

## General Discussion

This thesis set out to determine whether routine contrast-enhanced CT imaging can capture biologically and clinically meaningful intratumor heterogeneity in colorectal liver metastases. The work was structured around five interconnected research questions, each building on the findings of the previous one. Taken together, the results support a qualified affirmative answer: CT can capture meaningful tumor heterogeneity, but only under specific conditions. In this chapter, the main findings of the thesis are synthesized, their biological and clinical implications are discussed, and directions for future research are outlined.

### 9.1 Synthesis of Findings

The **central hypothesis** of this thesis was that CT-derived habitats could capture biologically meaningful heterogeneity and provide clinically relevant information beyond tumor volume. This hypothesis was supported, with important qualifications: CT habitats primarily reflect vascular organization rather than discrete histological compartments, and their prognostic value depends on context.

- **RQ1: Which CT handcrafted radiomics features are repeatable and reproducible enough to support stable habitat computation? (Chapter 6)**

Only a minority of handcrafted features met repeatability and reproducibility criteria in liver lesions. Out of 91 features evaluated, 26 were sufficiently precise to support robust habitat computation. Feature stability was more sensitive to kernel radius than to intensity discretization, and the set of robust features identified in liver tumors differed from those identified in lung. These findings highlight the site-specific nature of radiomics precision and the need for task- and anatomy-specific validation.

- **RQ2: Which CT data representation produces habitats that best separate tissue with different cellularity and vascularity? (Chapter 7)**

## GENERAL DISCUSSION

Contrary to expectations, handcrafted texture features produced habitats with stronger biological coherence than deep learning embeddings derived from both a liver tumor segmentation model and a foundation model. Handcrafted features yielded more spatially contiguous clusters and greater separation on co-registered mpMRI metrics of cellularity and perfusion (ADC,  $K^{trans}$ , and fv).

- **RQ3: What tissue phenotypes do CT-derived habitats represent? (Chapter 7)**

Rather than mapping onto necrosis, fibrosis, and viable tumor, CT habitats consistently reflected gradients of vascular organization: avascular cores, transitional cellular–perfused regions, and vascularized rims at the tumor–liver interface. Histopathological assessment confirmed qualitative spatial correspondence between habitats and tissue composition, but the dominant signal was vascular rather than categorical histology.

- **RQ4: Do CT-derived habitats provide prognostic information independent of tumor volume? (Chapter 8)**

Habitat-derived metrics added prognostic information beyond tumor volume in specific clinical contexts. Rim entropy predicted survival after neoadjuvant chemotherapy in resectable disease and prior to anti-angiogenic therapy in unresectable disease. In contrast, with cytotoxic chemotherapy alone, tumor volume remained the dominant predictor.

- **RQ5: If prognostic information exists, is it spatially localized (i.e. at the core or the rim), or is it uniformly distributed within a tumor? (Chapter 8)**

Across cohorts and treatment settings, metrics derived from a 2 mm outer rim consistently outperformed whole-tumor and core-based metrics. Exploratory pathology-correlated analyses suggested that rim entropy may reflect histopathological growth patterns, with higher entropy observed in desmoplastic growth compared to replacement growth.

## 9.2 Biological Interpretation of CT-Derived Habitats

The findings of this thesis indicate that CT-derived habitats in colorectal liver metastases do not primarily represent discrete histological compartments such as necrosis, fibrosis, and viable tumor. Instead, they capture gradients of vascular organization and perfusion, with a consistent spatial structure across lesions. This observation is biologically plausible given the physics of contrast-enhanced CT, which is inherently sensitive to iodine distribution within blood vessels.

Across cohorts, habitats consistently delineated an avascular tumor core, a transitional cellular–perfused zone, and a vascularized rim at the tumor–liver interface. This spatial organization aligns with known aspects of tumor biology. The invasive margin is the region where angiogenesis occurs, where tumor

cells interact with host tissue, and where histopathological growth patterns are defined. In colorectal liver metastases, desmoplastic growth is characterized by a fibrotic rim separating tumor cells from liver parenchyma, whereas replacement growth involves direct infiltration of tumor cells along hepatic sinusoids without an intervening fibrotic barrier. The consistent concentration of prognostic information at the invasive rim suggests that CT-derived habitats may be sensitive to these growth-related vascular architectures. Tumors with heterogeneous rims may reflect mixed or unstable angiogenic patterns, variable perfusion, or coexistence of distinct growth behaviors within the same lesion. When heterogeneity is averaged across the entire tumor volume, this signal is diluted, explaining the superior performance of rim-based metrics compared to whole-tumor measures.

These findings also provide a framework for understanding the context-dependent prognostic value of habitat-derived metrics. Cytotoxic chemotherapy primarily induces tumor cell death, leading to volumetric shrinkage that is well captured by size-based criteria. Anti-angiogenic therapies, in contrast, target the tumor vasculature, often altering tissue organization and perfusion without producing immediate changes in tumor diameter. In this setting, heterogeneity in contrast enhancement at the invasive margin captures treatment-relevant biology that is invisible to volumetric assessment.

A recurring ambition in imaging research is the concept of a “virtual biopsy”, implying that non-invasive imaging might replace tissue sampling in the future. While appealing, this framing ignores the complementary nature of radiology and pathology. These modalities operate at different spatial scales and capture different aspects of tumor biology.

Histopathology provides molecular specificity and cellular detail, enabling assessment of genetic alterations, microarchitecture, and immune contexture. Imaging, in contrast, provides spatial context across the entire tumor and its surroundings, non-invasively and longitudinally. As of today, a habitat map cannot identify a BRAF mutation, but it can characterize vascular organization across the whole lesion at every timepoint. This information cannot be provided by a single biopsy due to sampling constraints. CT-derived habitats should therefore be understood as a distinct representation of tumor biology rather than an imperfect surrogate for histology. As illustrated in Figure 9.1, different representations of the same object emphasize different features without one being inherently superior. The clinical value of a model lies not in its fidelity to histology, but in its usefulness for decision-making.

## 9.3 Technical and Translational Considerations

Several limitations must be acknowledged when considering the translational potential of habitat imaging. First, all clinical analyses performed in this thesis were retrospective. Although associations between rim-based habitat metrics and outcome were consistent across cohorts and treatment contexts, prospective validation is required before these metrics can be used to inform clinical decision-making.

A further limitation concerns the treatment of multiple liver metastases. In this thesis, habitat-derived metrics were aggregated at the patient level using volume-weighted averages, implicitly assuming that larger tumors carry more clinically relevant information. While this is a reasonable first approximation, it does not account for dissociated responses, where some metastases may respond biologically despite

## GENERAL DISCUSSION

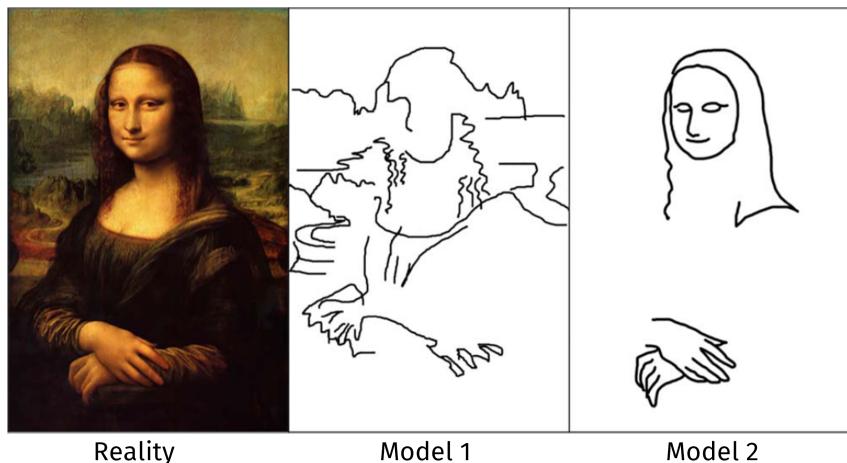
stable or increasing size in others. Before habitat imaging can be evaluated in clinical trials, it will be essential to study how *intertumor* heterogeneity (i.e. between tumors) should be combined, whether specific lesions dominate outcome, and how conflicting signals across metastases should be interpreted.

Moreover, habitat imaging relies on a multi-step computational pipeline involving tumor segmentation, feature extraction, and clustering. Each step introduces variability and design choices. One reason tumor size remains the dominant imaging biomarker in oncology is its simplicity: it is easy to measure and widely interpretable. In contrast, habitat imaging currently requires multiple parameters to be specified, which may limit robustness and reproducibility. An important byproduct of this thesis is the development of a complete, end-to-end CT habitat imaging pipeline implemented from first principles. Although methodological standardization was not an explicit objective, addressing the research questions posed here required constructing a transparent and reproducible framework. This represents the first fully described habitat imaging pipeline and constitutes an initial step toward standardization. Future work should aim to simplify this framework, establish robust defaults, and enable automated deployment suitable for clinical workflows.

Within this technical context, the choice of feature representation deserves particular attention. During the course of this thesis, expectations surrounding medical image analysis shifted rapidly with the adoption of deep learning and the emergence of foundation models [Paschali et al., 2025]. These have transformed representation learning in other fields such as pathology [Song et al., 2023] and therefore it was reasonable to expect that learned embeddings would outperform handcrafted radiomics features for habitat computation. In this task and with the datasets used in this thesis, however, handcrafted texture features produced more spatially coherent and biologically meaningful CT habitats than deep learning embeddings derived from pretrained segmentation and foundation models. This likely reflects constraints imposed by data size and task type: first, none of the evaluated models were optimized for spatial clustering or habitat discovery; second, in limited datasets, lower-dimensional and interpretable representations such as handcrafted features may also be less prone to overfitting.

These findings do not argue against deep learning, but rather emphasize that representation choice must be aware of both the task and the scale. Many clinically established imaging biomarkers, including tumor size and the multiparametric MRI maps used in this thesis, are handcrafted and remain valuable due to their interpretability and biological grounding. Learned representations may eventually surpass them, but such claims must be demonstrated empirically for each application.

Additional challenges include robustness across scanners and institutions, dependence on tumor segmentation accuracy, and extension of biological validation beyond qualitative histopathological correspondence. One promising approach to address this last challenge involves imaging resected liver metastases with ex vivo MRI and CT using preclinical scanners. Such datasets would enable accurate co-registration across pathology and radiology, providing a quantitative link between voxel-level imaging habitats and underlying tissue composition.



**Figure 9.1: Different representations capture different information.** Leonardo da Vinci's anatomical sketch and finished portrait of the Mona Lisa depict the same subject but emphasize different features. Similarly, histopathology and imaging habitats are both valid representations of tumor biology, suited to different purposes. Adapted from [Novikov et al., 2018].

## 9.4 Final Reflections

The global burden of cancer is projected to increase substantially over the coming decades. At the same time, advances in targeted therapies, immunotherapy, and supportive care raise the possibility that many cancers may increasingly be managed as chronic diseases rather than acute, fatal conditions. In this context, the need for robust, non-invasive, and informative biomarkers becomes even more urgent.

Medical imaging occupies a unique position in this landscape. It is routinely acquired, longitudinal, and captures the entire tumor and its microenvironment. As emphasized in recent perspectives on artificial intelligence in cancer research [Chang et al., 2025], the challenge is no longer the lack of data, but the ability to extract meaningful and clinically relevant information from complex datasets. This thesis argues that part of this challenge can be addressed by asking better questions of the imaging data already available. Habitat imaging seeks to move beyond the question of how large a tumor is, toward how it is organized and how it interacts with host tissue. Inevitably, this framework introduces new complexities and opens new methodological questions, which future PhD theses will need to address.

This work began with the image of a patient whose CT scan showed stable disease, offering little insight into whether her tumor was responding or adapting. One can envision a future in which intratumor heterogeneity assessment becomes a routine component of imaging interpretation, allowing radiologists to ask not only how large a tumor is, but how it is responding biologically. Narrowing the gap between what imaging captures and what clinicians need will require continued biological validation, technical rigor, and prospective testing, but the information is already present in the images we acquire every day.



# Chapter 10

## Conclusions

1. **Only a minority of handcrafted radiomics features are suitable for robust habitat computation.** Of the 91 handcrafted CT radiomics features evaluated, only 26 met repeatability and reproducibility criteria in liver lesions. Feature stability was strongly influenced by computational choices, with kernel radius having a greater impact than intensity discretization.
2. **Handcrafted texture features outperform deep learning embeddings for CT habitat computation.** Handcrafted features produced more spatially contiguous and biologically coherent habitats than embeddings from both a liver tumor segmentation model and a foundation model. The optimal representation depends on the downstream task, data type, and sample size, not model complexity alone.
3. **CT-derived habitats reflect vascular organization rather than discrete histological compartments.** While habitats were initially expected to distinguish necrosis, fibrosis, and viable tumor, they instead consistently mapped gradients of vascular organization: avascular cores, cellular-perfused regions, and vascularized rims.
4. **The prognostic value of CT habitat metrics depends on context.** Rim entropy was associated with survival after neoadjuvant chemotherapy and anti-angiogenic treatment, but not with cytotoxic chemotherapy alone, where tumor volume remained dominant. CT habitats may complement volumetric assessment when treatment effects are not captured by size change.
5. **Prognostic information is spatially concentrated at the invasive tumor rim.** Metrics derived from a 2 mm outer rim consistently outperformed whole-tumor and core-based measures for survival prediction. Rim heterogeneity may capture biologically relevant features related to histopathological growth patterns. Focusing analysis on the invasive margin, rather than averaging across the tumor, preserves this signal and improves prognostic performance.



# References

- Hugo J. W. L. Aerts, Emmanuel Rios Velazquez, Ralph T. H. Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, Frank Hoebers, Michelle M. Rietbergen, C. René Leemans, Andre Dekker, John Quackenbush, Robert J. Gillies, and Philippe Lambin. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5(1):4006, June 2014. ISSN 2041-1723. doi:[10.1038/ncomms5006](https://doi.org/10.1038/ncomms5006).
- American Cancer Society. Cancer Facts & Figures 2025. <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2025-cancer-facts-figures.html>, 2025.
- Jesper L.R. Andersson, Stefan Skare, and John Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, 20(2):870–888, October 2003. ISSN 10538119. doi:[10.1016/S1053-8119\(03\)00336-7](https://doi.org/10.1016/S1053-8119(03)00336-7).
- T. André, K.-K. Shiu, T.W. Kim, B.V. Jensen, L.H. Jensen, C.J.A. Punt, D. Smith, R. Garcia-Carbonero, J. Alcaide-Garcia, P. Gibbs, C. De La Fouchardiere, F. Rivera, E. Elez, D.T. Le, T. Yoshino, Y. Zuo, D. Fogelman, D. Adelberg, and L.A. Diaz. Pembrolizumab versus chemotherapy in microsatellite instability-high or mismatch repair-deficient metastatic colorectal cancer: 5-year follow-up from the randomized phase III KEYNOTE-177 study. *Annals of Oncology*, 36(3):277–284, March 2025. ISSN 09237534. doi:[10.1016/j.annonc.2024.11.012](https://doi.org/10.1016/j.annonc.2024.11.012).
- Melina Arnold, Mónica S Sierra, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*, 66(4):683–691, April 2017. ISSN 0017-5749, 1468-3288. doi:[10.1136/gutjnl-2015-310912](https://doi.org/10.1136/gutjnl-2015-310912).
- B Avants, C Epstein, M Grossman, and J Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, February 2008. ISSN 13618415. doi:[10.1016/j.media.2007.06.004](https://doi.org/10.1016/j.media.2007.06.004).
- Maria Balaguer-Montero, Adrià Marcos Morales, Marta Ligero, Christina Zatse, David Leiva, Luz M. Atlagich, Nikolaos Staikoglou, Cristina Viaplana, Camilo Monreal, Joaquin Mateo, Jorge Hernando, Alejandro García-Álvarez, Francesc Salvà, Jaume Capdevila, Elena Elez, Rodrigo Dienstmann, Elena Garralda, and Raquel Perez-Lopez. A CT-based deep learning-driven tool for automatic liver tumor detection and delineation in patients with cancer. *Cell Reports Medicine*, 6(4):102032, April 2025. ISSN 26663791. doi:[10.1016/j.xcrm.2025.102032](https://doi.org/10.1016/j.xcrm.2025.102032).
- Peter Bankhead, Maurice B. Loughrey, José A. Fernández, Yvonne Dombrowski, Darragh G. McArt, Philip D. Dunne, Stephen McQuaid, Ronan T. Gray, Liam J. Murray, Helen G. Coleman, Jacqueline A. James, Manuel Salto-Tellez, and Peter W. Hamilton. QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1):16878, December 2017. ISSN 2045-2322. doi:[10.1038/s41598-017-17204-5](https://doi.org/10.1038/s41598-017-17204-5).
- Shahla Bari, Marco Matejcic, Richard D. Kim, Hao Xie, Ibrahim H. Sahin, Benjamin D. Powers, Jamie K. Teer, Timothy A. Chan, Seth I. Felder, and Stephanie L. Schmit. Practice Patterns and Survival Outcomes of Immunotherapy for Metastatic Colorectal Cancer. *JAMA Network Open*, 8(3):e251186, March 2025. ISSN 2574-3805. doi:[10.1001/jamanetworkopen.2025.1186](https://doi.org/10.1001/jamanetworkopen.2025.1186).

## REFERENCES

- John J. Bartko. On various intraclass correlation reliability coefficients. *Psychological Bulletin*, 83(5):762–765, September 1976. ISSN 1939-1455, 0033-2909. doi:[10.1037/0033-2909.83.5.762](https://doi.org/10.1037/0033-2909.83.5.762).
- Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1):289–300, January 1995. ISSN 1369-7412, 1467-9868. doi:[10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x).
- Roberto Berenguer, María Del Rosario Pastor-Juan, Jesús Canales-Vázquez, Miguel Castro-García, María Victoria Villas, Francisco Mansilla Legorburu, and Sebastià Sabater. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. *Radiology*, 288(2):407–415, August 2018. ISSN 0033-8419, 1527-1315. doi:[10.1148/radiol.2018172361](https://doi.org/10.1148/radiol.2018172361).
- Kinga Bernatowicz, Francesco Grussu, Marta Ligero, Alonso Garcia, Eric Delgado, and Raquel Perez-Lopez. Robust imaging habitat computation using voxel-wise radiomics features. *Scientific Reports*, 11(1):20133, October 2021. ISSN 2045-2322. doi:[10.1038/s41598-021-99701-2](https://doi.org/10.1038/s41598-021-99701-2).
- Nigel C. Bird, David Mangnall, and Ali W. Majeed. Biology of colorectal liver metastases: A review. *Journal of Surgical Oncology*, 94(1):68–80, July 2006. ISSN 0022-4790, 1096-9098. doi:[10.1002/jso.20558](https://doi.org/10.1002/jso.20558).
- Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, 2006. ISBN 978-0-387-31073-2.
- Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, et al. On the Opportunities and Risks of Foundation Models, 2021.
- Freddie Bray, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Isabelle Soerjomataram, and Ahmedin Jemal. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3):229–263, May 2024. ISSN 0007-9235, 1542-4863. doi:[10.3322/caac.21834](https://doi.org/10.3322/caac.21834).
- Jan Budczies, Moritz von Winterfeld, Frederick Klauschen, Michael Bockmayr, Jochen K. Lennerz, Carsten Denkert, Thomas Wolf, Arne Warth, Manfred Dietel, Ioannis Anagnostopoulos, Wilko Weichert, Daniel Wittschieber, and Albrecht Stenzinger. The landscape of metastatic progression patterns across major human cancers. *Oncotarget*, 6(1):570–583, January 2015. ISSN 1949-2553. doi:[10.18632/oncotarget.2677](https://doi.org/10.18632/oncotarget.2677).
- Adrian Burton. RECIST: Right time to renovate? *The Lancet Oncology*, 8(6):464–465, June 2007. ISSN 14702045. doi:[10.1016/S1470-2045\(07\)70155-9](https://doi.org/10.1016/S1470-2045(07)70155-9).
- Richard A.D. Carano, Adrienne L. Ross, Jed Ross, Simon P. Williams, Hartmut Koeppen, Ralph H. Schwall, and Nicholas Van Bruggen. Quantification of tumor tissue populations by multispectral analysis. *Magnetic Resonance in Medicine*, 51(3):542–551, March 2004. ISSN 0740-3194, 1522-2594. doi:[10.1002/mrm.10731](https://doi.org/10.1002/mrm.10731).
- A. Cervantes, R. Adam, S. Roselló, D. Arnold, N. Normanno, J. Taïeb, J. Seligmann, T. De Baere, P. Osterlund, T. Yoshino, and E. Martinelli. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Annals of Oncology*, 34(1):10–32, January 2023. ISSN 09237534. doi:[10.1016/j.annonc.2022.10.003](https://doi.org/10.1016/j.annonc.2022.10.003).
- Christine L. Chaffer and Robert A. Weinberg. A Perspective on Cancer Cell Metastasis. *Science*, 331(6024):1559–1564, March 2011. ISSN 0036-8075, 1095-9203. doi:[10.1126/science.1203543](https://doi.org/10.1126/science.1203543).
- Hector H Li Chang, W Robert Leeper, Gabriel Chan, Douglas Quan, and David Kevin Driman. Infarct-like necrosis: A distinct form of necrosis seen in colorectal carcinoma liver metastases treated with perioperative chemotherapy. *Am. J. Surg. Pathol.*, 36(4):570–576, April 2012. ISSN 1532-0979. doi:[10.1097/PAS.0b013e31824057e7](https://doi.org/10.1097/PAS.0b013e31824057e7).

- Tian-Gen Chang, Seongyong Park, Alejandro A. Schäffer, Peng Jiang, and Eytan Ruppin. Hallmarks of artificial intelligence contributions to precision oncology. *Nature Cancer*, 6(3):417–431, March 2025. ISSN 2662-1347. doi:[10.1038/s43018-025-00917-2](https://doi.org/10.1038/s43018-025-00917-2).
- Baishali Chaudhury, Mu Zhou, Dmitry B. Goldgof, Lawrence O. Hall, Robert A. Gatenby, Robert J. Gillies, and Jennifer S. Drukesteinis. Identifying metastatic breast tumors using textural kinetic features of a contrast based habitat in DCE-MRI. In Lubomir M. Hadjiiski and Georgia D. Tourassi, editors, *SPIE Medical Imaging*, page 941415, Orlando, Florida, United States, March 2015. doi:[10.1117/12.2081386](https://doi.org/10.1117/12.2081386).
- Jianan Chen, Laurent Milot, Helen M. C. Cheung, and Anne L. Martel. Unsupervised Clustering of Quantitative Imaging Phenotypes using Autoencoder and Gaussian Mixture Model. 2019. doi:[10.48550/ARXIV.1909.02953](https://arxiv.org/abs/1909.02953).
- Lihua Chen, Min Liu, Jing Bao, Yunbao Xia, Jiuquan Zhang, Lin Zhang, Xuequan Huang, and Jian Wang. The Correlation between Apparent Diffusion Coefficient and Tumor Cellularity in Patients: A Meta-Analysis. *PLoS ONE*, 8(11):e79008, November 2013. ISSN 1932-6203. doi:[10.1371/journal.pone.0079008](https://doi.org/10.1371/journal.pone.0079008).
- Haesun Choi, Chuslip Charnsangavej, Silvana C. Faria, Homer A. Macapinlac, Michael A. Burgess, Shreyaskumar R. Patel, Lei L. Chen, Donald A. Podoloff, and Robert S. Benjamin. Correlation of Computed Tomography and Positron Emission Tomography in Patients With Metastatic Gastrointestinal Stromal Tumor Treated at a Single Institution With Imatinib Mesylate: Proposal of New Computed Tomography Response Criteria. *Journal of Clinical Oncology*, 25(13):1753–1759, May 2007. ISSN 0732-183X, 1527-7755. doi:[10.1200/JCO.2006.07.3049](https://doi.org/10.1200/JCO.2006.07.3049).
- Yun Shin Chun. Association of Computed Tomography Morphologic Criteria With Pathologic Response and Survival in Patients Treated With Bevacizumab for Colorectal Liver Metastases. *JAMA*, 302(21):2338, December 2009. ISSN 0098-7484. doi:[10.1001/jama.2009.1755](https://doi.org/10.1001/jama.2009.1755).
- Jacob Cohen. A power primer. *Psychological Bulletin*, 112(1):155–159, 1992. ISSN 1939-1455, 0033-2909. doi:[10.1037/0033-2909.112.1.155](https://doi.org/10.1037/0033-2909.112.1.155).
- D. R. Cox. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 34(2):187–202, January 1972. ISSN 1369-7412, 1467-9868. doi:[10.1111/j.2517-6161.1972.tb00899.x](https://doi.org/10.1111/j.2517-6161.1972.tb00899.x).
- Cameron Davidson-Pilon. Lifelines: Survival analysis in Python. *Journal of Open Source Software*, 4(40):1317, August 2019. ISSN 2475-9066. doi:[10.21105/joss.01317](https://doi.org/10.21105/joss.01317).
- Basar Demir, Lin Tian, Thomas Hastings Greer, Roland Kwitt, Francois-Xavier Vialard, Raul San Jose Estepar, Sylvain Bouix, Richard Jarrett Rushmore, Ebrahim Ebrahim, and Marc Niethammer. multiGradICON: A Foundation Model for Multimodal Medical Image Registration, February 2025.
- Aydin Demircioğlu. The effect of preprocessing filters on predictive performance in radiomics. *European Radiology Experimental*, 6(1):40, September 2022. ISSN 2509-9280. doi:[10.1186/s41747-022-00294-w](https://doi.org/10.1186/s41747-022-00294-w).
- Federica Di Nicolantonio, Pietro Paolo Vitiello, Silvia Marsoni, Salvatore Siena, Josep Tabernero, Livio Trusolino, Rene Bernards, and Alberto Bardelli. Precision oncology in metastatic colorectal cancer — from biology to medicine. *Nature Reviews Clinical Oncology*, 18(8):506–525, August 2021. ISSN 1759-4774, 1759-4782. doi:[10.1038/s41571-021-00495-z](https://doi.org/10.1038/s41571-021-00495-z).
- Mathew R. Divine, Prateek Katiyar, Ursula Kohlhofer, Leticia Quintanilla-Martinez, Bernd J. Pichler, and Jonathan A. Disselhorst. A Population-Based Gaussian Mixture Model Incorporating<sup>18</sup> F-FDG PET and Diffusion-Weighted MRI Quantifies Tumor Tissue Classes. *Journal of Nuclear Medicine*, 57(3):473–479, March 2016. ISSN 0161-5505, 2159-662X. doi:[10.2967/jnmed.115.163972](https://doi.org/10.2967/jnmed.115.163972).

## REFERENCES

- E.A. Eisenhauer, P. Therasse, J. Bogaerts, L.H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2): 228–247, January 2009. ISSN 09598049. doi:[10.1016/j.ejca.2008.10.026](https://doi.org/10.1016/j.ejca.2008.10.026).
- Jennie Engstrand, Henrik Nilsson, Cecilia Strömberg, Eduard Jonas, and Jacob Freedman. Colorectal cancer liver metastases – a population-based study on incidence, management and survival. *BMC Cancer*, 18(1):78, December 2018. ISSN 1471-2407. doi:[10.1186/s12885-017-3925-x](https://doi.org/10.1186/s12885-017-3925-x).
- Oscar Esteban, Christopher J. Markiewicz, Christopher Burns, et al. Nipy/nipype: 1.8.3. Zenodo, July 2022.
- James Ewing. *Neoplastic Diseases*. Philadelphia: W.B. Saunders Company., 1919.
- C. Federau, R. Meuli, K. O'Brien, P. Maeder, and P. Hagmann. Perfusion Measurement in Brain Gliomas with Intravoxel Incoherent Motion MRI. *American Journal of Neuroradiology*, 35(2):256–262, February 2014. ISSN 0195-6108, 1936-959X. doi:[10.3174/ajnr.A3686](https://doi.org/10.3174/ajnr.A3686).
- Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, John Buatti, Stephen Aylward, James V. Miller, Steve Pieper, and Ron Kikinis. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*, 30(9):1323–1341, November 2012. ISSN 0730725X. doi:[10.1016/j.mri.2012.05.001](https://doi.org/10.1016/j.mri.2012.05.001).
- Carlos Fernández Moro, Natalie Geyer, Sara Harrizi, Yousra Hamidi, Sara Söderqvist, Danyil Kuznyecov, Evelina Tidholm Qvist, Media Salmonson Schaad, Laura Hermann, Amanda Lindberg, Rainer L Heuchel, Alfonso Martín-Bernabé, Soniya Dhanjal, Anna C Navis, Christina Villard, Andrea C Del Valle, Lorand Bozóky, Ernesto Sparrelid, Luc Dirix, Carina Strell, Arne Östman, Bernhard Schnierer, Peter B Vermeulen, Jennie Engstrand, Béla Bozóky, and Marco Gerling. An idiosyncratic zonated stroma encapsulates desmoplastic liver metastases and originates from injured liver. *Nat. Commun.*, 14(1):5024, August 2023. ISSN 2041-1723. doi:[10.1038/s41467-023-40688-x](https://doi.org/10.1038/s41467-023-40688-x).
- Isabella Fornacon-Wood, Hitesh Mistry, Christoph J. Ackermann, Fiona Blackhall, Andrew McPartlin, Corinne Faivre-Finn, Gareth J. Price, and James P. B. O'Connor. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *European Radiology*, 30(11):6241–6250, November 2020. ISSN 0938-7994, 1432-1084. doi:[10.1007/s00330-020-06957-9](https://doi.org/10.1007/s00330-020-06957-9).
- Jean-Philippe Fortin, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylor, Irem Aselcioglu, Philip A. Cook, Phil Adams, Crystal Cooper, Maurizio Fava, Patrick J. McGrath, Melvin McInnis, Mary L. Phillips, Madhukar H. Trivedi, Myrna M. Weissman, and Russell T. Shinohara. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167:104–120, February 2018. ISSN 10538119. doi:[10.1016/j.neuroimage.2017.11.024](https://doi.org/10.1016/j.neuroimage.2017.11.024).
- James M. Franklin, Benjamin Irving, Bartlomiej W. Papiez, Jesper F. Kallehauge, Lai Mun Wang, Robert D. Goldin, Adrian L. Harris, Ewan M. Anderson, Julia A. Schnabel, Michael A. Chappell, Michael Brady, Ricky A. Sharma, and Fergus V. Gleeson. Tumour subregion analysis of colorectal liver metastases using semi-automated clustering based on DCE-MRI: Comparison with histological subregions and impact on pharmacokinetic parameter analysis. *European Journal of Radiology*, 126:108934, May 2020. ISSN 0720048X. doi:[10.1016/j.ejrad.2020.108934](https://doi.org/10.1016/j.ejrad.2020.108934).
- Sophia Frentzas, Eve Simoneau, Victoria L Bridgeman, Peter B Vermeulen, Shane Foo, Eleftherios Kostaras, Mark R Nathan, Andrew Wotherspoon, Zu-hua Gao, Yu Shi, Gert Van Den Eynden, Frances Daley, Clare Peckitt, Xianming Tan, Ayat Salman, Anthoula Lazaris, Patrycja Gazinska, Tracy J Berg, Zak Eltahir, Laila

- Ritsma, Jacco Van Rheenen, Alla Khashper, Gina Brown, Hanna Nyström, Malin Sund, Steven Van Laere, Evelyne Loyer, Luc Dirix, David Cunningham, Peter Metrakos, and Andrew R Reynolds. Vessel co-option mediates resistance to anti-angiogenic therapy in liver metastases. *Nature Medicine*, 22(11):1294–1302, November 2016. ISSN 1078-8956, 1546-170X. doi:[10.1038/nm.4197](https://doi.org/10.1038/nm.4197).
- Robert A. Gatenby, Olya Grove, and Robert J. Gillies. Quantitative Imaging in Cancer Evolution and Ecology. *Radiology*, 269(1):8–14, October 2013. ISSN 0033-8419, 1527-1315. doi:[10.1148/radiol.13122697](https://doi.org/10.1148/radiol.13122697).
- Mirjam Gerwing, Ken Herrmann, Anne Helfen, Christoph Schliemann, Wolfgang E. Berdel, Michel Eisenblätter, and Moritz Wildgruber. The beginning of the end for conventional RECIST — novel therapies require novel imaging approaches. *Nature Reviews Clinical Oncology*, 16(7):442–458, July 2019. ISSN 1759-4774, 1759-4782. doi:[10.1038/s41571-019-0169-5](https://doi.org/10.1038/s41571-019-0169-5).
- Richard Goldberg and Peter Kirkpatrick. Cetuximab. *Nature Reviews Drug Discovery*, 4(5):S10–S11, May 2005. ISSN 1474-1776, 1474-1784. doi:[10.1038/nrd1728](https://doi.org/10.1038/nrd1728).
- Ian Goodfellow, Aaron Courville, and Yoshua Bengio. *Deep Learning*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, Massachusetts, 2016. ISBN 978-0-262-33737-3.
- Francesco Grussu, Athanasios Grigoriou, Kinga Bernatowicz, Marco Palombo, Irene Casanova-Salas, Daniel Navarro-Garcia, Ignasi Barba, Sara Simonetti, Garazi Serna, Carlos Macarro, Anna Kira Voronova, Valezka Garay, Juan Francisco Corral, Marta Vidorreta, Pablo García-Polo García, Xavier Merino, Richard Mast, Núria Rosón, Manuel Escobar, Maria Vieito, Rodrigo Toledo, Paolo Nuciforo, Joaquin Mateo, Elena Garralda, and Raquel Perez-Lopez. Clinically feasible liver tumour cell size measurement through histology-informed *in vivo* diffusion MRI. *Communications Medicine*, November 2025. ISSN 2730-664X. doi:[10.1038/s43856-025-01246-2](https://doi.org/10.1038/s43856-025-01246-2).
- H. Gupta, M.P. Deek, T.R. McNutt, J. Lee, H. Quon, and K. Sheikh. Predicting Progression-Free Survival Using Primary and Nodal Radiomic Features in Head and Neck Cancer. *International Journal of Radiation Oncology\*Biology\*Physics*, 111(3):e130–e131, November 2021. ISSN 03603016. doi:[10.1016/j.ijropbp.2021.07.563](https://doi.org/10.1016/j.ijropbp.2021.07.563).
- Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, November 1973. ISSN 0018-9472, 2168-2909. doi:[10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314).
- Frank E. Harrell, Kerry L. Lee, and Daniel B. Mark. MULTIVARIABLE PROGNOSTIC MODELS: ISSUES IN DEVELOPING MODELS, EVALUATING ASSUMPTIONS AND ADEQUACY, AND MEASURING AND REDUCING ERRORS. *Statistics in Medicine*, 15(4):361–387, February 1996. ISSN 02776715, 10970258. doi:[10.1002/\(SICI\)1097-0258\(19960229\)15:4<361::AID-SIM168>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4).
- Michael P. Hartung, Adam Brown, and Michael P. Hartung and Adam Brown. Abdominal CT: Phases. <https://litfl.com/abdominal-ct-phases/>, February 2024.
- Erica C Henning, Chieko Azuma, Christopher H Sotak, and Karl G Helmer. Multispectral quantification of tissue types in a RIF-1 tumor model with histological validation. Part I. *Magn. Reson. Med.*, 57(3):501–512, March 2007. ISSN 0740-3194. doi:[10.1002/mrm.21161](https://doi.org/10.1002/mrm.21161).
- Kenneth R. Hess, Gauri R. Varadhachary, Sarah H. Taylor, Wei Wei, Martin N. Raber, Renato Lenzi, and James L. Abbruzzese. Metastatic patterns in adenocarcinoma. *Cancer*, 106(7):1624–1633, April 2006. ISSN 0008-543X, 1097-0142. doi:[10.1002/cncr.21778](https://doi.org/10.1002/cncr.21778).
- Niek Hugen, Gina Brown, Robert Glynne-Jones, Johannes H. W. De Wilt, and Iris D. Nagtegaal. Advances in the care of patients with mucinous colorectal cancer. *Nature Reviews Clinical Oncology*, 13(6):361–369, June 2016. ISSN 1759-4774, 1759-4782. doi:[10.1038/nrclinonc.2015.140](https://doi.org/10.1038/nrclinonc.2015.140).

## REFERENCES

- Alan Jackson, James P.B. O'Connor, Geoff J.M. Parker, and Gordon C. Jayson. Imaging Tumor Vascular Heterogeneity and Angiogenesis using Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *Clinical Cancer Research*, 13(12):3449–3459, June 2007. ISSN 1078-0432, 1557-3265. doi:[10.1158/1078-0432.CCR-07-0238](https://doi.org/10.1158/1078-0432.CCR-07-0238).
- Bruna V. Jardim-Perassi, Suning Huang, William Dominguez-Viqueira, Jan Poleszczuk, Mikalai M. Budzevich, Mahmoud A. Abdalah, Smitha R. Pillai, Epifanio Ruiz, Marilyn M. Bui, Debora A.P.C. Zuccari, Robert J. Gillies, and Gary V. Martinez. Multiparametric MRI and Coregistered Histology Identify Tumor Habitats in Breast Cancer Mouse Models. *Cancer Research*, 79(15):3952–3964, August 2019. ISSN 0008-5472, 1538-7445. doi:[10.1158/0008-5472.CAN-19-0213](https://doi.org/10.1158/0008-5472.CAN-19-0213).
- A. K. Jha, S. Mithun, V. Jaiswar, U. B. Sherkhane, N. C. Purandare, K. Prabhash, V. Rangarajan, A. Dekker, L. Wee, and A. Traverso. Repeatability and reproducibility study of radiomic features on a phantom and human cohort. *Scientific Reports*, 11(1):2055, January 2021. ISSN 2045-2322. doi:[10.1038/s41598-021-81526-8](https://doi.org/10.1038/s41598-021-81526-8).
- Javier Juan-Albarracín, Elies Fuster-Garcia, Alexandre Pérez-Girbés, Fernando Aparici-Robles, Ángel Alberich-Bayarri, Antonio Revert-Ventura, Luis Martí-Bonmatí, and Juan M. García-Gómez. Glioblastoma: Vascular Habitats Detected at Preoperative Dynamic Susceptibility-weighted Contrast-enhanced Perfusion MR Imaging Predict Survival. *Radiology*, 287(3):944–954, June 2018. ISSN 0033-8419, 1527-1315. doi:[10.1148/radiol.2017170845](https://doi.org/10.1148/radiol.2017170845).
- E. L. Kaplan and Paul Meier. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282):457–481, June 1958. ISSN 0162-1459, 1537-274X. doi:[10.1080/01621459.1958.10501452](https://doi.org/10.1080/01621459.1958.10501452).
- Prateek Katiyar, Mathew R. Divine, Ursula Kohlhofer, Leticia Quintanilla-Martinez, Bernhard Schölkopf, Bernd J. Pichler, and Jonathan A. Disselhorst. A Novel Unsupervised Segmentation Approach Quantifies Tumor Tissue Populations Using Multiparametric MRI: First Results with Histological Validation. *Molecular Imaging and Biology*, 19(3):391–397, June 2017. ISSN 1860-2002. doi:[10.1007/s11307-016-1009-y](https://doi.org/10.1007/s11307-016-1009-y).
- Prateek Katiyar, Johannes Schwenck, Leonie Frauenfeld, Mathew R. Divine, Vaibhav Agrawal, Ursula Kohlhofer, Sergios Gatidis, Roland Kontermann, Alfred Königsrainer, Leticia Quintanilla-Martinez, Christian La Fougère, Bernhard Schölkopf, Bernd J. Pichler, and Jonathan A. Disselhorst. Quantification of intratumoural heterogeneity in mice and patients via machine-learning models trained on PET–MRI data. *Nature Biomedical Engineering*, 7(8):1014–1027, June 2023. ISSN 2157-846X. doi:[10.1038/s41551-023-01047-9](https://doi.org/10.1038/s41551-023-01047-9).
- Yoshiki Kawata, Noboru Niki, Hironobu Ohmatsu, Masahiko Kusumoto, Takaaki Tsuchida, Kenji Eguchi, Masahiro Kaneko, and Noriyuki Moriyama. Quantitative classification based on CT histogram analysis of non-small cell lung cancer: Correlation with histopathological characteristics and recurrence-free survival: Quantitative classification based on CT histogram analysis. *Medical Physics*, 39(2):988–1000, February 2012. ISSN 00942405. doi:[10.1118/1.3679017](https://doi.org/10.1118/1.3679017).
- Elias Kellner, Bibek Dhital, Valerij G. Kiselev, and Marco Reisert. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic Resonance in Medicine*, 76(5):1574–1581, November 2016. ISSN 0740-3194, 1522-2594. doi:[10.1002/mrm.26054](https://doi.org/10.1002/mrm.26054).
- M. G. Kendall and B. Babington Smith. The Problem of \$\\$ Rankings. *The Annals of Mathematical Statistics*, 10 (3):275–287, September 1939. ISSN 0003-4851. doi:[10.1214/aoms/1177732186](https://doi.org/10.1214/aoms/1177732186).
- Larry G Kessler, Huiman X Barnhart, Andrew J Buckler, Kingshuk Roy Choudhury, Marina V Kondratovich, Alicia Toledano, Alexander R Guimaraes, Ross Filice, Zheng Zhang, Daniel C Sullivan, and QIBA Terminology Working Group. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Statistical Methods in Medical Research*, 24(1):9–26, February 2015. ISSN 0962-2802, 1477-0334. doi:[10.1177/0962280214537333](https://doi.org/10.1177/0962280214537333).

## REFERENCES

- Terry K. Koo and Mae Y. Li. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2):155–163, June 2016. ISSN 15563707. doi:[10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012).
- Scott Kopetz, Danielle A. Murphy, Jie Pu, Fortunato Ciardiello, Jayesh Desai, Eric Van Cutsem, Harpreet Singh Wasan, Takayuki Yoshino, Hedieh Saffari, Xiaosong Zhang, Phineas Hamilton, Tao Xie, Rona Yaeger, and Josep Tabernero. Molecular profiling of BRAF-V600E-mutant metastatic colorectal cancer in the phase 3 BEACON CRC trial. *Nature Medicine*, 30(11):3261–3271, November 2024. ISSN 1078-8956, 1546-170X. doi:[10.1038/s41591-024-03235-9](https://doi.org/10.1038/s41591-024-03235-9).
- Michael I. Koukourakis, Ioannis Mavani, George Kouklakis, Michael Pitiakoudis, George Minopoulos, Costantinos Manolas, and Costantinos Simopoulos. Early Antivascular Effects of Bevacizumab Anti-VEGF Monoclonal Antibody on Colorectal Carcinomas Assessed With Functional CT Imaging. *American Journal of Clinical Oncology*, 30(3):315–318, June 2007. ISSN 0277-3732. doi:[10.1097/01.coc.0000258119.90805.ca](https://doi.org/10.1097/01.coc.0000258119.90805.ca).
- H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2): 83–97, March 1955. ISSN 0028-1441, 1931-9193. doi:[10.1002/nav.3800020109](https://doi.org/10.1002/nav.3800020109).
- Paras Lakhani and Baskaran Sundaram. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*, 284(2):574–582, August 2017. ISSN 0033-8419, 1527-1315. doi:[10.1148/radiol.2017162326](https://doi.org/10.1148/radiol.2017162326).
- Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud G.P.M. Van Stiphout, Patrick Granton, Catharina M.L. Zegers, Robert Gillies, Ronald Boellard, André Dekker, and Hugo J.W.L. Aerts. Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4):441–446, March 2012. ISSN 09598049. doi:[10.1016/j.ejca.2011.11.036](https://doi.org/10.1016/j.ejca.2011.11.036).
- Philippe Lambin, Ralph T.H. Leijenaar, Timo M. Deist, Jurgen Peerlings, Evelyn E.C. De Jong, Janita Van Timmeren, Sebastian Sanduleanu, Ruben T.H.M. Larue, Aniek J.G. Even, Arthur Jochems, Yvonka Van Wijk, Henry Woodruff, Johan Van Soest, Tim Lustberg, Erik Roelofs, Wouter Van Elmpt, Andre Dekker, Felix M. Mottaghy, Joachim E. Wildberger, and Sean Walsh. Radiomics: The bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology*, 14(12):749–762, December 2017. ISSN 1759-4774, 1759-4782. doi:[10.1038/nrclinonc.2017.141](https://doi.org/10.1038/nrclinonc.2017.141).
- E. Latacz, O. Prior Palomares, N. Ruiz Roig, I. Puig Povedano, K. Mils, F. Salvà Ballabrera, I. Joye, S. Declercq, L. Dirix, F. Deckers, P. Cools, P. Boons, T. Apers, P. Dirix, R. Perez-Lopez, D. Garcia Mollevi, B. Gypen, P. Willemse, L. LLado, and P. Vermeulen. Prospective complete histopathological characterization of liver metastases from colorectal and breast carcinoma to predict the histopathological growth patterns by medical imaging (POEM). *European Journal of Surgical Oncology*, 50:109241, December 2024. ISSN 07487983. doi:[10.1016/j.ejsco.2024.109241](https://doi.org/10.1016/j.ejsco.2024.109241).
- Emily Latacz, Pieter-Jan Van Dam, Christian Vanhove, Laura Llado, Benedicte Descamps, Núria Ruiz, Ines Joye, Dirk Grünhagen, Steven Van Laere, Piet Dirix, David G. Mollevi, Cornelis Verhoef, Luc Dirix, and Peter Vermeulen. Can medical imaging identify the histopathological growth patterns of liver metastases? *Seminars in Cancer Biology*, 71:33–41, June 2021. ISSN 1044579X. doi:[10.1016/j.semcan.2020.07.002](https://doi.org/10.1016/j.semcan.2020.07.002).
- Emily Latacz, Diederik Höppener, Ali Bohlok, et al. Histopathological growth patterns of liver metastasis: Updated consensus guidelines for pattern scoring, perspectives and recent mechanistic insights. *Br. J. Cancer*, 127(6): 988–1013, October 2022. ISSN 0007-0920. doi:[10.1038/s41416-022-01859-7](https://doi.org/10.1038/s41416-022-01859-7).
- P. C. Lauterbur. Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, 242(5394):190–191, March 1973. ISSN 0028-0836, 1476-4687. doi:[10.1038/242190a0](https://doi.org/10.1038/242190a0).

## REFERENCES

- D Le Bihan, E Breton, D Lallemand, M L Aubin, J Vignaud, and M Laval-Jeantet. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*, 168(2):497–505, August 1988. ISSN 0033-8419, 1527-1315. doi:[10.1148/radiology.168.2.3393671](https://doi.org/10.1148/radiology.168.2.3393671).
- Denis Le Bihan, Jean-François Mangin, Cyril Poupon, Chris A. Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat. Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4):534–546, April 2001. ISSN 1053-1807, 1522-2586. doi:[10.1002/jmri.1076](https://doi.org/10.1002/jmri.1076).
- Breton E Le Bihan D. Imagerie de diffusion in vivo par résonance magnétique nucléaire. *C R Acad Sci Paris.*, (301:1109–1112), 1985.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. ISSN 0028-0836, 1476-4687. doi:[10.1038/nature14539](https://doi.org/10.1038/nature14539).
- Shaolei Li, Yongming Dai, Jiayi Chen, Fuhua Yan, and Yingli Yang. MRI-based habitat imaging in cancer treatment: Current technology, applications, and challenges. *Cancer Imaging*, 24(1), August 2024. ISSN 1470-7330. doi:[10.1186/s40644-024-00758-9](https://doi.org/10.1186/s40644-024-00758-9).
- Xiangrui Li, Paul S. Morgan, John Ashburner, Jolinda Smith, and Christopher Rorden. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, 264:47–56, May 2016. ISSN 01650270. doi:[10.1016/j.jneumeth.2016.03.001](https://doi.org/10.1016/j.jneumeth.2016.03.001).
- Gwilym S. Lodwick, Theodore E. Keats, and John P. Dorst. The Coding of Roentgen Images for Computer Analysis as Applied to Lung Cancer. *Radiology*, 81(2):185–200, August 1963. ISSN 0033-8419, 1527-1315. doi:[10.1148/81.2.185](https://doi.org/10.1148/81.2.185).
- F Loupakis, M Schirripa, C Caparello, N Funel, L Pollina, E Vasile, C Cremolini, L Salvatore, M Morvillo, C Antoniotti, F Marmorino, G Masi, and A Falcone. Histopathologic evaluation of liver metastases from colorectal cancer in patients treated with FOLFOXIRI plus bevacizumab. *Br. J. Cancer*, 108(12):2549–2556, June 2013. ISSN 0007-0920. doi:[10.1038/bjc.2013.245](https://doi.org/10.1038/bjc.2013.245).
- Fotios Loupakis, Chiara Cremolini, Gianluca Masi, Sara Lonardi, Vittorina Zagonel, Lisa Salvatore, Enrico Cortesi, Gianluca Tomasello, Monica Ronzoni, Rosella Spadi, Alberto Zaniboni, Giuseppe Tonini, Angela Buondonna, Domenico Amoroso, Silvana Chiara, Chiara Carlomagno, Corrado Boni, Giacomo Allegrini, Luca Boni, and Alfredo Falcone. Initial Therapy with FOLFOXIRI and Bevacizumab for Metastatic Colorectal Cancer. *New England Journal of Medicine*, 371(17):1609–1618, October 2014. ISSN 0028-4793, 1533-4406. doi:[10.1056/NEJMoa1403108](https://doi.org/10.1056/NEJMoa1403108).
- B. Maher, E. Ryan, M. Little, P. Boardman, and B. Stedman. The management of colorectal liver metastases. *Clinical Radiology*, 72(8):617–625, August 2017. ISSN 00099260. doi:[10.1016/j.crad.2017.05.016](https://doi.org/10.1016/j.crad.2017.05.016).
- H. B. Mann and D. R. Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1):50–60, March 1947. ISSN 0003-4851. doi:[10.1214/aoms/1177730491](https://doi.org/10.1214/aoms/1177730491).
- Kenneth A. Miles, Balaji Ganeshan, Matthew R. Griffiths, Rupert C. D. Young, and Christopher R. Chatwin. Colorectal Cancer: Texture Analysis of Portal Phase Hepatic CT Images as a Potential Marker of Survival. *Radiology*, 250(2):444–452, February 2009. ISSN 0033-8419, 1527-1315. doi:[10.1148/radiol.2502071879](https://doi.org/10.1148/radiol.2502071879).
- Eileen Morgan, Melina Arnold, A Gini, V Lorenzoni, C J Cabasag, Mathieu Laversanne, Jerome Vignat, Jacques Ferlay, Neil Murphy, and Freddie Bray. Global burden of colorectal cancer in 2020 and 2040: Incidence and mortality estimates from GLOBOCAN. *Gut*, 72(2):338–344, February 2023. ISSN 0017-5749, 1468-3288. doi:[10.1136/gutjnl-2022-327736](https://doi.org/10.1136/gutjnl-2022-327736).

## REFERENCES

- Van K. Morris, Erin B. Kennedy, Nancy N. Baxter, Al B. Benson, Andrea Cercek, May Cho, Kristen K. Ciombor, Chiara Cremolini, Anjee Davis, Dustin A. Deming, Marwan G. Fakih, Sepideh Gholami, Theodore S. Hong, Ishmael Jaiyesimi, Kelsey Klute, Christopher Lieu, Hanna Sanoff, John H. Strickler, Sarah White, Jason A. Willis, and Cathy Eng. Treatment of Metastatic Colorectal Cancer: ASCO Guideline. *Journal of Clinical Oncology*, 41(3):678–700, January 2023. ISSN 0732-183X, 1527-7755. doi:[10.1200/JCO.22.01690](https://doi.org/10.1200/JCO.22.01690).
- Margherita Mottola, Stephan Ursprung, Leonardo Rundo, Lorena Escudero Sanchez, Tobias Klatte, Iosif Men-dichovszky, Grant D Stewart, Evis Sala, and Alessandro Bevilacqua. Reproducibility of CT-based radiomic features against image resampling and perturbations for tumour and healthy kidney in renal cancer patients. *Scientific Reports*, 11(1):11542, June 2021. ISSN 2045-2322. doi:[10.1038/s41598-021-90985-y](https://doi.org/10.1038/s41598-021-90985-y).
- National Electrical Manufacturers Association (NEMA). *Digital Imaging and Communications in Medicine (DICOM) Standard*. National Electrical Manufacturers Association (NEMA), Rosslyn, VA, USA, 2025.
- Francesca Ng, Robert Kozarski, Balaji Ganeshan, and Vicky Goh. Assessment of tumor heterogeneity by CT texture analysis: Can the largest cross-sectional area be used as an alternative to whole tumor analysis? *European Journal of Radiology*, 82(2):342–348, February 2013. ISSN 0720048X. doi:[10.1016/j.ejrad.2012.10.023](https://doi.org/10.1016/j.ejrad.2012.10.023).
- Kåre Nielsen, Hans C Rolff, Rikke L Eefsen, and Ben Vainer. The morphological growth patterns of colorectal liver metastases are prognostic for overall survival. *Modern Pathology*, 27(12):1641–1648, December 2014. ISSN 08933952. doi:[10.1038/modpathol.2014.4](https://doi.org/10.1038/modpathol.2014.4).
- Mizuki Nishino. Tumor Response Assessment for Precision Cancer Therapy: Response Evaluation Criteria in Solid Tumors and Beyond. *American Society of Clinical Oncology Educational Book*, (38):1019–1029, May 2018. ISSN 1548-8748, 1548-8756. doi:[10.1200/EDBK\\_201441](https://doi.org/10.1200/EDBK_201441).
- Dmitry S. Novikov, Valerij G. Kiselev, and Sune N. Jespersen. On modeling. *Magnetic Resonance in Medicine*, 79(6):3172–3193, June 2018. ISSN 0740-3194, 1522-2594. doi:[10.1002/mrm.27101](https://doi.org/10.1002/mrm.27101).
- James P. B. O'Connor, Eric O. Aboagye, Judith E. Adams, Hugo J. W. L. Aerts, Sally F. Barrington, Ambros J. Beer, Ronald Boellaard, Sarah E. Bohndiek, Michael Brady, Gina Brown, David L. Buckley, Thomas L. Chenevert, Laurence P. Clarke, Sandra Collette, Gary J. Cook, Nandita M. deSouza, John C. Dickson, Caroline Dive, Jeffrey L. Evelhoch, Corinne Faivre-Finn, Ferdia A. Gallagher, Fiona J. Gilbert, Robert J. Gillies, Vicky Goh, John R. Griffiths, Ashley M. Groves, Steve Halligan, Adrian L. Harris, David J. Hawkes, Otto S. Hoekstra, Erich P. Huang, Brian F. Hutton, Edward F. Jackson, Gordon C. Jayson, Andrew Jones, Dow-Mu Koh, Denis La-combe, Philippe Lambin, Nathalie Lassau, Martin O. Leach, Ting-Yim Lee, Edward L. Leen, Jason S. Lewis, Yan Liu, Mark F. Lythgoe, Prakash Manoharan, Ross J. Maxwell, Kenneth A. Miles, Bruno Morgan, Steve Morris, Tony Ng, Anwar R. Padhani, Geoff J. M. Parker, Mike Partridge, Arvind P. Pathak, Andrew C. Peet, Shonit Pun-wani, Andrew R. Reynolds, Simon P. Robinson, Lalitha K. Shankar, Ricky A. Sharma, Dmitry Soloviev, Sigrid Stroobants, Daniel C. Sullivan, Stuart A. Taylor, Paul S. Tofts, Gillian M. Tozer, Marcel Van Herk, Simon Walker-Samuel, James Wason, Kaye J. Williams, Paul Workman, Thomas E. Yankelev, Kevin M. Brindle, Lisa M. McShane, Alan Jackson, and John C. Waterton. Imaging biomarker roadmap for cancer studies. *Nature Reviews Clinical Oncology*, 14(3):169–186, March 2017. ISSN 1759-4774, 1759-4782. doi:[10.1038/nrclinonc.2016.162](https://doi.org/10.1038/nrclinonc.2016.162).
- James P.B. O'Connor. Cancer heterogeneity and imaging. *Seminars in Cell & Developmental Biology*, 64:48–57, April 2017. ISSN 1084-9521. doi:[10.1016/j.semcdcb.2016.10.001](https://doi.org/10.1016/j.semcdcb.2016.10.001).
- James Pb O'Connor, Alan Jackson, Marie-Claude Asselin, David L Buckley, Geoff Jm Parker, and Gordon C Jayson. Quantitative imaging biomarkers in the clinical development of targeted therapeutics: Current and future perspectives. *The Lancet Oncology*, 9(8):766–776, August 2008. ISSN 1470-2045. doi:[10.1016/s1470-2045\(08\)70196-7](https://doi.org/10.1016/s1470-2045(08)70196-7).

## REFERENCES

- James P.B. O'Connor, Chris J. Rose, John C. Waterton, Richard A.D. Carano, Geoff J.M. Parker, and Alan Jackson. Imaging Intratumor Heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome. *Clinical Cancer Research*, 21(2):249–257, January 2015. ISSN 1078-0432, 1557-3265. doi:[10.1158/1078-0432.ccr-14-0990](https://doi.org/10.1158/1078-0432.ccr-14-0990).
- S Ourselin, A Roche, G Subsol, X Pennec, and N Ayache. Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing*, 19(1-2):25–31, January 2001. ISSN 02628856. doi:[10.1016/S0262-8856\(00\)00052-4](https://doi.org/10.1016/S0262-8856(00)00052-4).
- Kumi Ozaki, Shohei Higuchi, Hirohiko Kimura, and Toshifumi Gabata. Liver Metastases: Correlation between Imaging Features and Pathomolecular Environments. *RadioGraphics*, 42(7):1994–2013, November 2022. ISSN 0271-5333, 1527-1323. doi:[10.1148/rg.220056](https://doi.org/10.1148/rg.220056).
- S. Paget. The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Reviews*, 8(2): 98–101, 1889. ISSN 0167-7659.
- Suraj Pai, Dennis Bontempi, Ibrahim Hadzic, Vasco Prudente, Mateo Sokač, Tafadzwa L. Chaunzwa, Simon Bernatz, Ahmed Hosny, Raymond H. Mak, Nicolai J. Birkbak, and Hugo J. W. L. Aerts. Foundation model for cancer imaging biomarkers. *Nature Machine Intelligence*, 6(3):354–367, March 2024. ISSN 2522-5839. doi:[10.1038/s42256-024-00807-9](https://doi.org/10.1038/s42256-024-00807-9).
- Suraj Pai, Ibrahim Hadzic, Dennis Bontempi, Keno Bressem, Benjamin H. Kann, Andriy Fedorov, Raymond H. Mak, and Hugo J. W. L. Aerts. Vision Foundation Models for Computed Tomography, 2025.
- Liron Pantanowitz, Navid Farahani, and Anil Parwani. Whole slide imaging in pathology: Advantages, limitations, and emerging perspectives. *Pathology and Laboratory Medicine International*, page 23, June 2015. ISSN 1179-2698. doi:[10.2147/PLMI.S59826](https://doi.org/10.2147/PLMI.S59826).
- Magdalini Paschali, Zhihong Chen, Louis Blankemeier, Maya Varma, Alaa Youssef, Christian Bluethgen, Curtis Langlotz, Sergios Gatidis, and Akshay Chaudhari. Foundation Models in Radiology: What, How, Why, and Why Not. *Radiology*, 314(2):e240597, February 2025. ISSN 0033-8419, 1527-1315. doi:[10.1148/radiol.240597](https://doi.org/10.1148/radiol.240597).
- Steve Paxton, Michelle Peckham, and Adele Knibbs. The Leeds Histology Guide. 2003.
- R Peto, M C Pike, P Armitage, N E Breslow, D R Cox, S V Howard, N Mantel, K McPherson, J Peto, and P G Smith. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer*, 34(6):585–612, December 1976. ISSN 0007-0920, 1532-1827. doi:[10.1038/bjc.1976.220](https://doi.org/10.1038/bjc.1976.220).
- Elisabeth Pfahler, Ivan Zhovannik, Lise Wei, Ronald Boellaard, Andre Dekker, René Monshouwer, Issam El Naqa, Jan Bussink, Robert Gillies, Leonard Wee, and Alberto Traverso. A systematic review and quality of reporting checklist for repeatability and reproducibility of radiomic features. *Physics and Imaging in Radiation Oncology*, 20:69–75, October 2021. ISSN 24056316. doi:[10.1016/j.phro.2021.10.007](https://doi.org/10.1016/j.phro.2021.10.007).
- Robert A. Pooley. Fundamental Physics of MR Imaging. *RadioGraphics*, 25(4):1087–1099, July 2005. ISSN 0271-5333, 1527-1323. doi:[10.1148/rg.254055027](https://doi.org/10.1148/rg.254055027).
- George A. Poulsides, Fei Bao, Elliot L. Servais, Tina Hernandez-Boussard, Ronald P. DeMatteo, Peter J. Allen, Yuman Fong, Nancy E. Kemeny, Leonard B. Saltz, David S. Klimstra, William R. Jarnagin, Jinru Shia, and Michael I. D'Angelica. Pathologic Response to Preoperative Chemotherapy in Colorectal Liver Metastases: Fibrosis, not Necrosis, Predicts Outcome. *Annals of Surgical Oncology*, 19(9):2797–2804, September 2012. ISSN 1068-9265, 1534-4681. doi:[10.1245/s10434-012-2335-1](https://doi.org/10.1245/s10434-012-2335-1).

- Olivia Prior, Carlos Macarro, Víctor Navarro, Camilo Monreal, Marta Ligero, Alonso Garcia-Ruiz, Garazi Serna, Sara Simonetti, Irene Braña, María Vieito, Manuel Escobar, Jaume Capdevila, Annette T. Byrne, Rodrigo Di-enstmann, Rodrigo Toledo, Paolo Nuciforo, Elena Garralda, Francesco Grussu, Kinga Bernatowicz, and Raquel Perez-Lopez. Identification of Precise 3D CT Radiomics for Habitat Computation by Machine Learning in Cancer. *Radiology: Artificial Intelligence*, 6(2):e230118, March 2024. ISSN 2638-6100. doi:[10.1148/ryai.230118](https://doi.org/10.1148/ryai.230118).
- L. Rubbia-Brandt, E. Giostra, C. Brezault, A.D. Roth, A. Andres, V. Audard, P. Sartoretti, B. Dousset, P.E. Majno, O. Soubrane, S. Chaussade, G. Mentha, and B. Terris. Importance of histological tumor response assessment in predicting the outcome in patients with colorectal liver metastases treated with neo-adjuvant chemotherapy followed by liver surgery. *Annals of Oncology*, 18(2):299–304, February 2007. ISSN 09237534. doi:[10.1093/annonc/mdl386](https://doi.org/10.1093/annonc/mdl386).
- David Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982. ISSN 0006-3444, 1464-3510. doi:[10.1093/biomet/69.1.239](https://doi.org/10.1093/biomet/69.1.239).
- Erich Schubert. Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *ACM SIGKDD Explorations Newsletter*, 25(1):36–42, June 2023. ISSN 1931-0145, 1931-0153. doi:[10.1145/3606274.3606278](https://doi.org/10.1145/3606274.3606278).
- Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), March 1978. ISSN 0090-5364. doi:[10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136).
- Eran Segal, Claude B Sirlin, Clara Ooi, Adam S Adler, Jeremy Gollub, Xin Chen, Bryan K Chan, George R Matcuk, Christopher T Barry, Howard Y Chang, and Michael D Kuo. Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nature Biotechnology*, 25(6):675–680, June 2007. ISSN 1087-0156, 1546-1696. doi:[10.1038/nbt1306](https://doi.org/10.1038/nbt1306).
- Rebecca L. Siegel, Kimberly D. Miller, Ann Goding Sauer, Stacey A. Fedewa, Lynn F. Butterly, Joseph C. Anderson, Andrea Cerck, Robert A. Smith, and Ahmedin Jemal. Colorectal cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(3):145–164, May 2020. ISSN 0007-9235, 1542-4863. doi:[10.3322/caac.21601](https://doi.org/10.3322/caac.21601).
- Amber L. Simpson, Jacob Peoples, John M. Creasy, Gabor Fichtinger, Natalie Gangai, Krishna N. Keshavamurthy, Andras Lasso, Jinru Shia, Michael I. D'Angelica, and Richard K. G. Do. Preoperative CT and survival data for patients undergoing resection of colorectal liver metastases. *Scientific Data*, 11(1):172, February 2024. ISSN 2052-4463. doi:[10.1038/s41597-024-02981-2](https://doi.org/10.1038/s41597-024-02981-2).
- Andrew H. Song, Guillaume Jaume, Drew F. K. 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, October 2023. ISSN 2731-6092. doi:[10.1038/s44222-023-00096-8](https://doi.org/10.1038/s44222-023-00096-8).
- Perry Sprawls. *Physical Principles of Medical Imaging*. Medical Physics Publ, Madison, Wis, 2. ed edition, 1995. ISBN 978-0-944838-54-9.
- Daniel C. Sullivan, Nancy A. Obuchowski, Larry G. Kessler, David L. Raunig, Constantine Gatsonis, Erich P. Huang, Marina Kondratovich, Lisa M. McShane, Anthony P. Reeves, Daniel P. Barboriak, Alexander R. Guimaraes, Richard L. Wahl, and For the RSNA-QIBA Metrology Working Group. Metrology Standards for Quantitative Imaging Biomarkers. *Radiology*, 277(3):813–825, December 2015. ISSN 0033-8419, 1527-1315. doi:[10.1148/radiol.2015142202](https://doi.org/10.1148/radiol.2015142202).
- Alexey Surov, Hans Jonas Meyer, and Andreas Wienke. Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: A meta-analysis. *Oncotarget*, 8(35):59492–59499, August 2017. ISSN 1949-2553. doi:[10.18632/oncotarget.17752](https://doi.org/10.18632/oncotarget.17752).

## REFERENCES

- Patrick Therasse, Susan G. Arbuck, Elizabeth A. Eisenhauer, Jantien Wanders, Richard S. Kaplan, Larry Rubinstein, Jaap Verweij, Martine Van Glabbeke, Allan T. Van Oosterom, Michaele C. Christian, and Steve G. Gwyther. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. *JNCI: Journal of the National Cancer Institute*, 92(3):205–216, February 2000. ISSN 0027-8874, 1460-2105. doi:[10.1093/jnci/92.3.205](https://doi.org/10.1093/jnci/92.3.205).
- Choon Hua Thng. Perfusion magnetic resonance imaging of the liver. *World Journal of Gastroenterology*, 16(13):1598, 2010. ISSN 1007-9327. doi:[10.3748/wjg.v16.i13.1598](https://doi.org/10.3748/wjg.v16.i13.1598).
- Paul S. Tofts. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *Journal of Magnetic Resonance Imaging*, 7(1):91–101, January 1997. ISSN 1053-1807, 1522-2586. doi:[10.1002/jmri.1880070113](https://doi.org/10.1002/jmri.1880070113).
- Paul S. Tofts and Allan G. Kermode. Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magnetic Resonance in Medicine*, 17(2):357–367, February 1991. ISSN 0740-3194, 1522-2594. doi:[10.1002/mrm.1910170208](https://doi.org/10.1002/mrm.1910170208).
- Osamu Togao, Akio Hiwatashi, Koji Yamashita, Kazufumi Kikuchi, Daichi Momosaka, Koji Yoshimoto, Daisuke Kuga, Masahiro Mizoguchi, Satoshi O Suzuki, Toru Iwaki, Marc Van Cauteren, Koji Iihara, and Hiroshi Honda. Measurement of the perfusion fraction in brain tumors with intravoxel incoherent motion MR imaging: Validation with histopathological vascular density in meningiomas. *The British Journal of Radiology*, 91(1085):20170912, May 2018. ISSN 0007-1285, 1748-880X. doi:[10.1259/bjr.20170912](https://doi.org/10.1259/bjr.20170912).
- Maciej Tomczak and Ewa Tomczak. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, 21(1):19–25, 2014. ISSN 2299-9590.
- 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, November 2018. ISSN 03603016. doi:[10.1016/j.ijrobp.2018.05.053](https://doi.org/10.1016/j.ijrobp.2018.05.053).
- Diamantis I. Tsilimigras, Pnina Brodt, Pierre-Alain Clavien, Ruth J. Muschel, Michael I. D'Angelica, Itaru Endo, Rowan W. Parks, Majella Doyle, Eduardo De Santibañes, and Timothy M. Pawlik. Liver metastases. *Nature Reviews Disease Primers*, 7(1):27, April 2021. ISSN 2056-676X. doi:[10.1038/s41572-021-00261-6](https://doi.org/10.1038/s41572-021-00261-6).
- Alan I Valderrama-Treviño, Baltazar Barrera-Mera, Jesús C Ceballos-Villalva, and Eduardo E Montalvo-Javé. Hepatic Metastasis from Colorectal Cancer. *EuroAsian Journal of Hepato-Gastroenterology*, 7(2):166–175, December 2017. ISSN 2231-5047, 2231-5128. doi:[10.5005/jp-journals-10018-1241](https://doi.org/10.5005/jp-journals-10018-1241).
- Gert G Van den Eynden, Ali W Majeed, Martin Illemann, Peter B Vermeulen, Nigel C Bird, Gunilla Hoyer-Hansen, Rikke Løvendahl Eefsen, Andrew R Reynolds, and Pnina Brodt. The multifaceted role of the microenvironment in liver metastasis: Biology and clinical implications. *Cancer Res.*, 73(7):2031–2043, April 2013. ISSN 0008-5472. doi:[10.1158/0008-5472.CAN-12-3931](https://doi.org/10.1158/0008-5472.CAN-12-3931).
- Joost J.M. Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina G.H. Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo J.W.L. Aerts. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*, 77(21):e104–e107, November 2017. ISSN 0008-5472, 1538-7445. doi:[10.1158/0008-5472.CAN-17-0339](https://doi.org/10.1158/0008-5472.CAN-17-0339).
- Jelle Veraart, Dmitry S. Novikov, Daan Christiaens, Benjamin Ades-aron, Jan Sijbers, and Els Fieremans. Denoising of diffusion MRI using random matrix theory. *NeuroImage*, 142:394–406, November 2016. ISSN 10538119. doi:[10.1016/j.neuroimage.2016.08.016](https://doi.org/10.1016/j.neuroimage.2016.08.016).
- Mueez Waqar, Petra J. Van Houdt, Eline Hessen, Ka-Loh Li, Xiaoping Zhu, Alan Jackson, Mudassar Iqbal, James O'Connor, Ibrahim Djoukhadar, Uulke A. van der Heide, David J. Coope, and Gerben R. Borst. Visualising spatial heterogeneity in glioblastoma using imaging habitats. *Frontiers in Oncology*, 12:1037896, 2022. ISSN 2234-943X. doi:[10.3389/fonc.2022.1037896](https://doi.org/10.3389/fonc.2022.1037896).

- N A C S Wong and L P Neville. Specificity of intra-acinar necrosis as a marker of colorectal liver metastasis. *Histopathology*, 51(5):725–727, November 2007. ISSN 0309-0167, 1365-2559. doi:[10.1111/j.1365-2559.2007.02837.x](https://doi.org/10.1111/j.1365-2559.2007.02837.x).
- Lei Xu, Pengfei Yang, Eric Alexander Yen, Yidong Wan, Yangkang Jiang, Zuozhen Cao, Xiaoyong Shen, Yan Wu, Jing Wang, Chen Luo, and Tianye Niu. 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, November 2019. ISSN 1361-6560. doi:[10.1088/1361-6560/ab489f](https://doi.org/10.1088/1361-6560/ab489f).
- Shangyuan Ye, Jeong Youn Lim, and Wei Huang. Statistical considerations for repeatability and reproducibility of quantitative imaging biomarkers. *BJR|Open*, 4(1):20210083, January 2022. ISSN 2513-9878. doi:[10.1259/bjro.20210083](https://doi.org/10.1259/bjro.20210083).
- Miho I. Yoshikawa, Shozo Ohsumi, Shigenori Sugata, Masaaki Kataoka, Shigemitsu Takashima, Teruhito Mochizuki, Hirohiko Ikura, and Yutaka Imai. Relation between cancer cellularity and apparent diffusion coefficient values using diffusion-weighted magnetic resonance imaging in breast cancer. *Radiation Medicine*, 26(4):222–226, May 2008. ISSN 0288-2043, 1862-5274. doi:[10.1007/s11604-007-0218-3](https://doi.org/10.1007/s11604-007-0218-3).
- Giulia A. Zamboni, Giovanni Cappello, Damiano Caruso, Sofia Gourtsoyianni, Clemens Cyran, Heinz-Peter Schlemmer, Melvin D'Anastasi, Laure Fournier, and Emanuele Neri. ESR Essentials: Response assessment criteria in oncologic imaging—practice recommendations by the European Society of Oncologic Imaging. *European Radiology*, 35(2):674–683, August 2024. ISSN 1432-1084. doi:[10.1007/s00330-024-11006-w](https://doi.org/10.1007/s00330-024-11006-w).
- Fadl A. Zeineddine, Mohammad A. Zeineddine, Abdelrahman Yousef, Yue Gu, Saikat Chowdhury, Arvind Dasari, Ryan W. Huey, Benny Johnson, Bryan Kee, Michael S. Lee, Maria Pia Morelli, Van K. Morris, Michael J. Overman, Christine Parseghian, Kanwal Raghav, Jason Willis, Robert A. Wolff, Yoshikuni Kawaguchi, Jean-Nicolas Vauthey, Ryan Sun, Scott Kopetz, and John Paul Shen. Survival improvement for patients with metastatic colorectal cancer over twenty years. *npj Precision Oncology*, 7(1):16, February 2023. ISSN 2397-768X. doi:[10.1038/s41698-023-00353-4](https://doi.org/10.1038/s41698-023-00353-4).
- Kelly H. Zou, Simon K. Warfield, Aditya Bharatha, Clare M.C. Tempany, Michael R. Kaus, Steven J. Haker, William M. Wells, Ferenc A. Jolesz, and Ron Kikinis. Statistical validation of image segmentation quality based on a spatial overlap index1. *Academic Radiology*, 11(2):178–189, February 2004. ISSN 10766332. doi:[10.1016/S1076-6332\(03\)00671-8](https://doi.org/10.1016/S1076-6332(03)00671-8).
- Alex Zwanenburg, Stefan Leger, Linda Agolli, Karoline Pilz, Esther G. C. Troost, Christian Richter, and Steffen Löck. Assessing robustness of radiomic features by image perturbation. *Scientific Reports*, 9(1):614, January 2019. ISSN 2045-2322. doi:[10.1038/s41598-018-36938-4](https://doi.org/10.1038/s41598-018-36938-4).
- Alex Zwanenburg, Martin Vallières, Mahmoud A. Abdalah, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*, 295(2):328–338, May 2020. ISSN 0033-8419, 1527-1315. doi:[10.1148/radiol.2020191145](https://doi.org/10.1148/radiol.2020191145).



# Appendix A

## Image Acquisition Details

### A.1 MRI Acquisition Details (PREDICT)

Patients were scanned on either a 1.5T Siemens Avanto or a 3T GE SIGNA Pioneer system. Each patient was scanned on only one of the two scanners. The protocol included anatomical imaging (T2-weighted and T1-weighted), diffusion MRI, variable flip angle spoiled gradient echo (SGrE) imaging for T1 mapping, and dynamic contrast-enhanced (DCE) MRI.

#### 1.5T Siemens Avanto system

The protocol included high-resolution anatomical T2w and T1w scans, diffusion MRI and different spoiled gradient echo (SGrE) sequences, such as those for T1 mapping and dynamic contrast enhanced (DCE) MRI.

- Anatomical T2w scan: turbo spin echo, TE = 82 ms, TR = 4500 ms, turbo factor of 29, echo spacing 8.2 ms, NEX = 8, 2 concatenations, resolution of 1.4mm × 1.4mm, slice thickness of 5 mm, GRAPPA = 2, acquisition in free breathing.
- Anatomical T1w scan: turbo spin echo, TE = 6.3 ms, TR = 470 ms, turbo factor of 11, echo spacing 6.26 ms, NEX = 6, 6 concatenations, resolution of 1.4mm × 1.4mm, slice thickness of 5 mm, GRAPPA = 2, acquisition in free breathing.
- Diffusion MRI: single-shot twice-refocused spin echo EPI,  $b = \{0, 50, 100, 400, 900, 1200, 1600\}$  s/mm<sup>2</sup>, TR = 7900 ms, averaging of 3 mutually-orthogonal directions, NEX = 2, 1 concatenation, resolution of 1.9mm × 1.9mm, slice thickness of 6 mm, SPAIR fat suppression, GRAPPA = 2, EPI factor 150, echo spacing 0.82 ms, each b-value acquired at TE = {93 ms, 105 ms, 120 ms}, acquisition in free breathing. Additionally, one b = 0 image at TE = 93 ms was acquired with reversed phase encoding polarity.
- SGrE for T1 mapping: FLASH, TE = 1.76 ms, TR = 4.59 ms, NEX = 1, 1 concatenation, resolution of 2.7mm × 2.7mm, slice thickness of 6 mm, flip angles of {5°, 15°, 20°}, GRAPPA = 2, acquisition

## IMAGE ACQUISITION DETAILS

in free breathing.

- SGrE for DCE: same acquisition as for T1 mapping with fixed flip angle of 15°; 26 dynamic acquisitions with temporal resolution of 10s, Gadovist with dose of 0.5ml/Kg injected at 3ml/s followed by a bolus of physiological solution of 20ml at 3ml/s, injection delay of 10s, acquisition in free breathing.

### **3T GE SIGNA Pioneer system**

The protocol included high-resolution anatomical T2w and T1w scans, diffusion MRI and different spoiled gradient echo (SGrE) sequences, as those for T1 mapping and DCE imaging.

- Anatomical T2w scan: fast spin echo, TE = 50 ms, TR = 3750 ms, turbo factor of 16, NEX = 2, resolution of 1.4mm × 1.4mm, slice thickness of 6 mm, respiratory-gated acquisition.
- Anatomical T1w scan: navigated SGrE LAVA-Flex providing water/fat images, TE = 2.60 ms, TR = 5.38 ms, resolution of 1.4mm × 1.4mm, slice thickness of 6 mm, flip angle of 12°, acquisition in free-breathing after liver motion measurement.
- Diffusion MRI: single-shot pulsed gradient spin echo EPI,  $b = \{0, 50, 100, 400, 900, 1200, 1500\}$  s/mm<sup>2</sup>, TR = 3500ms, averaging of 3 mutually-orthogonal directions, NEX = 2, resolution of 2.4mm × 2.4mm, slice thickness of 6 mm, ASPIR fat suppression, ASSET = 2, echo spacing 0.80 ms, each b-value acquired at TE = {75 ms, 90 ms, 105 ms}, respiratory-gated acquisition.
- SGrE for T1 mapping: LAVA, TE = 1.2 ms, TR = 2.72 ms, NEX = 1, resolution of 2.4mm × 2.4mm; slice thickness of 6 mm; flip angles of {5°, 10°, 15°}, ASSET = 2, acquisition of two separate images in breathhold, acquisition of the vendor's B1 map.
- SGrE for DCE: same acquisition as for T1 mapping with fixed flip angle of 12°; 26 dynamic acquisitions with temporal resolution of 10s, Clariscan 0.5 mmol/ml with dose of 0.2ml/Kg injected at 0.5ml/kg at 3ml/s followed by a bolus of physiological solution of 20ml at 3ml/s, injection delay of 10s, acquisition in free breathing.

## A.2 mpMRI Maps Biological Ranges

**Table A.1: mpMRI maps derived with biological ranges.** Thirteen quantitative maps were derived from diffusion-relaxation MRI, variable flip angle T1 mapping, and dynamic contrast-enhanced MRI.

mpMRI metric	Computed from	Units	Biological Range
$ADC_t$	Tissue ADC	$\mu\text{m}^2/\text{ms}$	0.0–3.0
$ADC_v$	Vascular ADC	$\mu\text{m}^2/\text{ms}$	3.0–150.0
$K_t$	Tissue kurtosis excess	Dimensionless	0.0–5.0
$f_v$	Vascular signal fraction	Normalized	0.0–1.0
$T_{2t}$	Tissue $T_2$	ms	0.0–800.0
$D_0$	Intrinsic diffusivity	$\mu\text{m}^2/\text{ms}$	0.8–3.0
$vCS$	Volume-weighted cell size	$\mu\text{m}$	5.0–40.0
$f_{in}$	Intracellular fraction	Normalized	0.0–1.0
$CD$	Cell density	Cells/ $\text{mm}^3$	0.0–8e6
$T_1$	$T_1$	ms	0.0–5000.0
$T_2^*$	$T_2^*$	ms	0.0–200.0
$K^{trans}$	Capillary permeability	$\text{min}^{-1}$	0.0–4.0
$v_e$	EES volume	Normalized	0.0–1.0



# Appendix B

## Identification of Precise Handcrafted Features

### B.1 Cohort Characteristics

**Table B.1:** Total number of patients, images, and lesions per cohort and lesion location.

Primary tumor	Patients		Images		Lesions	
	Liver	Lung	Liver	Lung	Liver	Lung
Colorectal	63	12	186	29	959	122
Lung	13	72	22	102	89	141
Neuroendocrine	86	0	86	0	447	0
Mixed	44	41	93	87	366	312
<b>Total</b>	206	125	387	218	1861	575
	<b>331</b>		<b>605</b>		<b>2436</b>	

**Table B.2:** Image acquisition parameters per cohort. (\*) Median [IQR]

	Colorectal (n=215)	Lung (n=124)	Neuroendocrine (n=86)	Mixed (n=180)
<b>Manufacturers</b>				
SIEMENS/PHILIPS/ TOSHIBA/GE	138/58/9/10	63/40/0/21	22/35/6/23	144/23/3/10
<b>Tube Voltage (kVP)</b>				
100/110/120/ 130/140/unknown	25/17/161/ 0/0/12	6/1/105/ 3/0/9	10/3/70/ 1/2/0	14/7/158/ 0/1/0
<b>Recon. kernel</b>				
SOFT/STANDARD/B B20f/B30f/B31f I31s/I50s/unknown	1/7/96 16/38/12 15/6/20	14/7/43 11/23/0 0/0/22	1/19/36 2/2/4 0/0/19	2/11/21 9/112/4 0/0/21
<b>Slice thickness (mm)*</b>	2.0 [2.0–5.0]	2.5 [2.0–5.0]	2.0 [2.0–3.0]	5.0 [1.0–5.0]
<b>Pixel spacing (mm)*</b>	0.92 [0.77–0.98]	0.91 [0.81–0.98]	0.75 [0.70–0.82]	0.98 [0.82–0.98]

**Table B.3:** List of primary tumor types included within the mixed cohort.

Primary tumor	Patients		Images		Lesions	
	Liver	Lung	Liver	Lung	Liver	Lung
Adrenal	3	0	5	0	35	5
Biliary Tract	11	5	24	11	71	66
Bladder	3	3	5	5	41	20
Bone	0	1	0	3	0	21
Breast	4	2	9	6	32	6
Cervix	2	2	3	3	21	21
Esophagus	1	2	2	4	8	9
Head & Neck	2	4	4	11	13	13
Kidney	1	2	2	3	4	18
Liver	2	1	3	1	11	1
Ovary	1	2	2	4	4	26
Pancreas	2	0	4	0	12	0
Penis	1	0	3	0	9	0
Skin	6	11	13	19	68	39
Stomach	4	0	10	0	29	0
Thymus	1	0	4	0	8	0
Thyroid	0	6	0	17	0	52
<b>Total</b>	44	41	93	87	366	312
		<b>85</b>		<b>180</b>		<b>678</b>

## B.2 Radiomics Features and Computation

**Table B.4:** List of radiomics features analyzed in this study. 93 voxel-wise features were computed. 91 were analyzed after excluding GLCM\_MCC and FirstOrder\_TotalEnergy. Feature definitions are available in the IBSI reference manual [Zwanenburg et al., 2020].

Class	Feature	Class	Feature
<b>First Order</b>	10Percentile, 90Percentile, Energy, Entropy, InterquartileRange, Kurtosis, Maximum, MeanAbsoluteDeviation, Mean, Median, Minimum, Range, RobustMeanAbsoluteDeviation, RootMeanSquared, Skewness, TotalEnergy, Uniformity, Variance	<b>GLRLM</b>	GrayLevelNonUniformity, GrayLevelNonUniformityNorm., GrayLevelVariance, HighGrayLevelRunEmphasis, LongRunEmphasis, LongRunHighGrayLevelEmphasis, LongRunLowGrayLevelEmphasis, LowGrayLevelRunEmphasis, RunEntropy, RunLengthNonUniformity, RunLengthNonUniformityNorm., RunPercentage, RunVariance, ShortRunEmphasis, ShortRunHighGrayLevelEmphasis, ShortRunLowGrayLevelEmphasis
<b>GLSZM</b>	GrayLevelNonUniformity, GrayLevelNonUniformityNorm., GrayLevelVariance, HighGrayLevelZoneEmphasis, LargeAreaEmphasis, LargeAreaHighGrayLevelEmphasis, LargeAreaLowGrayLevelEmphasis, LowGrayLevelZoneEmphasis, SizeZoneNonUniformity, SizeZoneNonUniformityNorm., SmallAreaEmphasis, SmallAreaHighGrayLevelEmphasis, SmallAreaLowGrayLevelEmphasis, ZoneEntropy, ZonePercentage, ZoneVariance	<b>GLDM</b>	DependenceEntropy, DependenceNonUniformity, DependenceNonUniformityNorm., DependenceVariance, GrayLevelNonUniformity, GrayLevelVariance, HighGrayLevelEmphasis, LargeDependenceEmphasis, LargeDependenceHighGrayLevelEmphasis, LargeDependenceLowGrayLevelEmphasis, LowGrayLevelEmphasis, SmallDependenceEmphasis, SmallDependenceHighGrayLevelEmphasis, SmallDependenceLowGrayLevelEmphasis
<b>GLCM</b>	Autocorrelation, ClusterProminence, ClusterShade, ClusterTendency, Contrast, Correlation, DifferenceAverage, DifferenceEntropy, DifferenceVariance, Id, Idm, Idmn, Idn, Imc1, Imc2, InverseVariance, JointAverage, JointEnergy, JointEntropy, MCC (Excluded), MaximumProbability, SumAverage, SumEntropy, SumSquares	<b>NGTDM</b>	Busyness, Coarseness, Complexity, Contrast, Strength

**Table B.5:** Image processing and radiomics feature computation parameters.

Parameter	Value/Description
<b>Image Processing</b>	
Software	PyRadiomics v3.0.1, installed in Python 3.7.10
Bounding box	Defined by the segmentation, extended by default padding distance.
Resampled voxel spacing	1 x 1 x 1 mm
Image interpolation	B-spline
Intensity rounding	None
ROI interpolation	Nearest neighbor
Resegmentation	None
<b>Feature Computation</b>	
Kernel radius	1 / 3 mm
Discretization	12 / 25 HU (Fixed Bin Size)
Image filter	None
maskedKernel	True (only voxels in kernel also in ROI used)
Initvalue	NaN (voxels outside ROI considered transparent)
Distance weighting	No weighting
GLCM Symmetry	Symmetric
Distance Metrics	Chebyshev distance $\delta = 1$
NGTDM Coarseness	Coarseness parameter $\alpha = 0$

## B.3 Image Perturbation

Image perturbation was carried out in three ways: rotation, translation, and noise addition. While the first two emulate changes in patient positioning, the latter represents the noise present in different voxel intensities in CT images. Perturbations were performed as described in [Bernatowicz et al. \[2021\]](#), where the authors demonstrated that the combination of these three perturbations simulate the retest scenario. Briefly, we added Gaussian noise (mean 0, standard deviation as present in the image); for translation, we shifted the voxel grid by a fraction of the image voxel spacing following; finally, we rotated the image around the z-axis by an angle of 0.5°.

## B.4 Habitat Computation

To take into account intravoxel heterogeneity, we decided to choose a probabilistic model, Gaussian Mixture Models (GMMs), for clustering rather than a deterministic approach. GMMs, which have been previously used in similar contexts [[Chen et al., 2019](#), [Jardim-Perassi et al., 2019](#)], are generative probabilistic models that find a mixture of multiple Gaussian probability distributions that best fit the data. The Expectation-Maximization (EM) algorithm is used to estimate the model parameters [[Bishop, 2006](#)]. A GMM is represented by the following formula:

$$P(x) = \sum_k (\pi_k N(x | \mu_k, \Sigma_k))$$

where

$P(x)$  : probability density of the data point  $x$

$\pi_k$ : mixing coefficient for the  $k$ th Gaussian component

$N(x | \mu_k, \Sigma_k)$ :  $k$ th Gaussian component with mean  $\mu_k$  and covariance matrix  $\Sigma_k$

To determine the optimal number of habitats ( $k$ ), we used the Bayesian Information Criterion (BIC). The formula for BIC is:

$$BIC = -2 \log(L) + d \log(n)$$

where

$L$  : likelihood of the data given the model

$d$ : number of parameters

$n$ : number of data points

The BIC score is a measure of the trade-off between model complexity and goodness of fit. It penalizes

models with more parameters, such as GMMs with more clusters. In general, lower BIC scores indicate better model fit. However, depending on data characteristics, a clear minimum in BIC scores might not be observed and thus, the gradient can be used to determine the optimal number of clusters. This was our case and therefore we performed a GMM fit for different values of clusters ( $k$ ): {2, 3, 4 and 5}. The maximum number of 5 clusters was determined by being the maximum number of tissue types observed in histology by an experienced pathologist. The optimal value of  $k$  was the one where the change in BIC score with respect to  $k$  was maximal, which was an indication that adding more clusters after that point does not improve the model fit significantly. A cluster number was automatically selected by BIC using the precise original radiomics data and was given as a parameter to the GMM model to compute imaging habitats in both the original and perturbed data. GMM was implemented using Python package scikit-learn (v1.0.2) with a random seed of 123, and default parameters (except for the number of clusters), specifically maximum iteration of 100, convergence threshold of  $10^{-3}$ , full covariance type and initialization with kmeans.

In addition, The Hungarian algorithm (also known as the Kuhn-Munkres algorithm) [Kuhn, 1955], was used to match habitats between original and perturbed data. The Hungarian algorithm is a combinatorial optimization algorithm that solves the assignment problem in polynomial time. It finds an optimal one-to-one matching between two sets by minimizing the total cost (in our case, the difference in cluster assignments).

Finally, to quantify habitat stability, we computed the Dice Similarity Coefficient (DSC) [Zou et al., 2004] between original and perturbed habitats for each habitat within a lesion, across all lesions. The DSC is a widely used metric for evaluating the overlap between two sets, with a higher DSC indicating greater similarity.

All codes are publicly available at <https://github.com/radiomicsgroup/precise-habitats>.

## B.5 Intraclass Correlation Coefficient

An Intraclass Correlation Coefficient (ICC) value of 1 indicates that a feature is highly repeatable/reproducible whereas a value of 0 implies no reliability. Negative ICC values were truncated at 0 as proposed and done previously [Bartko, 1976, Fornacon-Wood et al., 2020]. The ICC is calculated by mean squares obtained through the analysis of variance (ANOVA). In this study, we use two versions of the ICC that are based on a two-way mixed effects ANOVA model, following Koo's guidelines. Below we describe the formulas used to compute the ICC formulas. More information regarding such formulas can be found in the highly cited paper from McGraw and Wong.

To compute the ANOVA model let's consider a dataframe with dimensions  $n \times k$  dataframe where  $n$  is the total number of voxels (rows) for one region of interest (ROI) and  $k$  is the total number of conditions or measurements (columns). In our case,  $k=2$ . For repeatability the two conditions are original-perturbed (test-retest) and for reproducibility against kernel size the two conditions are computation with radius kernel 1mm or radius kernel 3mm (or bin size 12HU or 25HU in the case of reproducibility against bin size). Each voxel measurement is indexed as  $Y_{ij}$  where  $i$  denotes the voxel ( $i = 1, \dots, n$ ) and  $j$  denotes

the measurement under the repeatability/reproducibility condition ( $j = 1 \dots k$ ). We define the following concepts:

$\bar{Y}_i$ : mean of all voxel values in a column

$$\bar{Y}_i = \frac{\sum_{j=1}^k Y_{ij}}{k}$$

$\bar{Y}_j$  : mean of all voxel values in a column

$$\bar{Y}_j = \frac{\sum_{i=1}^n Y_{ij}}{n}$$

$\mu$  : mean of all values (also called *grand mean*)

$$\mu = \frac{\sum_{j=1}^k \sum_{i=1}^n Y_{ij}}{n * k}$$

$\sigma_w^2$  : Within-voxel variance, the estimated variance of repeated measurements

$$\sigma_w^2 = \frac{\sum_{j=1}^k (\bar{Y}_j - \bar{Y}_i)^2}{k - 1}$$

$\sigma_w$  : Within-voxel standard deviation, the standard deviation we get if we measure the voxel multiple times. Calculated by averaging the within-subject sample variances. Since we have a variance per voxel and we can't meaningfully take the average of a list of standard deviations, we first calculate the variance for each voxel, and then compute the average of those, and finally square root that mean variance [Ye et al., 2022].

$$\sigma_w = \sqrt{\frac{\sum_{i=1}^n \frac{\sum_{j=1}^k (\bar{Y}_{ij} - \bar{Y}_i)^2}{k-1}}{n}}$$

The degrees of freedom, sum squares and mean square expectations that correspond to a two-way mixed ANOVA model are summarized below.

## IDENTIFICATION OF PRECISE HANDCRAFTED FEATURES

**Table B.6: Two-way Mixed ANOVA Model.** MSC: mean square columns, MSR: mean square rows, MSE: mean square error, SSC= sum of squares columns, SSR= sum of squares rows, SST= sum of squares total, SSE= sum of squares error, dfc= degrees of freedom columns, dfr=degrees of freedom rows, dfe=degrees of freedom errors

Source of Variation	Degrees of freedom	Sum Squares	Mean Square Expectations
Conditions (columns)	$dfc = k - 1$	$SSC = \sum_{j=1}^k n \times (\bar{Y}_j - \mu)^2$	$MSC = \frac{SSC}{dfc \times n}$
Voxels (rows)	$dfr = n - 1$	$SSR = \sum_{i=1}^n k \times (\bar{Y}_i - \mu)^2$	$MSR = \frac{SSR}{dfr}$
Total		$SST = \sum_{j=1}^k \sum_{i=1}^n (Y_{ij} - \mu)^2$	=
Error (or residual)	$dfe = (n - 1)(k - 1)$	$SSE = SST - SSC - SSR$	$MSE = \frac{SSE}{dfe}$

We compute the two versions of ICC for repeatability and reproducibility:

Repeatability ICC(3A,1): ICC based on single-measurement, absolute-agreement, two-way mixed-effects model.

$$ICC(3A,1) = \frac{MSR - MSE}{MSR + dfc \times MSE + \frac{k}{n} \times (MSC - MSE)}$$

Reproducibility ICC(3C,1): ICC based on single-measurement, consistency, two-way mixed-effects model.

$$ICC(3C,1) = \frac{MSR - MSE}{MSR + dfc \times MSE}$$

We compute the lower bound of the 95% CI of the ICC (LCL) and the upper bound (UCL):

$$LCL = \frac{\frac{FR}{F} - 1}{\frac{FR}{F} + k - 1} \quad UCL = \frac{(FR \times F) - 1}{(FR \times F) + k - 1}$$

Where F is the  $(1 - \frac{\alpha}{2}) \times 100^{\text{th}}$  percentile of the F distribution with  $n-1$  numerator degrees of freedom and  $(n-1)(k-1)$  denominator degrees of freedom and FR is the F-statistic for voxels computed as:  $FR = \frac{MSR}{MSE}$ .

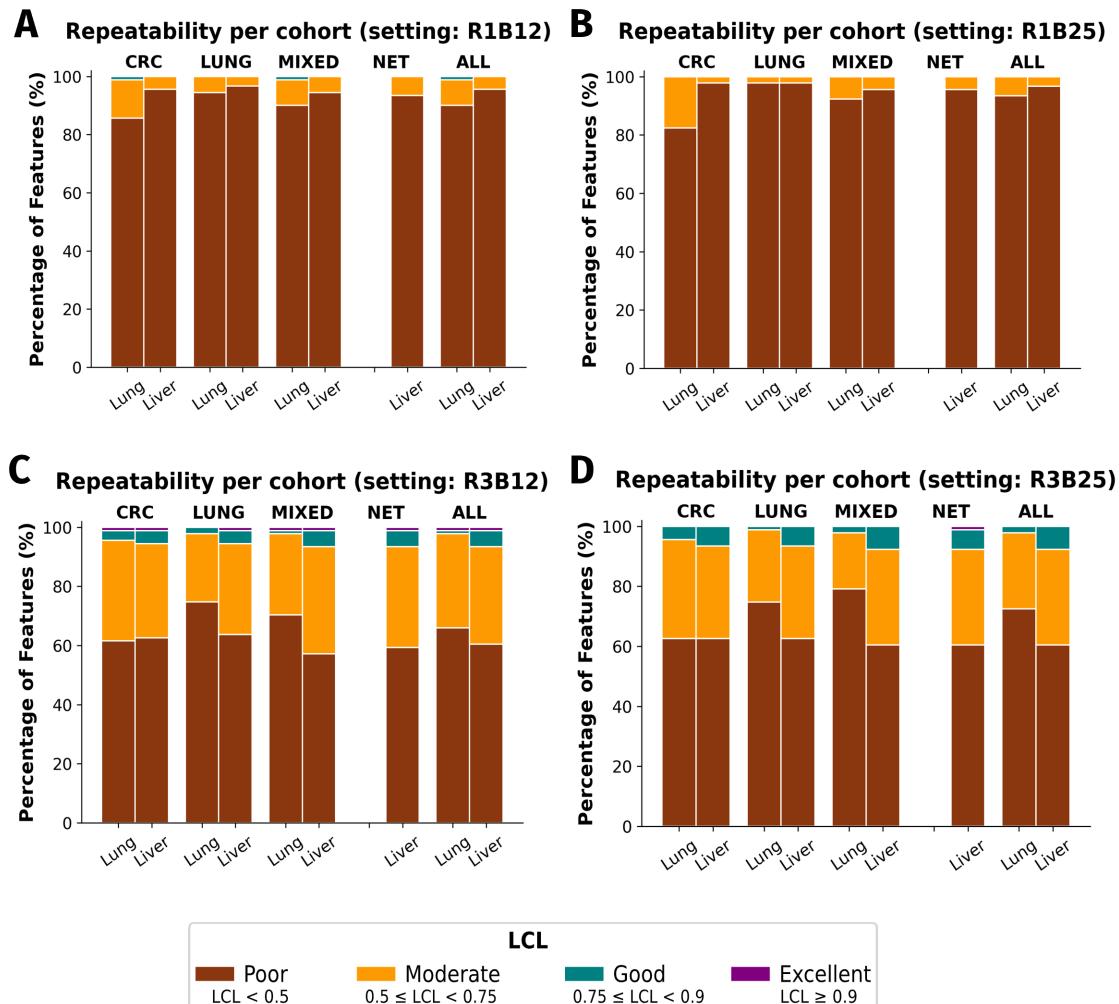
Custom codes used to calculate ICC (3A,1) and ICC (3C,1), based on Nipype's [Esteban et al., 2022] module ICC (v1.8.5) and approved by a statistician (VN) are available at <https://github.com/radio-micsgroup/precise-habitats>.

## B.6 Justification for NGTDM Coarseness Inclusion

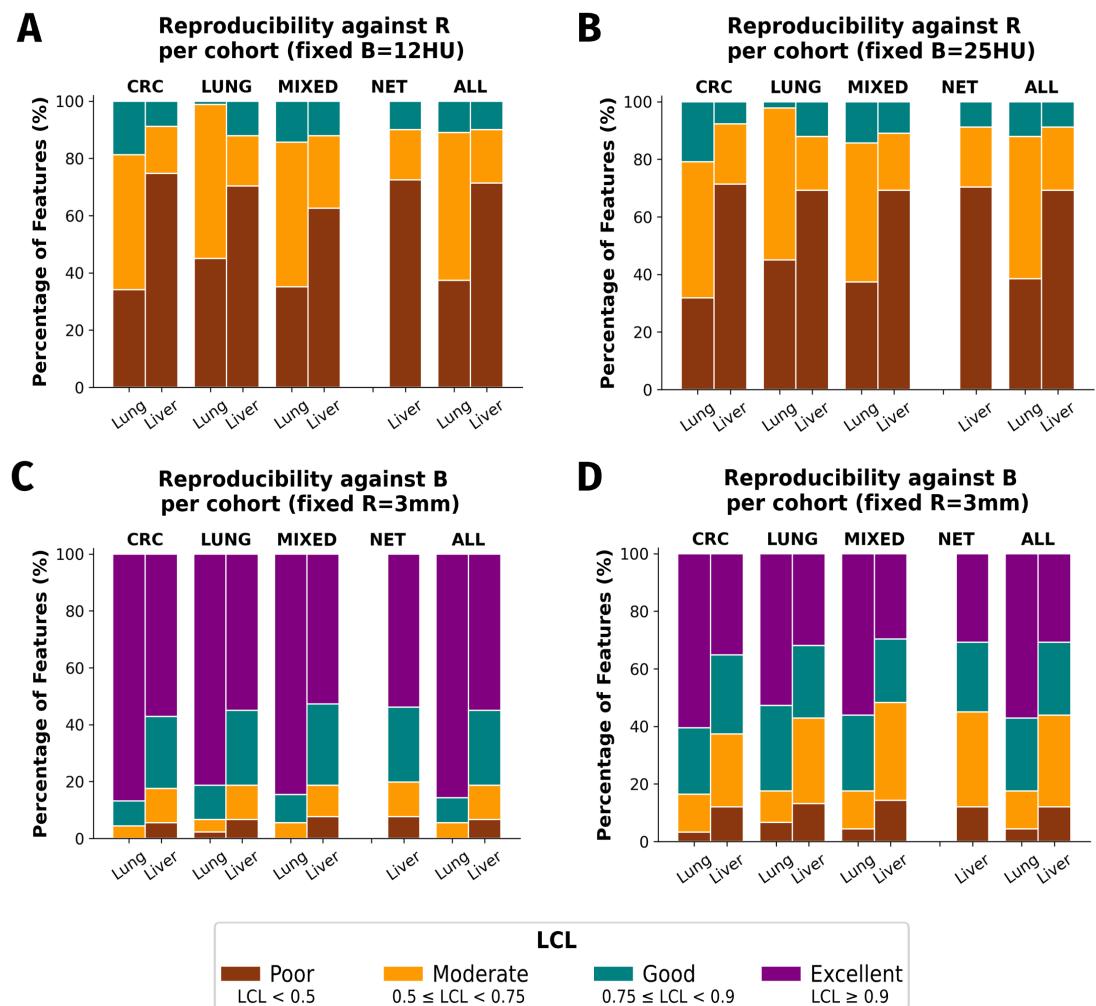
NGTDM (Neighborhood Gray-Tone-Difference Matrix) coarseness describes the roughness (i.e. how fine or coarse) the texture of an image is. In the radiomics literature, evidence has been found regarding its usefulness to characterize heterogeneity and predict progression-free survival in oncology [[Gupta et al., 2021](#)].

In our study, we identified precise features by linking repeatability and reproducibility results. That is, for every feature, we considered results obtained in the three relevant experiments: repeatability (setting R3B12), reproducibility against R (fixed B=12HU), and reproducibility against B (fixed R=3mm). A feature was selected as precise if it presented  $LCL \geq 0.50$  (i.e. moderate, good or excellent repeatability/reproducibility) in the three experiments. NGTDM Coarseness presented excellent repeatability and reproducibility against bin size, but was not selected as precise as it presented poor reproducibility against kernel radius. However, by the nature of its definition, the poor reproducibility against kernel radius is acceptable: the feature captures the distribution of differences in gray-tone values between pairs of neighboring pixels. Considering the feature's excellent results in two out of three experiments, its potential usefulness and in light of the fact that we were already being stringent, first by using LCL rather than ICC and second by linking results of three different experiments, we decided to include it as a precise feature for both liver and lung lesions.

## B.7 Primary Tumor Has No Effect on Precision



**Figure B.1: Repeatability by primary tumor type.** Repeatability distribution of radiomics features computed with setting R1B12 (A), R1B25 (B), R3B12 (C) and R3B25 (D) per cohort for lung and liver lesions separately. Primary tumor has no effect on repeatability. LCL, 95% lower confidence limit of the Intraclass Correlation Coefficient; R1B12, features computed with kernel radius 1mm and bin size 12HU; R1B25, features computed with kernel radius 1mm and bin size 25HU; R3B12, features computed with kernel radius 3mm and bin size 12HU; R3B25, features computed with kernel radius 3mm and bin size 25HU; CRC: colorectal cohort; NET: neuroendocrine cohort; ALL: all cohorts combined.



**Figure B.2: Reproducibility by primary tumor type.** Reproducibility distribution against R of radiomics features computed with fixed bin size of 12HU (A) and fixed bin size of 25HU (B) per cohort for lung and liver lesions separately. Similarly, (C) and (D) depict the reproducibility distribution against B of radiomics features computed with fixed radius of 1mm (C) and 3mm (D) per cohort for lung and liver lesions separately. LCL, 95% lower confidence limit of the Intraclass Correlation Coefficient; CRC: colorectal cohort; NET: neuroendocrine cohort; ALL: all cohorts combined.



# Appendix C

## Development and Validation of an mpMRI-anchored CT Habitat Model

### C.1 CT-mpMRI Co-Registration: Extended Methods

Prior to registration, all images were cropped to a tumor-centered bounding box with a 5–7 mm margin beyond the segmentation boundary. Cropping served two purposes: it reduced computational cost and focused the registration algorithm on the tumor region, avoiding spurious alignment driven by distant anatomical structures (e.g., ribs, spine) that may differ in position between CT and MRI acquisitions.

Images were resampled to 2x2x2 mm isotropic resolution to match the T2-weighted reference. Resampling to a common grid is necessary for voxelwise comparison; we chose the T2w resolution as the reference because it represented the coarsest native resolution among the sequences and avoided artificial upsampling of MRI data.

Three pipelines were required:

- **CT to T2w:** Aligns contrast-enhanced CT to the anatomical MRI reference. This is the most challenging registration due to differences in tissue contrast between modalities.
- **DWI to T2w:** Aligns diffusion-weighted images (and derived ADC maps) to T2w. DWI and T2w are both MRI sequences, but DWI suffers from geometric distortion, particularly near air-tissue interfaces.
- **DCE to T2w:** Aligns dynamic contrast-enhanced images (and derived perfusion maps) to T2w. DCE images were acquired with a GRE sequence at different resolution than T2w.

All registrations were performed using NiftyReg (reg\_aladin for rigid/affine, reg\_f3d for B-spline deformable registration). The pipeline proceeded sequentially: rigid registration was performed first; if affine registration improved DSC, it replaced the rigid result; if B-spline registration improved DSC further, it replaced the affine result. This conservative approach avoided overfitting from unnecessary

deformable registration.

Registration quality was assessed by computing the Dice similarity coefficient between the tumor mask on the fixed image (T2w) and the warped tumor mask from the moving image. Tumors with CT to T2w DSC < 0.50 were excluded from analysis, as low overlap indicates registration failure that would propagate errors into habitat comparisons.

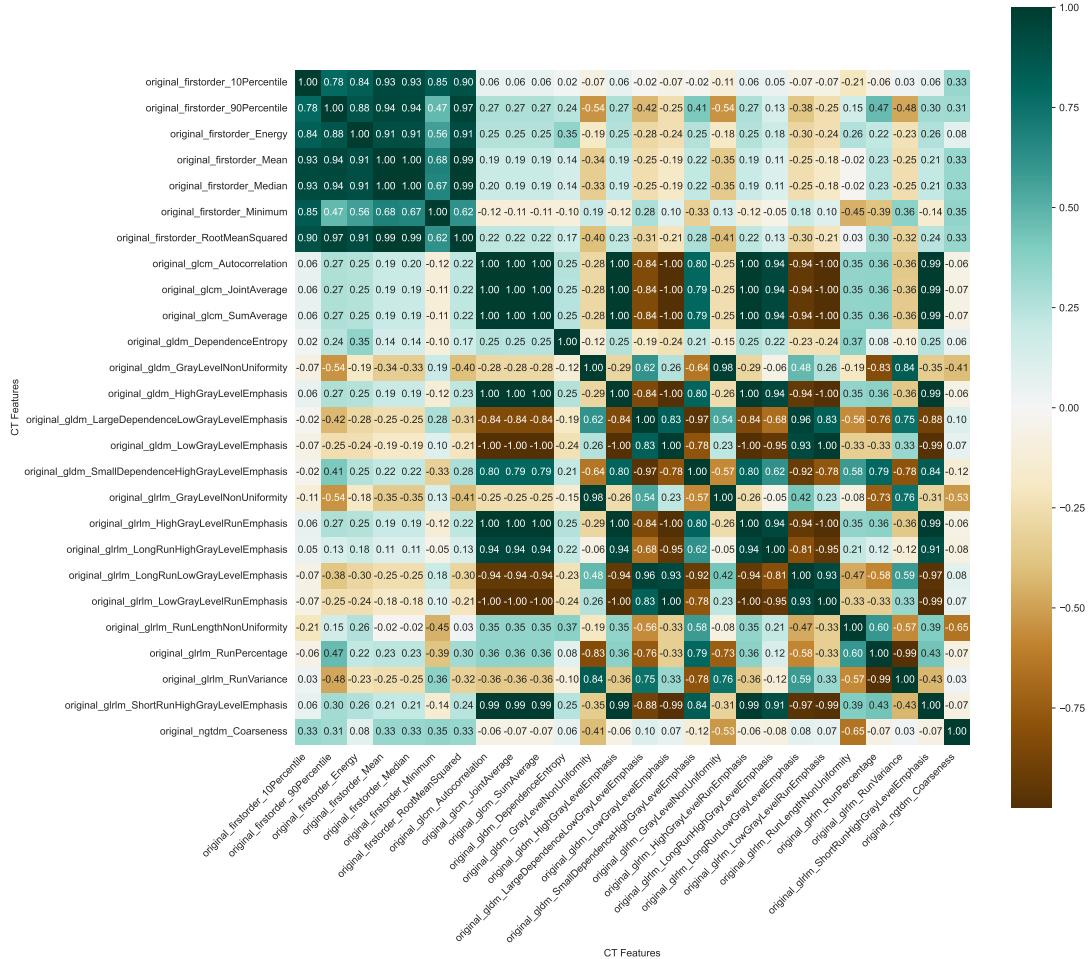
## C.2 Handcrafted Features - Correlation Analysis

Chapter 6 identified 26 radiomics features with acceptable repeatability and reproducibility for liver lesions. However, many of these features are highly correlated, capturing overlapping information. Clustering on redundant features can distort distance metrics and bias cluster assignments toward the correlated feature set.

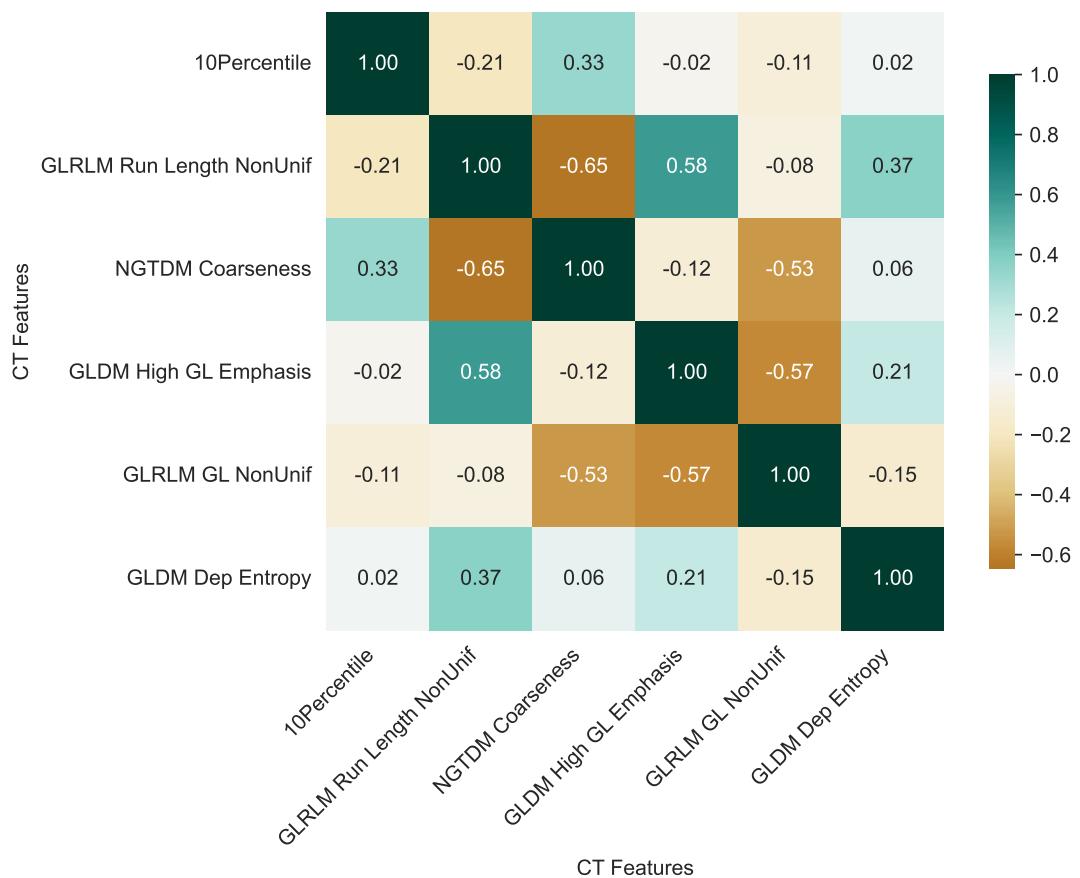
To obtain a non-redundant feature set, we computed pairwise Spearman correlations across all voxels in the PREDICT cohort. Features with  $|\rho| \geq 0.80$  were considered redundant. From each correlated pair, we retained the feature with higher mean repeatability (ICC) from the Chapter 6 analysis. This procedure reduced the 26 precise features to 6 non-redundant features:

1. 10th percentile intensity (first-order)
2. GLDM dependence entropy
3. GLDM small dependence high gray level emphasis
4. GLRLM gray level non-uniformity
5. GLRLM run length non-uniformity
6. NGTDM coarseness

Figure C.2 shows the correlation matrix of the original 26 features and Figure C.3 shows the retained features, which span different texture families (GLDM, GLRLM, NGTDM) and capture distinct aspects of local intensity variation, ensuring that the clustering input is diverse rather than dominated by a single texture property.



**Figure C.1: Exploratory correlations between CT habitat proportions and histological tissue percentages.** Exploratory correlations between CT habitat proportions and histological tissue percentages (POEM cohort, N = 6). Each panel shows the relationship between a habitat proportion (rows: H1, H2, H3) and a histological tissue percentage (columns: necrosis, fibrosis, viable tumor). Spearman correlation coefficients ( $\rho$ ) and p-values are shown. All correlations were non-significant, reflecting the small sample size and CT's inability to distinguish necrosis from fibrosis within the avascular compartment.



**Figure C.2: Correlation matrix of the 26 precise handcrafted radiomics features.** Correlation matrix of the 26 precise handcrafted radiomics features identified in Chapter 6. Pairwise Spearman correlations computed across all voxels in the PREDICT cohort. Features with  $|\rho| \geq 0.80$  were considered redundant.

### C.3 Sensitivity Analysis for Number of Habitats (K)

To determine the optimal number of habitats, we compared  $K = 2$ ,  $3$ , and  $4$  using the handcrafted feature representation. For each  $K$ , we computed habitats and assessed their separation of mpMRI-derived biophysical metrics using Kendall's W effect size.

- **$K = 2$**  achieved the highest effect sizes for most metrics but collapsed biologically distinct compartments into a single “vascularized” cluster. The two-habitat solution could not distinguish the cellular-perfused tumor ( $H_2$ ) from the vascular interface ( $H_3$ ).
- **$K = 3$**  provided the best interpretive balance. It separated the avascular core ( $H_1$ ) from two distinct vascularized phenotypes: a cellular-perfused compartment with high  $K^{trans}$  ( $H_2$ ) and a vascular compartment with high  $f_v$  but moderate  $K^{trans}$  ( $H_3$ ). This three-way distinction aligns with the known histological architecture of colorectal liver metastases.
- **$K = 4$**  introduced a fourth habitat by splitting  $H_3$  into two subgroups. However, this split showed no clear biological rationale—both subhabitats had similar mpMRI profiles—and the fourth cluster showed unstable membership across bootstrap resampling (ARI = 0.91 vs. 0.97 for  $K = 3$ ).

Based on these findings,  $K = 3$  was selected as the final model configuration.

**Table C.1: Biophysical separation (Kendall's W) across different numbers of habitats (K).**

K	$f_v$		$K^{trans}$		$ADC_t$		Interpretation
	W	p	W	p	W	p	
2	0.98	0.002	0.78	0.014	0.62	0.049	Strong separation but oversimplified; merges biologically distinct vascular phenotypes
3	0.67	0.005	0.52	0.018	0.16	0.328	Best balance; separates vascular gradient while identifying cellular-perfused compartment
4	0.85	0	0.37	0.025	0.32	0.039	Fourth habitat splits $H_3$ without biological justification; unstable across bootstrap

### C.4 Technical Validation: Extended Results

- **Initialization stability:** All representations achieved ARI > 0.96, indicating that clustering solutions were reproducible regardless of random seed initialization.
- **Data stability:** All representations showed high bootstrap stability (median ARI > 0.96), indicating that habitat definitions were not driven by a few influential patients.
- **Spatial coherence:** Handcrafted features produced the highest Moran's I (0.804), indicating strong spatial autocorrelation—habitats formed contiguous regions rather than scattered voxels. DL-SALSA

and DL-FM produced lower spatial coherence (0.513–0.622), reflecting more fragmented “salt-and-pepper” patterns. Raw HU showed the lowest spatial coherence (0.420).

The combination of high initialization stability, adequate data stability, and superior spatial coherence supported the selection of handcrafted features for the final CT habitat model.

**Table C.2:** Technical robustness of candidate CT feature representations. ARI = Adjusted Rand Index.

	Initialization Stability (ARI)	Data Stability (ARI) [Median (IQR)]	Moran’s I [Mean ± SD]
Raw HU	0.965	<b>0.991 (0.980–0.997)</b>	0.420 ± 0.143
Handcrafted	<b>0.997</b>	0.966 (0.959–0.985)	<b>0.804 ± 0.056</b>
DL-SALSA	0.984	0.985 (0.977–0.988)	0.513 ± 0.110
DL-FM	0.978	0.989 (0.976–0.992)	0.622 ± 0.119

## C.5 mpMRI Characterization: Pairwise Comparison

For metrics with significant Friedman test ( $p < 0.05$ , BH-corrected), we report Wilcoxon signed-rank tests comparing habitat pairs. Effect size  $r = Z/\sqrt{N}$ . All p-values are BH-corrected within each metric. N = 10 patients.

**Table C.3:** Post-hoc pairwise comparisons for mpMRI metrics across CT habitats. BH = Benjamini-Hochberg corrected p-values; r = effect size.

Metric	H1 vs H2		H1 vs H3		H2 vs H3	
	p (BH)	r	p (BH)	r	p (BH)	r
$ADC_t$	0.029	0.82	0.24	0.44	1	0
$ADC_v$	0.126	0.55	0.111	0.66	0.232	0.38
$f_v$	0.375	0.28	0.006	0.98	0.006	0.91
$D_0$	0.049	0.62	0.003	0.98	0.003	0.98
$T_2^*$	0.02	0.74	0.003	0.98	0.003	0.98
$T_1$	0.029	0.74	0.012	0.91	0.037	0.66
$K^{trans}$	0.015	0.87	0.015	0.82	0.16	0.44

### Key findings:

- **Vascular gradient (H1 to H3):**  $f_v$ ,  $D_0$ ,  $T_2^*$ , and  $T_1$  showed significant differences between H1 and H3, with large effect sizes ( $r > 0.90$ ), confirming a vascular gradient from the avascular core to the vascular rim.

- **Cellular peak (H2):** ADCt was significantly lower in H2 than H1 ( $p = 0.029$ ,  $r = 0.82$ ), indicating H2 as the most cellular habitat. Ktrans was significantly higher in H2 than H1 ( $p = 0.015$ ,  $r = 0.87$ ), consistent with leaky tumor neovessels.
- **H2 vs H3 distinction:** fv distinguished H2 from H3 ( $p = 0.006$ ), but Ktrans did not ( $p = 0.160$ ). This supports the interpretation that H3 represents a vascular compartment with mature (less leaky) vessels, possibly reflecting partial volume with normal liver.

## C.6 Habitat-Histology Correlations

As an exploratory analysis, we computed Spearman correlations between whole-tumor habitat proportions and histological tissue percentages in the POEM cohort ( $N = 6$  tumors). Direct voxel-to-voxel co-registration between CT and histology was not feasible; correlations therefore reflect whole-tumor associations only.

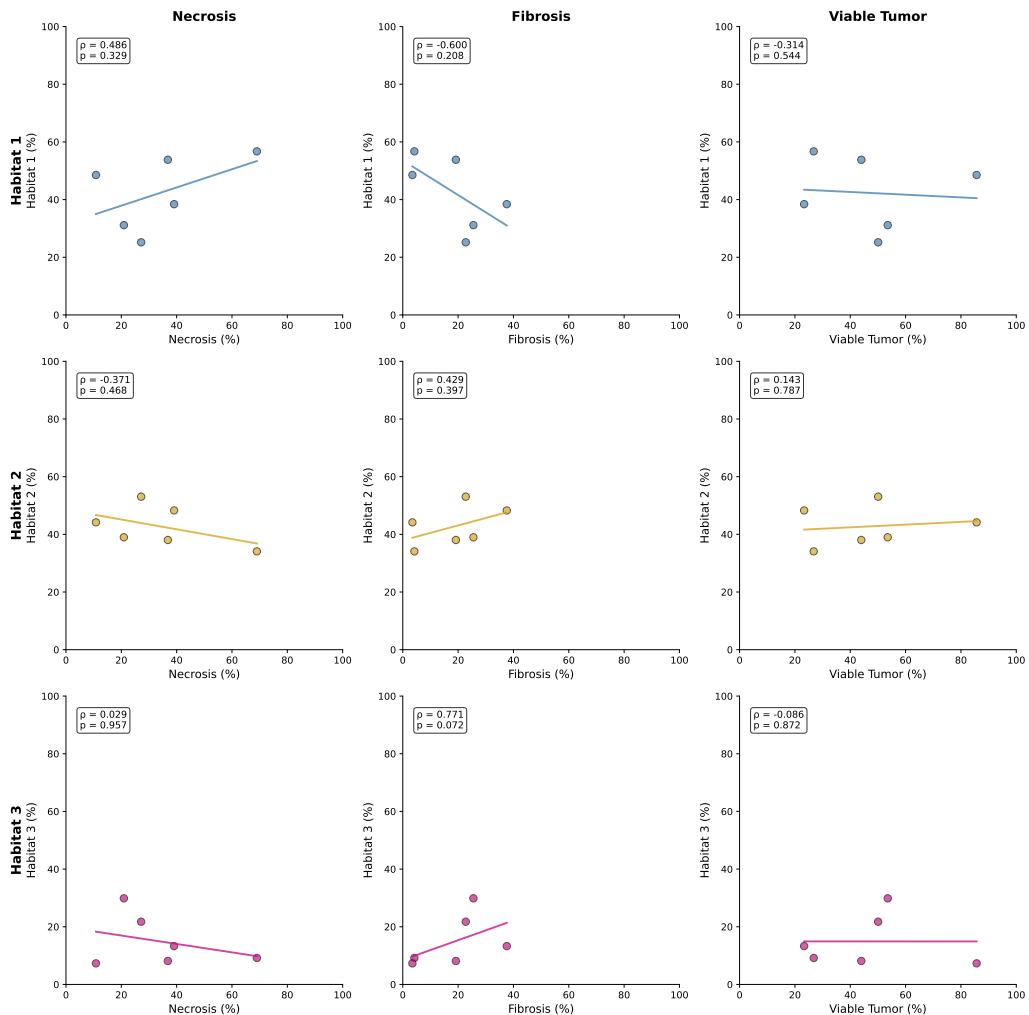
**Table C.4:** Spearman correlations between habitat proportions and histological tissue percentages.

<b>Habitat</b>	<b>Necrosis</b>		<b>Fibrosis</b>		<b>Viable Tumor</b>	
	<b><math>\rho</math></b>	<b><math>p</math></b>	<b><math>\rho</math></b>	<b><math>p</math></b>	<b><math>\rho</math></b>	<b><math>p</math></b>
H1	0.49	0.329	-0.60	0.208	-0.31	0.544
H2	-0.37	0.468	0.43	0.397	0.14	0.787
H3	0.03	0.957	0.77	0.072	-0.09	0.872

All correlations were weak to moderate and non-significant. Several trends were observed:

- **H1 and necrosis:** A positive trend ( $\rho = 0.49$ ) suggests that the avascular habitat may partially capture necrotic tissue, consistent with its biological profile (low vascularity, low cellularity).
- **H1 and fibrosis:** A negative trend ( $\rho = -0.60$ ) suggests that H1 does not specifically correspond to fibrosis. This is expected: CT cannot distinguish necrosis from fibrosis, and both may appear as avascular tissue.
- **H3 and fibrosis:** The strongest trend observed ( $\rho = 0.77$ ,  $p = 0.072$ ) suggests a potential association between the vascular habitat and fibrotic tissue. However, this finding is difficult to interpret biologically and may reflect confounding by tumor size or treatment history.

The sample size ( $N = 6$ ) provides insufficient power to detect moderate correlations. Additionally, the scale mismatch between voxel-level imaging (mm resolution) and microscopic histology (um resolution), combined with tissue deformation during resection and processing, limits the interpretability of whole-tumor correlations. These exploratory findings should be interpreted with caution and require validation in larger cohorts with spatially co-registered imaging and histology.



**Figure C.3: Correlation matrix of the 6 selected precise handcrafted radiomics features.** Correlation matrix of the 6 selected precise handcrafted radiomics features identified in Chapter 6. The 6 non-redundant features retained for habitat computation are highlighted: 10th percentile intensity, GLDM dependence entropy, GLDM small dependence high gray level emphasis, GLRLM gray level non-uniformity, GLRLM run length non-uniformity, and NGTDM coarseness.

# Appendix D

## Clinical Relevance of CT Habitats

### D.1 Habitat Metrics and Tumor Volume

**Table D.1:** Spearman correlations between habitat metrics and tumor volume.

\*p<0.05, \*\*p<0.01, \*\*\*p<0.001.

Metric	TCIA ( $\rho$ )	VHIO ( $\rho$ )
<b>Rim metrics</b>		
Rim entropy vs Volume	-0.77***	-0.71***
Rim cellular-perfused vs Volume	0.49***	0.45***
Rim avascular vs Volume	-0.32***	-0.28***
Rim vascular vs Volume	-0.21**	-0.19**
<b>Whole-tumor metrics</b>		
Whole entropy vs Volume	-0.44***	-0.62***
Whole avascular vs Volume	0.44***	0.70***
Whole cellular-perfused vs Volume	0.12	0.08
Whole vascular vs Volume	-0.38***	-0.52***
<b>Core metrics</b>		
Core entropy vs Volume	-0.28***	-0.41***
Core avascular vs Volume	0.31***	0.55***
Core cellular-perfused vs Volume	-0.18*	-0.22**

## D.2 TCIA: Neoadjuvant Chemotherapy Remodels Tumor Composition

**Table D.2:** Habitat metric comparisons: treatment-naive vs neoadjuvant-treated (TCIA). Mann-Whitney U test. FDR correction using Benjamini-Hochberg procedure. Bold indicates FDR-adjusted p<0.05.

<b>Metric</b>	<b>Treatment-naive (n=74)</b>	<b>Neoadjuvant (n=115)</b>	<b>Unadjusted p</b>	<b>FDR- adjusted p</b>
	<b>Median [IQR]</b>	<b>Median [IQR]</b>		
Tumor volume ( $mm^3$ )	14,639 [5,812–38,421]	8,620 [3,284–24,108]	0.018	0.057
Whole entropy	1.45 [1.38–1.52]	1.51 [1.44–1.57]	<0.001	<b>0.001</b>
Whole avascular	0.398 [0.31–0.49]	0.354 [0.27–0.44]	0.019	0.057
Whole cellular-perf.	0.412 [0.34–0.48]	0.408 [0.35–0.47]	0.75	0.75
Whole vascular	0.168 [0.12–0.22]	0.214 [0.16–0.27]	<0.001	<b>0.001</b>
Rim entropy	1.13 [1.02–1.24]	1.25 [1.14–1.35]	<0.001	<b>0.001</b>
Rim avascular	0.152 [0.09–0.22]	0.138 [0.08–0.20]	0.28	0.42
Rim cellular-perf.	0.485 [0.41–0.56]	0.512 [0.44–0.58]	0.057	0.11
Rim vascular	0.342 [0.28–0.41]	0.328 [0.27–0.39]	0.35	0.47
Core entropy	1.08 [0.95–1.21]	1.12 [1.01–1.24]	0.09	0.16

## D.3 Cox Regression Analyses by Treatment (VHIO)

**Table D.3:** Cox regression: Chemotherapy alone (n=122, 89 events).

<b>Variable</b>	<b>Univariable HR [95% CI]</b>	<b>p</b>	<b>Multivariable HR [95% CI]</b>	<b>p</b>
Tumor volume	5.46 [3.19–9.34]	<0.001	5.21 [2.98–9.12]	<0.001
Rim entropy	0.88 [0.54–1.43]	0.61	—	—
Extrahepatic disease	1.82 [1.18–2.81]	0.007	1.54 [0.98–2.42]	0.06
Age	1.24 [0.92–1.67]	0.16	—	—
<b>C-index</b>			0.703	

**Table D.4:** Cox regression: Chemotherapy + Bevacizumab (n=133, 95 events).

Variable	Univariable HR [95% CI]	p	Multivariable HR [95% CI]	p
Tumor volume	1.69 [1.31–2.17]	<0.001	1.58 [1.21–2.06]	0.001
Rim entropy	0.68 [0.52–0.88]	0.004	0.71 [0.54–0.93]	0.012
Extrahepatic disease	1.44 [0.95–2.18]	0.09	1.38 [0.90–2.12]	0.14
Left primary	0.58 [0.38–0.89]	0.012	0.62 [0.40–0.96]	0.031
C-index			0.674	

**Table D.5:** Cox regression: RAS-mutant patients (n=195, 138 events).

Variable	Univariable HR [95% CI]	p	Multivariable HR [95% CI]	p
Tumor volume	1.89 [1.41–2.53]	<0.001	1.78 [1.32–2.41]	<0.001
Rim entropy	0.80 [0.65–1.00]	0.047	0.82 [0.66–1.02]	0.07
Extrahepatic disease	1.61 [1.14–2.27]	0.007	1.48 [1.04–2.11]	0.03
Age	1.18 [0.96–1.45]	0.11	—	—
C-index			0.700	

## D.4 Longitudinal Analysis Summary Statistics

**Table D.6:** Rim entropy change by RECIST category (n=38).

RECIST Category	n	Median $\Delta$ Rim Entropy	IQR	n	n	Median OS (months)
All patients	38	+0.013	[-0.04, +0.10]	21	17	27.1
PR	23	+0.053	[-0.02, +0.10]	14	9	32.8
SD	8	+0.023	[-0.06, +0.16]	4	4	18.3
PD	7	-0.041	[-0.25, +0.03]	3	4	9.0

