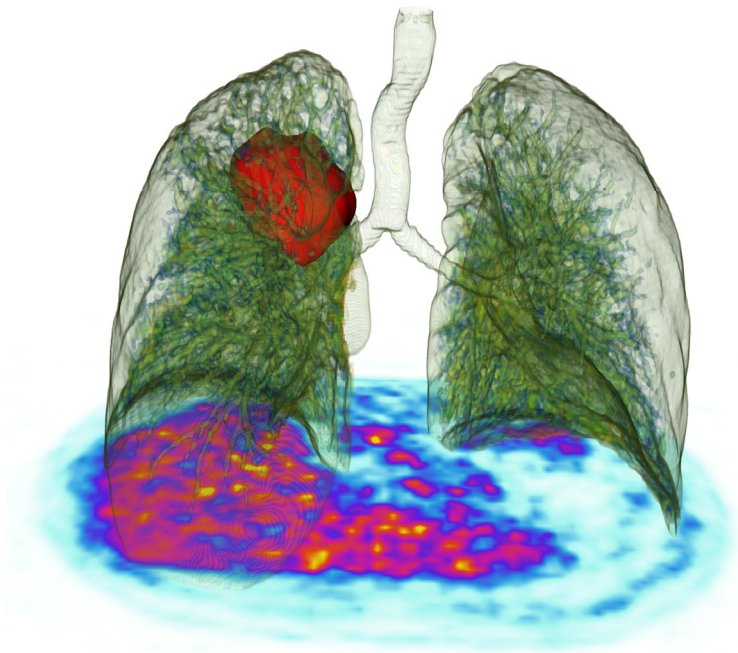


Doctoral Thesis in Medical Technology

Advanced Machine Learning Methods for Oncological Image Analysis

MEHDI ASTARAKI



Stockholm, Sweden 2022



**Karolinska
Institutet**

Advanced Machine Learning Methods for Oncological Image Analysis

MEHDI ASTARAKI

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology and the Karolinska Institute, is submitted for public defence for the Degree of Doctor of Philosophy on Friday the 30th of September 2022, at 1:00 p.m. in T2, Hälsovägen 11C, Huddinge, Sweden.

Doctoral Thesis in Medical Technology
KTH Royal Institute of Technology
Karolinska Institutet
Stockholm, Sweden 2022

© Mehdi Astaraki

ISBN 978-91-8040-313-9
TRITA-CBH-FOU-2022:38

Printed by: Universitetsservice US-AB, Sweden 2022

Abstract

Cancer is a major public health problem, accounting for an estimated 10 million deaths worldwide in 2020 alone. Rapid advances in the field of image acquisition and hardware development over the past three decades have resulted in the development of modern medical imaging modalities that can capture high-resolution anatomical, physiological, functional, and metabolic quantitative information from cancerous organs. Therefore, the applications of medical imaging have become increasingly crucial in the clinical routines of oncology, providing screening, diagnosis, treatment monitoring, and non/minimally-invasive evaluation of disease prognosis. The essential need for medical images, however, has resulted in the acquisition of a tremendous number of imaging scans. Considering the growing role of medical imaging data on one side and the challenges of manually examining such an abundance of data on the other side, the development of computerized tools to automatically or semi-automatically examine the image data has attracted considerable interest. Hence, a variety of machine learning tools have been developed for oncological image analysis, aiming to assist clinicians with repetitive tasks in their workflow.

This thesis aims to contribute to the field of oncological image analysis by proposing new ways of quantifying tumor characteristics from medical image data. Specifically, this thesis consists of six studies, the first two of which focus on introducing novel methods for tumor segmentation. The last four studies aim to develop quantitative imaging biomarkers for cancer diagnosis and prognosis.

The main objective of Study I is to develop a deep learning pipeline capable of capturing the appearance of lung pathologies, including lung tumors, and integrating this pipeline into the segmentation networks to leverage the segmentation accuracy. The proposed pipeline was tested on several comprehensive datasets, and the numerical quantifications show the superiority of the proposed prior-aware DL framework compared to the state of the art. Study II aims to address a crucial challenge faced by supervised segmentation models: dependency on the large-scale labeled dataset. In this study, an unsupervised segmentation approach is proposed based on the concept of image inpainting to segment lung and head-neck tumors in images from single and multiple modalities. The proposed autoinpainting pipeline shows great potential in synthesizing high-quality tumor-free images and

outperforms a family of well-established unsupervised models in terms of segmentation accuracy.

Studies III and IV aim to automatically discriminate the benign from the malignant pulmonary nodules by analyzing the low-dose computed tomography (LDCT) scans. In Study III, a dual-pathway deep classification framework is proposed to simultaneously take into account the local intra-nodule heterogeneities and the global contextual information. Study IV seeks to compare the discriminative power of a series of carefully selected conventional radiomics methods, end-to-end Deep Learning (DL) models, and deep features-based radiomics analysis on the same dataset. The numerical analyses show the potential of fusing the learned deep features into radiomic features for boosting the classification power.

Study V focuses on the early assessment of lung tumor response to the applied treatments by proposing a novel feature set that can be interpreted physiologically. This feature set was employed to quantify the changes in the tumor characteristics from longitudinal PET-CT scans in order to predict the overall survival status of the patients two years after the last session of treatments. The discriminative power of the introduced imaging biomarkers was compared against the conventional radiomics, and the quantitative evaluations verified the superiority of the proposed feature set. Whereas Study V focuses on a binary survival prediction task, Study VI addresses the prediction of survival rate in patients diagnosed with lung and head-neck cancer by investigating the potential of spherical convolutional neural networks and comparing their performance against other types of features, including radiomics. While comparable results were achieved in intra-dataset analyses, the proposed spherical-based features show more predictive power in inter-dataset analyses.

In summary, the six studies incorporate different imaging modalities and a wide range of image processing and machine-learning techniques in the methods developed for the quantitative assessment of tumor characteristics and contribute to the essential procedures of cancer diagnosis and prognosis.

Keywords: Medical Image Analysis, Machine Learning, Deep Learning, Survival Analysis, Early Response Assessment, Tumor Classification, Tumor Segmentation

Sammanfattning

Cancer är en global hälsoutmaning som uppskattas ansvara för cirka 10 miljoner dödsfall i hela världen, bara under året 2020. Framsteg inom medicinsk bildtagning och hårdvaruutveckling de senaste tre decennierna har banat vägen för moderna medicinska bildgivande system vars upplösningsförmåga tillåter att fånga information om tumörers anatomi, fysiologi, funktion samt metabolism. Medicinsk bildanalys har därför fått en mer betydelsefull roll i klinikers dagliga rutiner inom onkologin, för bland annat screening, diagnostik, uppföljning av behandling samt icke-invasiv utvärdering av sjukdomsprognoser. Sjukvårdens behov av medicinska bilder har lett till att det nu på sjukhusen finns en enorm mängd medicinska bilder på alla moderna sjukhus. Med hänsyn till den viktiga roll medicinsk bilddata spelar i dagens sjukvård, samt den mängd manuellt arbete som behöver göras för att analysera den mängd data som genereras varje dag, så har utvecklingen av digitala verktyg för att för att automatiskt eller semi-automatiskt analysera bilddatan alltid haft stort intresse. Därför har en rad maskininlärningsverktyg utvecklats för analys av onkologisk data, för att gripa sig an läkares repetitiva vardagssysslor.

Den här avhandlingen syftar att bidra till fältet “onkologisk bildanalys” genom att föreslå nya sätt att kvantifiera tumörers egenskaper från medicinsk bilddata. Specifikt, är denna avhandling baserad på sex artiklar där de första två har fokus att presentera nya metoder för segmentering av tumörer, och de resterande fyra ämnar att utveckla kvantitativa biomarkörer för cancerdiagnostik och prognos.

Huvudsyftet för “Studie I” har varit att utveckla en djupinlärnings-pipeline vars syfte är att fånga lungpatologiers anatomier (inklusive lungtumörer) samt integrera detta med djupa neurala nätverk för segmentering för att nyttja det första nätverkets utfall för att förbättra segmenteringskvalitén. Den föreslagna pipeline testades på flertalet dataset och numeriska analyser visar en överlägsna resultat för den föreslagna “prior-medvetna” djupinlärningsmetoden. “Studie II” ämnar att ta sig an ett viktigt problem som övervakade segmenteringsmetoder ställs inför: ett beroende av enorma annoterade dataset. I denna studie föreslås en icke-övervakad segmenteringsmetod som baseras på konceptet “ifyllnad” (“inpainting”) för att segmentera tumörer i områdena: lungor samt huvud och hals i bilder från olika modaliteter. Den föreslagna metoden lyckas bättre än en familj väletablerade icke-övervakade segmenteringsmodeller.

“Studie III” och “Studie IV” försöker automatiskt diskriminera benigna lungtumörer från maligna tumörer genom att analysera bilder från LDCT (lågdos-CT). I “Studie III” föreslås ett djupt neuralt nätverk för klassificering vars grafstruktur tillåter lokal analys av tumörens inbördes heterogeniteter samt en helhetsbild från global kontextuell information. “Studie IV” försöker utvärdera noggrant utvalda metoder som grundar sig på att extrahera anatomiska särdrag från medicinska bilder. I studien jämförs konventionella “radiomics”-metoder med särdrag från neurala nätverk samt en kombination av båda på samma dataset. Resultat från studien visar att en kombination av särdrag från djupa neurala nätverk samt “radiomics” kan ge bättre resultat i klassificeringsproblemet.

“Studie V” har fokus på tidig bedömning av lungtumörers respons på behandling genom att utveckla ett set nya fysiologisk observerbara särdrag. Den presenterade metoden har använts för att kvantifiera förändringar i tumörers karaktär i PET-CT-undersökningar för att predicera patienters prognos två år efter senaste behandling. Metoden jämförts mot konventionella “radiomics” och utvärderingen visar att den föreslagna metoden ger förbättrade resultat. Till skillnad från “Studie V”, som fokuserar på att lösa ett binärt klassificeringsproblem, så försöker “Studie VI” predicera överlevnadsgraden hos patienter med lung- samt huvud och hals-cancer genom att undersöka neurala nätverk med sfäriska faltningsoperationer. Metoden jämförs mot, bland annat, “radiomics” och visar liknande resultat för analys på samma dataset, men bättre resultat för analys på olika dataset.

Sammanfattningsvis så utnyttjar de sex studierna olika medicinska bildgivande system samt en mängd olika bildbehandling- och maskininlärningstekniker för att utveckla verktyg för att kvantifiera tumörers egenskaper, som kan underlätta fastställande av diagnos och prognos.

Acknowledgments

In the Name of the Father

Planning of conducting a doctoral study initiated for me in 2011; a dream which shaped in my mind before becoming a Master's student! I remember the moment that Örjan called me over Skype on May 12th 2017, and kindly offered me the announced PhD position... I felt felicitous for achieving the fruit that I had put effort into for around six years. At that moment, I was confident that I could manage to surmount the challenges a doctoral study might introduce; however, what I missed to take into account was to brace myself for "Culture shock"! Leaving all my heart and soul with my dear ones behind, moving into another country, and diving into its distinct culture, lifestyle, and language were the factors that I did not consider, in advance.

I expected a doctoral study to start first by passing some courses, and gradually moving into the research topics. However, on my first day, August 8th 2017, I came to the group's office at Flemingsberg, and Chunliang transferred some multimodal image data to my temporary workstation straight to the point, and asked to directly start the first project. "Brainstorming, implementing, and failing" were the most frequent workflow of my doctoral journey. Turning the "failing" into a "functioning properly" as well as dealing with the homesickness could not be possible without all the great people who supported me in such challenges.

In the beginning, I would like to express my deep sense of gratitude to all my supervisors, **Örjan, Chunliang, and Iuliana**. I am genuinely grateful for your invaluable advice, continuous support, encouragement, and feedback that have had navigated me over my doctoral education. Along with the technical and pedagogical mentorship, I am incredibly thankful for your positive attitude during all our interactions which opened me space to flourish. Indeed, it provided me the opportunity to increase my confidence in developing my career as an independent researcher.

Next, a big thank goes to all my colleagues and fellow doctoral students at the Biomedical Imaging Division. I would like to thank Rodrigo for always being there when I needed your support. Fabian, I was happy of working with you in the same office. As a multi-skilled person, I learned a lot from you, and will miss the frequent loud laughing over open

discussions we had. A big salute to Irene for being a supportive colleague and friend. We had to take more courses together, like the deep learning one! Though it started a bit late, I am very excited regarding the collaboration we established, Simone. I was always impressed not only by your soft computing skills but also by your sense of humor! I will never forget your priceless help with the exercises of my favorite(!) course Jingru. Gabriel, the nerd! It was always fun to discuss with you the details of machine learning and model development. Many thanks for translating the abstract part of my thesis. Amirreza, the Jesus! In your few months of stay at KTH back in 2017, your determination and dedication to the research were inspirational to me. Dr. Mike, I am fortunate to have your friendship. I enjoyed listening to your enthusiasm for new adventures in practical ways. I also need to thank my former colleagues, including Yongjun and Martin. Here, I would like to give a special thanks to Daniel. Among all the good memories, the Spring of 2020, when we spent almost every evening in the seminar room on the 9th floor, formed the highlights of our friendship at KTH. I look forward to take a road trip in Canada with Simone.

During the last year of my Ph.D. study, I had the opportunity to visit a fabulous group at the Technical University of Munich, the CAMP, that is led by the distinguished scientist in the medical imaging field. I am very pleased by Nassir for hosting me in his group and including me in their group activities. I also would like to express my sincere appreciation to Thomas, the most proactive researcher I have ever met. Your boundless enthusiasm for supervision and research activities is a gift for all the IFLers. Although I could not work at the office regularly, I really enjoyed interacting with Francesca, Yousef, Azadeh, Ashkan, and Mohammad. I met Shahrooz, for the first time, in January 2012 at the University of Tehran. It was exciting to meet you in TranslaTUM, far from Amirabad.

Despite the fact that Stockholm is around 5000km far from Tehran, I feel less homesick because of my good Persian friends. Farhad and Shirin, I am very thankful for your supportive role in different aspects of my life in Sweden. I need to give thanks to Abdolamir, Reza, Pooya, and Morteza, as well.

Hossein, we started this journey together where we spent days and nights researching and working at SRBIAU, the national library, Arian tower, Sedkhandan, etc. Ramin, we have been into the most important events of our lives together, including our never-ending discussion on the top of Baargaah! There are quite many different reasons that I would like to give a very special heartfelt thank you to Narguess. Indeed, a part of my character is truly influenced by your attitudes.

Finally, I want to express my deepest gratitude to my family. I will be indebted forever to my mom, **Manzar Banoo**, and my late but eternal dad, **Gholamreza**, for being always there for me and providing me with the opportunities to follow my passions. I wish to always make you proud of my achievements, who I am and who I would become.

Mehdi Astaraki

Stockholm, 20 August 2022

List of Scientific Papers

This thesis is based on the following six scientific publications, which are referred to using Roman numerals. At the end of this thesis, all publications are appended in their published or prepared manuscript form:

- I. Prior-aware Autoencoders for Lung Pathology Segmentation**
M. Astaraki, Ö. Smedby, C. Wang
Medical Image Analysis, vol. 80, 2022; <https://doi.org/10.1016/j.media.2022.102491>
- II. Unsupervised Tumor Segmentation**
M. Astaraki, F. De Benetti, Y. Yeganeh, I. Toma-Dasu, Ö. Smedby, C. Wang, N. Navab, T. Wendler
Manuscript.
- III. Benign-malignant pulmonary nodule classification in low-dose CT with convolutional features**
M. Astaraki, Y. Zakko, I. Toma Dasu, Ö. Smedby, C. Wang
Physica Medica, vol. 83, p. 146-153, 2021; <https://doi.org/10.1016/j.ejmp.2021.03.013>
- IV. A Comparative Study of Radiomics and Deep-Learning Based Methods for Pulmonary Nodule Malignancy Prediction in Low Dose CT Images**
M. Astaraki, G. Yang, Y. Zakko, I. Toma Dasu, Ö. Smedby, C. Wang
Frontiers in Oncology, vol. 11, 2021; <https://doi.org/10.3389/fonc.2021.737368>
- V. Early survival prediction in non-small cell lung cancer from PET/CT images using an intra-tumor partitioning method**
M. Astaraki, C. Wang, G. Buizza, I. Toma-Dasu, M. Lazzeroni, Ö. Smedby
Physica Medica, vol. 60, p. 58-65, 2019; <https://doi.org/10.1016/j.ejmp.2019.03.024>
- VI. Spherical Convolutional Neural Networks for Survival Rate Prediction in Cancer Patients**
F. Sinzinger, **M. Astaraki**, Ö. Smedby, R. Moreno
Frontiers in Oncology, vol. 12, 2022; <https://doi.org/10.3389/fonc.2022.870457>

Division of Works Between Authors

Paper I

The main scientific hypothesis was jointly formulated by **Mehdi Astaraki (MA)** and Chunliang Wang (CW). The normal appearance autoencoder model was proposed by **MA**. **MA** designed the methodology, implemented the models, analyzed the data. Comments and feedback were provided by CW. **MA** drafted the original manuscript, followed by critical comments by Örjan Smedby (ÖS). **MA** handled the correspondence with the journal.

Paper II

The main scientific hypothesis was jointly formulated by **MA** and Thomas Wendler (TW) and Francesca De Benetti (F D-B) and Nassir Navab (NN). **MA** formulated the hypothesis regarding the autoinpainting model. **MA** designed the general study to be conducted. CW, Iuliana Toma-Dasu (I T-D), and ÖS provided clinical insights into the problem. F D-B and Yousef Yeganeh (YY) conducted supplementary experiments. **MA** designed the methodology, implemented the models, preprocessing pipeline, and analysis steps. Critical comments and feedback were provided by TW, F D-B, and NN. **MA** drafted the original manuscript, followed by comments from the co-authors.

Paper III

MA, CW, I T-D, and ÖS formulated the scientific hypothesis. **MA** designed the data processing pipeline. CW designed the data annotation software. Yousuf Zakko (YZ) annotated the imaging data. **MA** designed and implemented the preprocessing pipeline, deep learning models and the analysis steps. Technical and clinical feedback was provided by all the co-authors. **MA** drafted the original manuscript. All the co-authors provided critical feedback. **MA** handled the correspondence with the journal.

Paper IV

MA, CW, I T-D, and ÖS formulated the scientific hypothesis. **MA** designed the data processing pipeline. CW designed the data annotation software. YZ annotated the imaging data. **MA** designed and implemented the preprocessing pipeline, deep learning models, radiomics pipeline, and analysis step. Comments and feedback were provided by all the co-

authors. **MA** drafted the original manuscript. All the co-authors provided critical feedback. **MA** handled the correspondence with the journal.

Paper V

IT-D, CW, and ÖS formulated the main scientific hypothesis. **MA**, Giulia Buizza (GB), CW, and I T-D designed the overall research to be conducted. I T-D and Marta Lazzeroni (ML) were responsible for data transferring and guiding the clinical perspectives. **MA** designed the image processing, machine learning methodology, and statistical analysis steps with frequent input from GB, CW, and ÖS. GB preprocessed the image data. **MA** implemented the models and conducted all the image analysis and machine learning implementations. **MA** drafted the original manuscript, which was modified by all the co-authors. CW handled the correspondence with the journal.

Paper VI

Fabian Sinzinger (FS), Rodrigo Moreno (RM), and ÖS formulated the main research hypothesis. **MA** and FS formulated the hypothesis regarding the tumor segmentation part of the work. The segmentation part of the study was designed, implemented, and analyzed by **MA**. FS designed, implemented, and analyzed the preprocessing steps, spherical convolution and survival analysis methods. FS performed the survival analysis with different methods based on the tumor segmentation masks which were provided by **MA**. FS wrote the original draft of the manuscript, except for the image segmentation part that was written jointly by **MA** and FS. All co-authors provided feedback. FS handled the correspondence with the journal.

Other Scientific Contributions

Journal Papers

- **CTV Delineation for High-Grade Gliomas: Is There Agreement With Tumor Cell Invasion Models?**

W. Häger, M. Lazzeroni, **M. Astaraki**, I. Toma-Dasu
Advances in Radiation Oncology, vol. 7, 2022; <https://doi.org/10.1016/j.adro.2022.100987>

Peer-reviewed Conference Papers

- **Normal Appearance Autoencoder for Lung Cancer Detection and Segmentation**

M. Astaraki, I. Toma-Dasu, Ö. Smedby, C. Wang
Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, p. 249–256; https://doi.org/10.1007/978-3-030-32226-7_28

- **Multimodal Brain Tumor Segmentation with Normal Appearance Autoencoder**

M. Astaraki, C. Wang, G. Carrizo, I. Toma-Dasu, Ö. Smedby
International MICCAI Brainlesion Workshop, BrainLes 2019, p. 316–323;
https://doi.org/10.1007/978-3-030-46643-5_31

Peer-reviewed Conference Abstracts

- **Tumor Detection in PET/CT Using Multimodal Image Inpainting**

M. Astaraki, F. De Benetti, I. Toma-Dasu, Ö. Smedby, C. Wang, N. Navab, T. Wendler
Nuklearmedizin, 61(02): 164, 2022

- **Early survival prediction in non-small cell lung cancer with PET/CT size aware longitudinal pattern**

M. Astaraki, C. Wang, G. Buizza, I. Toma-Dasu, M. Lazzeroni, Ö. Smedby
European Society Radiation Oncology – ESTRO 2019, vol. 133, p. 208-209

- **Model based target definition for high-grade gliomas: implementation and sensitivity analysis**

W. Häger, M. Lazzeroni, **M. Astaraki**, I. Toma-Dasu
European Society Radiation Oncology – ESTRO 2020, vol. 152, p. 843-844

- **Diffusion model based target definition for high-grade gliomas: robustness analysis**

W. Häger, M. Lazzeroni, **M. Astaraki**, I. Toma-Dasu
European Society Radiation Oncology – ESTRO 2021, vol. 161, p. 1520-1622

Preprints

- **Development and evaluation of a 3D annotation software for interactive COVID-19 lesion segmentation in chest CT**

S. Bendazzoli, I. Brusini, **M. Astaraki**, M. Persson, J. Yu, B. Connolly, S. Nyrén, F. Strand, Ö. Smedby, C. Wang
arXiv:2012.14752, 2020

Contents

| | |
|--|-----------|
| Introduction | 1 |
| Research Aims | 5 |
| 2.1 <i>Prior-Aware Supervised Tumor Segmentation (Paper I)</i> | 5 |
| 2.2 <i>Unsupervised Tumor Segmentation (Paper II)</i> | 6 |
| 2.3 <i>Deep Features for Benign-Malignancy Classification (Paper III)</i> | 7 |
| 2.4 <i>Hybrid Imaging Biomarkers for Benign-Malignancy Classification (Paper IV)</i> | 7 |
| 2.5 <i>Imaging Biomarkers for Early Assessment of Tumor Response (Paper V)</i> | 8 |
| 2.6 <i>Imaging Biomarkers for Survival Rate Prediction (Paper VI)</i> | 9 |
| Background | 10 |
| 3.1 <i>Cancer Epidemiology</i> | 10 |
| 3.2 <i>Cancer Biology</i> | 10 |
| 3.3 <i>Cancer Diagnosis and Treatment</i> | 13 |
| 3.3.1 Medical Imaging for Cancer Diagnosis | 13 |
| 3.3.2 Cancer Treatment Methods..... | 18 |
| 3.3.3 Treatment Response Evaluation Methods | 21 |
| 3.4 <i>Imaging Biomarkers</i> | 23 |
| 3.4.1 Precision Medicine in Medical Imaging | 23 |
| 3.4.2 Development of Imaging Biomarkers..... | 25 |
| 3.4.3 Imaging Biomarkers and Quantitative Imaging..... | 26 |
| 3.4.4 Imaging Biomarkers and Medical Image Analysis | 27 |
| 3.4.5 Validation of Imaging Biomarkers | 29 |
| 3.5 <i>Radiomics</i> | 31 |
| 3.5.1 Radiomic Feature Analysis..... | 35 |
| 3.5.1.1 Feature Selection Methods | 36 |
| 3.5.1.2 Learning Algorithms..... | 38 |
| 3.6 <i>Deep Learning Methods for Medical Image Analysis</i> | 39 |
| 3.6.1 Deep Neural Networks | 40 |
| 3.6.2 Convolutional Neural Networks..... | 42 |
| 3.6.3 Deep Classification | 46 |
| 3.6.4 Deep Segmentation..... | 48 |
| 3.6.5 Deep Inpainting..... | 52 |
| 3.6.6 Deep Representation Learning | 58 |
| Methods | 62 |
| 4.1 <i>Deep Learning Methods for Tumor Segmentation (Papers I, II)</i> | 62 |
| 4.1.1 Overview of the Studies | 62 |
| 4.1.2 Image Preparation and Preprocessing | 63 |

| | | |
|---------------------|--|------------|
| 4.1.3 | Prior-Aware Tumor Segmentation | 64 |
| 4.1.4 | Unsupervised Tumor Segmentation | 69 |
| 4.2 | <i>Imaging Biomarkers for Cancer Diagnosis and Prognosis (Papers III, IV, V, and VI)</i> | 72 |
| 4.2.1 | Overview of the Studies | 73 |
| 4.2.2 | Image Preprocessing | 74 |
| 4.2.3 | Tumor Segmentation | 75 |
| 4.2.4 | Feature Set Development/Extraction | 76 |
| 4.2.5 | Benign-Malignancy Prediction | 83 |
| 4.2.6 | Survival Prediction | 86 |
| Results | | 88 |
| 5.1 | <i>Deep Learning Methods for Tumor Segmentation</i> | 88 |
| 5.1.1 | Paper I | 88 |
| 5.1.2 | Paper II | 90 |
| 5.2 | <i>Imaging Biomarkers for Cancer Diagnosis and Prognosis</i> | 93 |
| 5.2.1 | Paper III | 93 |
| 5.2.2 | Paper IV | 95 |
| 5.2.3 | Paper V | 97 |
| 5.2.4 | Paper VI | 99 |
| Discussion | | 101 |
| 6.1 | <i>Deep Learning and Tumor Segmentation</i> | 104 |
| 6.1.1 | Prior-aware segmentation model | 104 |
| 6.1.2 | Unsupervised segmentation model | 106 |
| 6.2 | <i>Imaging Biomarkers and Benign-Malignancy Classification</i> | 109 |
| 6.3 | <i>Imaging Biomarkers and Evaluation of Early Response to Therapy</i> | 112 |
| 6.4 | <i>Imaging Biomarkers and Survival Rate Prediction</i> | 114 |
| 6.5 | <i>Advances and Limitations of Machine Learning Methods in Cancer Screening</i> | 116 |
| 6.5.1 | Imaging datasets | 117 |
| 6.5.2 | Volumetric analysis | 118 |
| 6.5.3 | Incorporating domain knowledge | 118 |
| 6.5.4 | Methodology | 119 |
| Conclusions | | 121 |
| Bibliography | | 124 |

List of Abbreviations

| | |
|-------|--|
| 2D | Two-dimensional |
| 3D | Three-dimensional |
| Adab | Adaptive boosting |
| AE | Autoencoder |
| ANN | Artificial Neural Network |
| AUROC | Area Under the Receiver Operating Characteristic Curve |
| CAD | Computer Aided Diagnosis |
| CCC | Concordance Correlation Coefficient |
| CNN | Convolutional Neural Network |
| CPH | Cox Proportional Hazards |
| CT | Computed Tomography |
| DL | Deep Learning |
| DT | Decision Tree |
| FDG | Fluorodeoxyglucose |
| FFS | Forward Feature Selection |
| GAN | Generative Adversarial Network |
| Gconv | Gated convolution |
| KNN | K-nearest neighbor |
| LDA | Linear Discriminant Analysis |
| LDCT | Low-Dose Computed Tomography |
| LOO | Leave One Out |
| ML | Machine Learning |
| MLP | Multi-Layer Perceptron |
| MR | Magnetic Resonance |
| MRI | Magnetic Resonance Imaging |
| MSE | Mean Square Error |
| NAA | Normal Appearance Autoencoder |
| NSCLC | Non-Small Cell Lung Cancer |
| PCA | Principal Component Analysis |
| PCNN | Partial Convolutional Neural Network |

| | |
|---------|--|
| Pconv | Partial Convolution |
| PERCIST | Positron Emission Tomography Response Criteria in Solid Tumors |
| PET | Positron Emission Tomography |
| PSNR | Peak Signal-to-Noise Ratio |
| QDA | Quadratic Discriminant Analysis |
| QIB | Quantitative Imaging Biomarkers |
| RECIST | Response Evaluation Criteria in Solid Tumors |
| RF | Random Forest |
| ROI | Region Of Interest |
| SALoP | Size-Aware Longitudinal Pattern |
| SMOTE | Synthetic Minority Over-sampling Technique |
| SphCNN | Spherical Convolutional Neural Network |
| SSIM | Structural Similarity Index |
| SUV | Standardized Uptake Value |
| SVM | Support Vector Machine |
| UAD | Unsupervised Anomaly Detection |
| VAE | Variational Autoencoder |

Chapter 1

Introduction

Over the past two decades, the impact of medical imaging technologies in clinical oncology has been substantially expanded from primarily diagnostic tools to a more central role in the domain of individualized medicine. Medical imaging contains invaluable information which can be used as complementary data with respect to other clinical resources such as genomics, pathology, and blood biomarkers. The combination of these data resources can provide unprecedented information that can improve individualized monitoring and therapy methods [1]. Furthermore, in contrast to genomic and proteomic examinations, which require sampling from the heterogeneous tumor masses through invasive biopsies, medical imaging can provide a more comprehensive view of the entire tumor in a non- or minimally-invasive procedure. Accordingly, medical imaging modalities play a key role in the different stages of cancer screening, including cancer diagnosis, treatment and surgery planning, cancer characterization, cancer monitoring, and tumor response evaluation.

Impressive innovations in the medical imaging hardware and imaging agents on one side and standardizing image acquisition protocols on the other side have led to capturing high-resolution quantitative images, which can be used in a variety of oncology-related applications. However, such widespread cancer screening facilities resulted in the accumulation of a large number of medical images that need to be manually examined by clinical experts. To reduce the heavy burden on clinicians, numerous computerized methods have been developed to assist clinicians with different tasks. These tools are often built upon image processing techniques, statistical methods, and Machine Learning (ML) models to automatically or semi-automatically draw inferences from medical images.

Regardless of the type of tumors, the oncological image analysis pipeline consists of the following general steps: (1) image preprocessing, (2) target tumor detection/segmentation,

(3) quantitative feature extraction, and (4) diagnosis, prognosis, or prediction. (See Figure 1.1.)

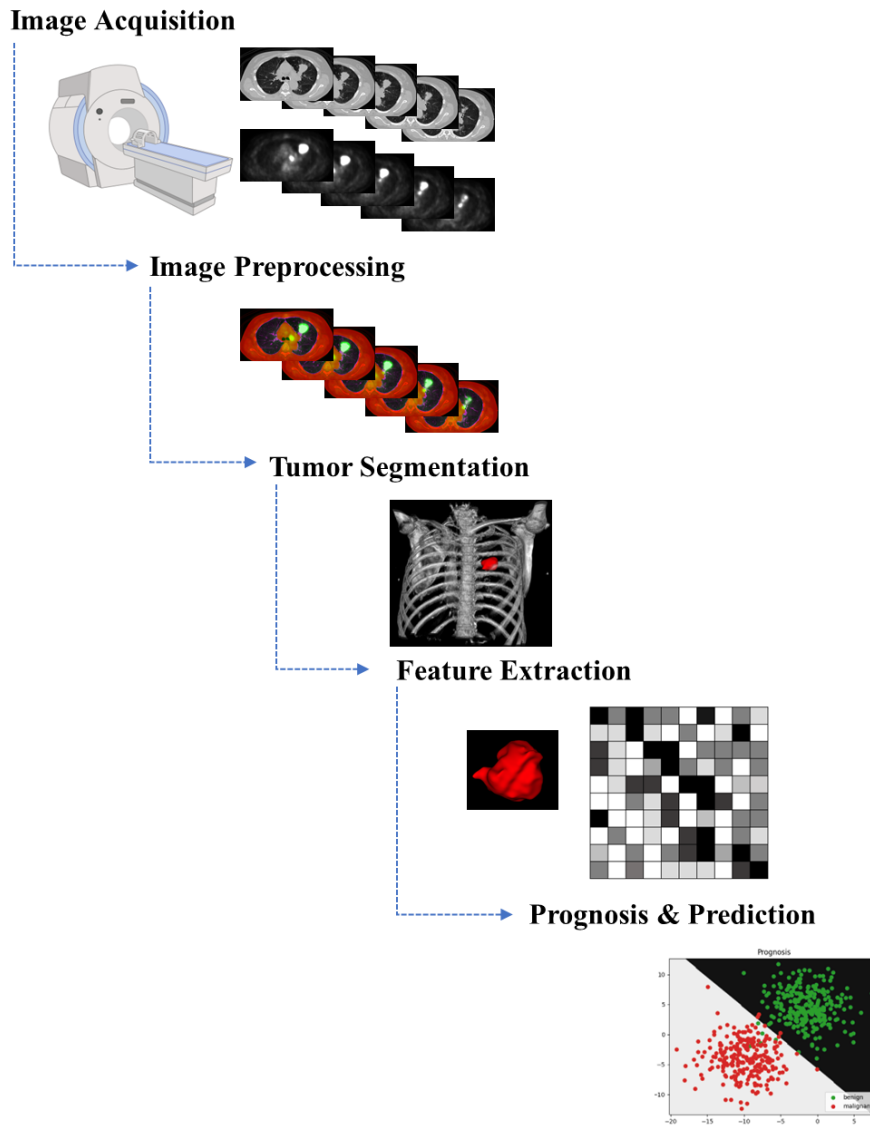


Figure 1.1. Schematic illustration of the oncological image analysis pipeline. Standard image preprocessing techniques are applied to the acquired images to prepare them for subsequent analyses, including tumor segmentation. Delineated tumors are then represented by quantitative features, which will be used to infer a decision. (A part of the figure was created with Biorender.com.)

Image preprocessing techniques can increase the perceptibility to the human observer of the anatomical structures and tissues within the images, or they can be considered as preparation steps by which the subsequent analysis can be applied. Typical preprocessing techniques include, but are not limited to, contrast and edge enhancement, noise reduction, artifact reduction, spatial resolution resampling, image registration, and fusion of multiple modalities [2].

Tumor segmentation, identifying the voxels of cancerous regions from background structures, is considered to be one of the most challenging tasks in medical image analysis due to the heterogeneous characteristics of the tumors. In fact, even two tumors within the same organ can maintain significantly different characteristics—a phenomenon referred to as *inter-tumor* heterogeneity. Moreover, large variability exists among the cancer cells within a single tumor of the same patient, which is referred to as *intra-tumor* heterogeneity. As a result, tumors appear with a variety of textural patterns in different sizes at various locations with similar intensity and textural patterns as their surrounding healthy structures. Numerous segmentation methods have been proposed over the last three decades for segmenting different types of tumors in different imaging modalities, such as region growing [3], graph cuts [4], active contours [5], and hidden Markov random field [6]. Due to the rapid development of DL techniques, medical image segmentation no longer requires handcrafted features because Convolutional Neural Networks (CNNs) can efficiently learn hierarchical feature representation directly from the images. As a result, outstanding segmentation performances have been achieved for different types of tumors such as lung [7], brain [8], and liver [9].

Quantitative feature extraction refers to transferring the raw voxel intensities of the segmented tumor volumes into quantitative attributes, which may represent certain characteristics of the tumors. In other words, it has been hypothesized that medical imaging provides critical information regarding tumor phenotypes, which could be exploited by extracting and mining the quantitative features [10]. This is the case because image-based quantitative features may reflect the general characteristics of cancers, such as proliferation, angiogenesis, metabolism, or agnostic features [11].

Diagnosis, prognosis, or prediction step aims at employing the extracted quantitative features in statistical or ML models to infer a decision such as benign-malignancy classification [12], lymph node metastasis prediction [13], overall survival prediction [14], cancer stage classification [15], and cancer recurrence prediction [16].

This thesis aims at contributing to the field of oncological image analysis by covering all the mentioned steps and proposing new ways to quantify the tumor characteristics from the medical image data. Specifically, this thesis consists of six studies, the first two of which introduce novel models for tumor segmentation; the last four studies focus on developing quantitative features for cancer diagnosis and prognosis. From a technical perspective, the presented works in this thesis pursue two objectives: first, to integrate the clinical knowledge into the well-established image processing and ML tools in order to improve the efficacy of

cancer diagnosis and prognosis methods; second, to identify the limitations of the currently available methods and to propose new approaches for addressing the recognized limitations in order to improve the performance of the computerized methods.

The structure of this thesis is organized as follows. In Chapter 2, the underlying motivation and aim of each of the six studies are specified. Chapter 3 consists of brief descriptions regarding the background theories of the methods employed in the six studies. Next, the methods and results of each of the six studies are outlined in Chapters 4 and 5. Analytical reasoning of the results is then presented in Chapter 6, followed by conclusions in Chapter 7. Lastly, the relevant papers of each of the six studies are appended at the end of this thesis. It should be noted that the two terms “study” and “paper” are used interchangeably in the rest of this thesis.

Chapter 2

Research Aims

Visually inspecting and subjectively assessing the medical images, which expert clinicians do in daily practice, is recognized as a demanding and error-prone task. The overall aim of the present doctoral thesis is to strengthen the important role of computerized image analysis techniques for diagnosis, prognosis, and prediction purposes, with a focus on oncological images in general and lung cancer images in specific. This is done by employing advanced image processing and ML techniques, modifying them, and introducing new technical contributions to the medical image analysis community with particular emphasis on the development of clinically relevant Quantitative Imaging Biomarkers (QIBs).

This thesis is built upon six scientific papers, which are appended to the end of this thesis. It should be noted that the order of the papers was chosen to match it with the pipeline of oncological image analysis, which starts with tumor segmentation and follows with quantification and prognosis prediction. The underlying motivations and research aims of each study are outlined in the following sections.

2.1 Prior-Aware Supervised Tumor Segmentation (Paper I)

Motivation: Segmentation of anatomical organs and/or pathologies is recognized as one of the central tasks in medical image analysis. In fact, a precise and accurate segmentation of the target region can provide important information that would be used for further diagnosis, prognosis, or treatment steps. In the lungs, precise segmentation of lung pathologies plays a vital role in different applications. For instance, accurate detection and segmentation of pulmonary nodules can be further used in the follow-up sessions to track the growth of the nodules, which is associated with nodule malignancy. In addition, to deliver the effective radiation dose to the lung tumors, a precise segmentation of the tumor boundaries is required for radiotherapy planning procedures. Gaining from the hierarchical feature learning directly from the input images through the convolutional layers, Convolutional Neural Networks (CNNs) have shown promising results in various medical image segmentation tasks. In this domain, the encoder-decoder U-Net architecture [17] is

considered a milestone as an abundance of the recent advances in segmentation networks have been inspired by the plain U-Net model. A segmentation prior encodes certain types of prior knowledge, such as the shape and appearance of a target region, into a segmentation model, which has been recognized as a helpful strategy to improve segmentation accuracy [18]. Various methods have been proposed to obtain prior knowledge and integrate it into the segmentation models; however, these methods are mainly designed to deal with organ segmentation. In fact, obtaining prior knowledge about pathological regions is a more challenging task due to the stochastic nature of the pathologies.

Aim: *To propose a deep autoencoder-based pipeline for obtaining prior knowledge of lung pathologies in CT images in order to improve the segmentation accuracy of pathological regions.*

2.2 Unsupervised Tumor Segmentation (Paper II)

Motivation: Due to rapid advances in the field of DL, the performance of the medical image segmentation tasks has improved considerably. Despite the promising results achieved by deep segmentation models, which could reach clinical-level accuracy for some applications, their feasibility is limited by the required large amount of carefully labeled data due to their supervised training fashion. However, supervised training of such data-greedy models suffers from two types of limitations: (a) the number of training medical images is often limited because of costly slice-by-slice pixel-level annotation; and (b) even if a large-scale labeled data is available, the generalization power of the model over the unseen classes is poor [19]. Unsupervised DL methods tend to be a natural fit for gaining insights into medical image analysis tasks as their optimizations do not entail labeled datasets. In this domain, Unsupervised Anomaly Detection (UAD) is an active field of research that aims to identify the data that do not fit into the learned distribution from normal data [20]. Despite the encouraging results achieved by UAD methods in some applications, these models suffer from several limitations. First, they often face difficulties in learning the distribution of healthy anatomies from high-resolution images. Second, these methods are often employed to detect anomalies with different intensity patterns with respect to nearby normal tissues, such as glioma and MS lesions, in specific sequences of MR images [21]. A third (and significant) limitation of the current UAD techniques is the difficulty of preserving the anatomical constraints within the generated images.

Aim: *To develop an autoinpainting-based UAD method for segmenting tumors in full resolution single/multi-modal images.*

2.3 Deep Features for Benign-Malignancy Classification (Paper III)

Motivation: Lung cancer is notorious for high morbidity and mortality worldwide. Most lung cancer patients are diagnosed with advanced stages of lung cancer. Although the five-year survival rate of locally advanced lung cancer is less than 5%, early diagnosis of asymptomatic lung cancer plays a vital role in treatment planning that can substantially improve the overall five-year survival by up to 60% [22]. Most lung cancers emerge from the tiny malignant pulmonary nodules even though the majority of lung nodules belong to the benign class. However, distinguishing malignant from benign modules is one of the challenging steps in lung cancer screening as they share highly similar visual characteristics. In clinical practice, radiologists examine the CT volumes slice-by-slice to estimate the malignancy likelihood by measuring the morphological attributes of the nodules. Moreover, it has been shown that intensity distribution and relative position of the nodules are strongly associated with lesion malignancy [23]. Therefore, the classification of malignant lung nodules from benign nodules in CT images by developing computerized methods is feasible and could assist clinicians and support their clinical decisions for potential interventions.

Aim: *To develop a robust method for automatic benign-malignant nodule classification in CT volumes by simultaneous analysis of intra-nodule heterogeneity and contextual features.*

2.4 Hybrid Imaging Biomarkers for Benign-Malignancy Classification (Paper IV)

Motivation: To classify malignant lung nodules from benign nodules, a large number of computerized models have recently been proposed. Taking a general view, these models can be categorized into two groups: radiomics-based and DL-based models. Radiomic descriptors are designed to quantify the nodule characteristics such as morphological attributes and textural features. The extracted feature set is often processed to identify the radiomic signatures—the most informative features—which are then used to train a conventional learning algorithm such as random forest and support vector machines [24]. Therefore, the radiomics pipeline is constructed based on several independent steps. By contrast, DL-based models provide an end-to-end structure to directly learn the nodule characteristics from the input CT images, process the features, and classify them into benign or malignant classes—all in one single framework. Even though radiomics and DL methods are fundamentally different approaches, both have shown promising and comparable results in predicting lung nodule malignancy from CT images. However, there is a lack of evidence

in the literature to determine which method can potentially perform better than the others in a specific condition, as most of the previous studies were conducted on different datasets. This makes it challenging to compare them objectively.

Aim: *To compare the classification performance of a series of carefully selected conventional radiomics methods, end-to-end DL methods, deep feature-based methods, and the potential of fusing radiomic and deep features, all processed and analyzed with similar fine-tuning techniques.*

2.5 Imaging Biomarkers for Early Assessment of Tumor Response (Paper V)

Motivation: Considerable attention has been directed in oncology communities to evaluating the early response of tumors to the applied therapy as it can be helpful to adapt the treatment and, potentially, improve patient clinical outcomes. In clinical practice, invasive biopsy sampling and/or non-invasive medical imaging can be used for tumor response assessment. Size-based metrics such as the World Health Organization’s (WHO) criteria or Response Evaluation Criteria in Solid Tumors (RECIST) are the standard strategies to assess the tumor response from medical image data [25]. Nevertheless, there are several scenarios in which such metrics would not be precise. For instance, there are certain types of cancers in which the tumors may not shrink significantly despite the supposedly effective treatment applied, such as lymphoma, sarcoma, hepatomas, mesothelioma, and gastrointestinal stromal tumors [26], [27]. In addition, some chemotherapy regimens may result in tumor hemorrhage, necrosis, or cavitation, which does not necessarily represent tumor shrinkage. Accordingly, effective changes in structure, functionality, or metabolic activity of a tumor caused by treatment may not be captured accurately by using only size-based measurements [28]. By converting the image data into quantitative descriptors, numerous studies over the past few years have focused on employing/developing radiomic features as QIBs for a variety of applications such as early tumor assessment, distance metastasis, and survival prediction. However, a major challenge faced by the radiomics field is the difficulty of establishing clinical relevance between the extracted features and the biological characteristics of the tumors. In this domain, other challenges include sensitivity to imaging acquisition parameters, sensitivity to image artifacts, and model overfitting issues.

Aim: *To develop imaging biomarkers that are physiologically interpretable by taking into account the concept of tumor heterogeneity in order to predict the long-term survival*

status of lung cancer patients by assessing the early response of tumors to the applied treatments.

2.6 Imaging Biomarkers for Survival Rate Prediction (Paper VI)

Motivation: The term “survival analysis” refers to the study of the time period it takes for an event to occur. In the domain of cancer survival, this analysis can be applied to different applications, the most common of which correspond either to the time between diagnosis and death caused by the disease or the time between the applied treatments and the cancer recurrence. Ideally, all the patients in survival studies would be followed up with until the target event is recorded; nevertheless, patients are often lost in the clinical follow-ups prior to the event of interest. In some cases, the event may not even occur at all if the patient dies from a different cause. Therefore, when the event is not observed, the last contact time point is referred to as “censoring time.” Due to the presence of censored data, survival analysis cannot technically be considered an ordinary regression task. One of the most common classical statistical approaches for modeling survival data with censored observations is the Cox Proportional Hazards (CPH) model [29]. However, the CPH method suffers from high-dimensional data, and regression cannot learn any complex non-linear functions. In clinical trials, survival analysis is often conducted using clinical data such as molecular profiling, which are not always available. As a potential alternative, QIBs such as radiomics have been used in CPH and other ML-based methods [30]. In recent years, the potential of DL-based models, mainly CNNs, for survival analysis has been widely studied. One major drawback of the CNN-based models is that the learned features strongly depend on the spatial orientation of the tumoral regions. In other words, a rotated tumor can potentially result in a different survival prediction when using traditional CNNs.

Aim: *To propose an automatic pipeline for survival analysis of lung and head-neck cancer patients using spherical convolutional neural networks as rotationally invariant models.*

Chapter 3

Background

3.1 Cancer Epidemiology

Cancer remains among the leading causes of death worldwide and is an important barrier to increasing life expectancy. The estimated number of new cancer cases worldwide in 2020 was 19.3 million, with an estimated 10 million mortalities [31]. Moreover, the number of new cases is expected to increase by 20% in 2040 [32], [33].

Since a major part of this thesis focuses on lung cancer image analysis, general concepts of lung cancer epidemiology are briefly outlined. Among all types of cancers, lung cancer is recognized as the second most frequently occurring cancer after breast cancer. In 2020, 11.4% of all newly diagnosed cancer cases were lung cancer, which equals 2.2 million new cases. More importantly, approximately 1.8 million deaths or 18% of all cancer mortality are caused by lung cancer, making it the most lethal type of cancer [31]. Data collected between 2010 and 2014 show that most patients are diagnosed with advanced stages of lung cancer; therefore, survival after five years from the initial diagnosis varies between 10 to 20% in most countries [34]. By contrast, early diagnosis of asymptomatic lung cancer plays a critical role in treatment planning that can remarkably improve the overall five-year survival by up to 60% [22].

Identifying the underlying reasons for such a rapid growth of incidence and mortality is quite challenging, but it likely reflects the effects of aging, growth of population, and prevalence of the main risk factors of cancer, most of which are associated with socioeconomic development.

3.2 Cancer Biology

It was a bone of contention whether cancer is one single disease or a set of various diseases. In 2000, Hanahan and Weinberg published a distinguished review, the “hallmarks of cancer” [35], in which they asserted that the highly complex characteristics of the cancer

mechanism could be simplified into a small number of underlying principles. In that review, the authors argued that all types of cancers share six distinctive and complementary biological capabilities—that is, *hallmarks*—that control the formation of malignant cells from normal ones. These hallmarks are briefly explained below [35].

Self-sufficiency in growth signals. To transit from a quiescent state into an active proliferative state, normal cells inevitably need mitogenic growth signals that are transmitted into the cell by transmembrane receptors. However, cancer cells are capable of growing without such external signals; in fact, cancer cells can produce these signals by themselves. They can either permanently activate the signaling pathways to respond positively to such activations or impair the factors that are responsible for preventing the excessive growth of such signals.

Insensitivity to antigrowth signals. Within normal tissues, several antiproliferative signals operate to maintain cellular quiescence and tightly control cell division procedures. These antigrowth signals are regulated by proteins called Tumor Suppressor Genes (TSGs). Incipient cancer cells attack the antigrowth signals; therefore, the altered TSGs in cancer cells will not be able to prevent over-division of the cells.

Evading apoptosis. The ability of the tumor cells to expand their numbers depends not only on the rate of cell proliferation but also on cell attrition. Normal cells have a mechanism named “apoptosis,” by which the cells are programmed to die automatically in the event of damage(s). Cancerous cells, on the other hand, alter this mechanism by impairing the proper signaling so that the apoptosis mechanism cannot be activated.

Limitless replicative potential. All three steps mentioned above lead to detaching a cell’s growth program from the other signals in its environment. Normal cells have a limited number of divisions so that they cannot divide indefinitely (i.e., senescence state) due to the existence of a DNA named “telomeres.” The length of the telomeres is diminished in the wake of every cell division until they become too short to activate senescence. By manipulating specific enzymes to increase the length of telomeres, cancer cells bypass this obstacle and will be able to continue dividing indefinitely.

Sustained angiogenesis. The oxygen and nutrients delivered by the circulatory system are vital nutrients for the functioning and survival of the cells. Therefore, the formation of normal tissue is followed by the growth of new blood vessels—the process of angiogenesis. Like normal tissue, an expanding tumor needs oxygen and other nutrients from the blood

vessels as well. Cancerous cells can exploit this normal functionality for their own benefits. As a result, cancer cells are capable of regulating and promoting the formation of new blood vessels.

Tissue invasion and metastasis. Most primary tumors are prone to invade nearby tissues or even spread into distant organs where they may succeed in establishing new colonies. This process is called “metastasis” and accounts for 90% of cancer mortalities. This is a multistep process that begins with invading nearby tissues and then continues into the blood vessels. After surviving the circulatory system, cancer cells leave that system and start dividing inside a new tissue or organ.

Eleven years after the original publication of cancer hallmarks, in 2011, the authors published the next generation of hallmarks in which they added two more hallmarks as well as two emerging characteristics (see Figure 3.1). This update clearly indicates that as cancer researchers gain more insight into cancer biology, more complexity of cancer phenotypes and genotypes will be revealed [36], [37]. In 2022, the same author published the new dimensions of cancer hallmarks [38] in which he listed some of the limitations of the current hallmark traits that cannot address the complexity of cancers, presenting several prospective new hallmarks and enabling characteristics.

While all types of cancers are characterized by an uncontrolled rate of cell division and the potential to invade nearby or even distant tissues, they may hold varying levels of heterogeneity in phenotyping features related to the cancer hallmarks. Today, cancer subtyping is done based on organs, tissues, or the type of cells from which the cancer initiated. However, even two cancers from the same organ can maintain significantly different tumor characteristics. This inter-tumor heterogeneity is integrated by large variabilities among the cancer cells within a single tumor of the same patient (intra-tumor heterogeneity). Heterogeneity within the tumors and biological behavior corresponding to the cancer hallmarks have been proposed as potential reasons for treatment failures. Accordingly, the concept of heterogeneity has become an essential field of research in the context of personalized medicine [39], [40].

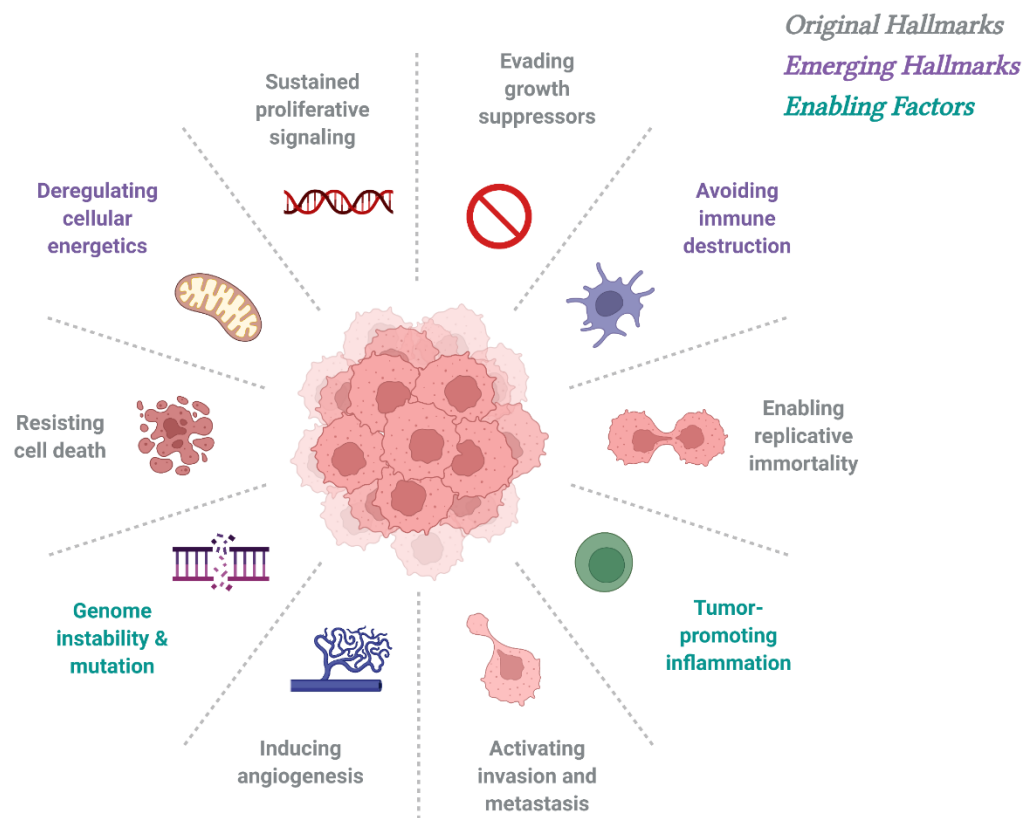


Figure 3.1. Illustration of cancer hallmarks. In addition to the original six hallmarks, two more hallmarks were identified as factors involved in the pathogenesis of cancers. One hallmark allows cancer cells to evade immunological destruction while the other involves the ability to modify or reprogram cellular metabolism. In addition, two consequential characteristics of neoplasia facilitate the acquisition of original and emerging hallmarks. (The figure was created with Biorender.com)

3.3 Cancer Diagnosis and Treatment

3.3.1 Medical Imaging for Cancer Diagnosis

The history of medical imaging modalities in the oncology field can be summarized as the elaboration of advanced technologies to project the human anatomies into an image plane in which details of the disease can be seen. Cancer imaging techniques have recently developed far beyond the initial goal of pathology detection. In fact, modern imaging modalities make it possible to characterize the disease structurally, functionally, physiologically, and biochemically. Therefore, medical imaging modalities have become an integral part of the diagnosis, prognosis, planning, and monitoring of cancer diseases [41]. Based on the nature of image acquisition, the imaging modalities can be broadly categorized into the following groups: X-ray-based modalities, Magnetic Resonance (MR) modalities, nuclear medicine imaging modalities, and ultrasound-based systems. Figure 3.2 depicts examples of different imaging modalities in cancer screening. While the physics of medical

imaging falls outside the scope of this thesis, a brief overview of some applications of the imaging systems in cancer screening will nonetheless be outlined below. This overview focuses on the first three modalities mentioned above, as they have been employed in the different research studies of this Ph.D. study.

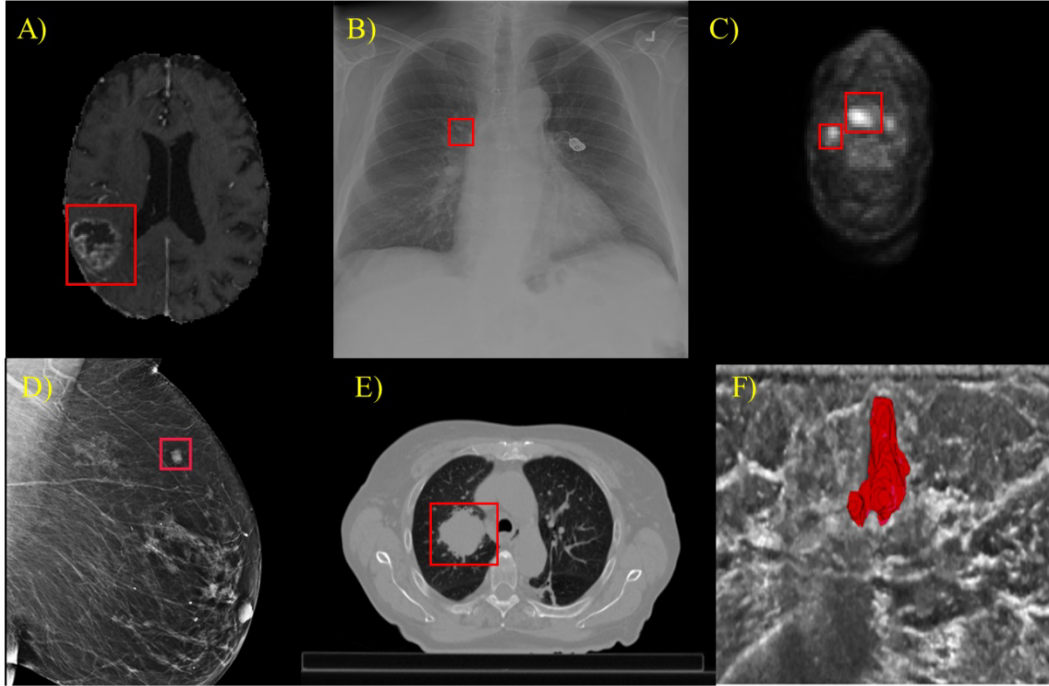


Figure 3.2. Different imaging modalities are used for screening different types of tumors. A) imaging glioblastoma with MRI modality; B) lung cancer detection in chest X-ray; C) imaging head-neck tumor with PET, i.e., nuclear medicine modality; D) breast cancer detection in digital mammography; E) lung cancer screening with CT imaging; and F) breast tumor segmentation in 3D ultrasound images.

X-ray-based modalities. The advances in digital imaging technology and the widespread use of Picture Archiving and Communications Systems (PACS) have extended the X-ray-based modalities in cancer imaging. Dual-energy X-ray machines can use two stacked detectors separated by a plate made of copper. These machines can use one X-ray exposure or one detector with dual X-ray exposure. Accordingly, they can acquire images of low and high-energy X-rays together so that they can obtain both soft-tissue images and hard-tissue scans such as bone and calcium. From the bone images, for example, calcifications in hilar lymph nodes can be captured. As another example, in conventional radiographic images of the chest region, rib defects such as sclerotic metastases or calcified pleural plaques share similar attributes with soft-tissue abnormalities; however, such defects may be accurately characterized by bone images. Another application of X-ray systems is related to full-field digital mammography, which acquires images from dense tissues of the breast with a high dynamic range of intensities, thus leading to diagnosis with

high sensitivity. A limitation of 2D mammograms is the possibility of missing the cancer regions as the normal structures (such as glandular tissue) would overlap and obscure tumoral regions. To tackle this limitation, the tomosynthesis technique can reconstruct a 3D volume of the breast by acquiring multiple low-dose X-rays taken from different angles. Malignant tumors often exhibit a rapid wash-in and wash-out of iodine, while benign tissues have a slow iodine uptake. When this characteristic is kept in mind, contrast-enhanced mammography with tomosynthesis can be used to distinguish malignant breast tumors from benign ones by following the distribution of the injected contrast-enhancing materials [42].

Computed Tomography (CT) machines using rotating X-ray tubes can reconstruct high-resolution volumetric images of up to 0.5 mm of slice thickness [43]. Among many other applications, this is the standard imaging modality for lung cancer screening. While identifying the lung nodules in CT images is not an easy task, distinguishing the malignant nodules from the benign ones is even more challenging (see Figure 3.3). As a solution, dynamic contrast-enhanced CT can be employed to identify malignant pulmonary nodules having increased vascularity due to angiogenesis [44]. Other applications of CT in cancer imaging include but are not limited to CT virtual colonography, CT perfusion imaging, and whole-body CT that can be used for screening colorectal cancer, metastases to lymph nodes and liver, and evaluating the therapy response as well [45].

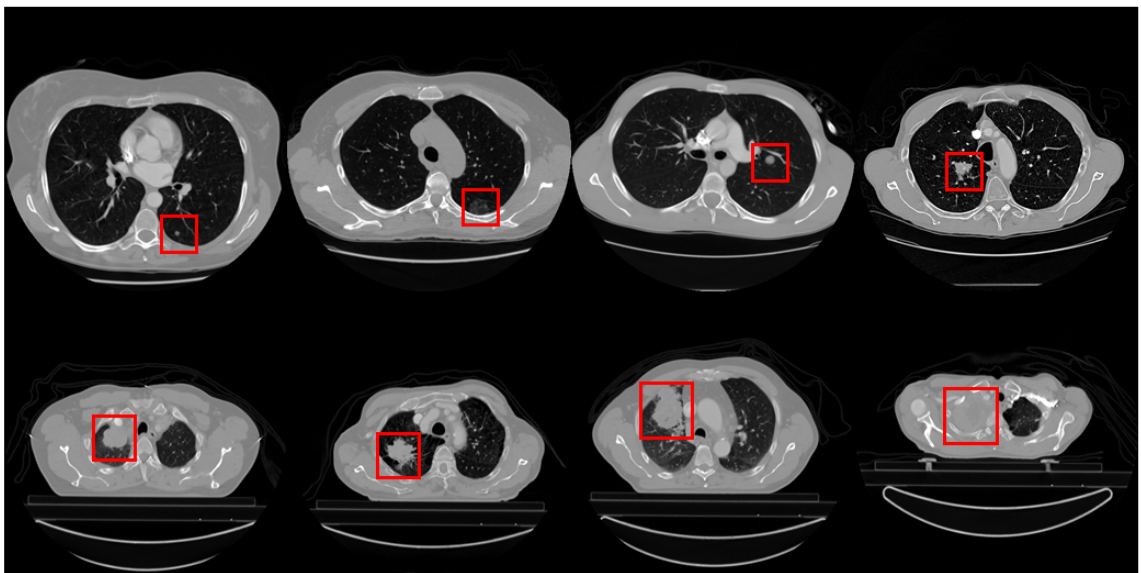


Figure 3.3. Lung cancer can appear in various sizes and with various characteristics, from tiny pulmonary nodules and ground glass opacities to extremely large-size non-small lung cancers.

Magnetic Resonance Imaging Modalities. Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique that can be used for cancer screening in different clinical

situations, including cancer detection, cancer staging, response evaluation, biopsy, and therapy guidance interventions. Structural MR imaging of the breast tissues can acquire high-resolution image data from chest wall involvement, multi-focal tumors, and lymph node metastasis. Therefore, it is possible to identify the stage of breast cancer with this modality. Structural MRI sequences play an essential role in brain tumor screening. Multi-modal scans can be acquired by using standard sequences, including Fluid-Attenuated Inversion Recovery (FLAIR), pre-gadolinium T1-weighted, and post-gadolinium T1-weighted (see Figure 3.4). The primary roles of structural MRI in brain tumor evaluation include detection and localization of the lesion(s), which can be further used for treatment and biopsy planning, and evaluating the mass effect on the brain tissues, ventricular system, and vasculatures.

Dynamic features of wash-in and wash-out can be examined effectively by using contrast-enhanced MRI images. Gadolinium is a paramagnetic element that changes the magnetic state of hydrogen atoms in water molecules. Tissues with a high level of gadolinium uptake in T1-weighted images appear with hyperintensity signals. The effect of gadolinium concentration is maximized as the first pass of a bolus of contrast agent after the rapid intravenous injection. On T2-weighted images, this causes hypointensity signals in areas of tissue that are highly perfused. In breast cancer screening, dynamic contrast-enhanced MR images with gadolinium agents are used to examine neo-angiogenesis, which is correlated with histopathological examinations [46]. Contrast-enhancing agents can indicate the local breakdown of the blood-brain barrier, which is a common characteristic in many brain tumors. Hyperintensity signals caused by contrast-enhancing agents within gliomas are associated with high-grade tumors. In addition, peri-tumoral edema appears with hyperintense signals in T2-weighted/FLAIR sequences surrounding the tumor cores.

Diffusion-Weighted Imaging (DWI) measures the diffusion of water molecules and is a promising technique for tumor detection and metastasis identification. Water molecule diffusion in tumoral tissues does not function normally; therefore, a lower Apparent Diffusion Coefficient (ADC) leads to high signal acquisition in DWI images. Accordingly, increases in the values of ADC represent a positive response to the applied treatment. In fact, such an increase in ADC is associated with the number of cells killed and is thought to relate to the liberation of water into the extracellular space as a result of cell necrosis [47]. DWI-based ADC values have been shown to be decreased in highly cellular tumors such as lymphoma and high-grade glioma [48].

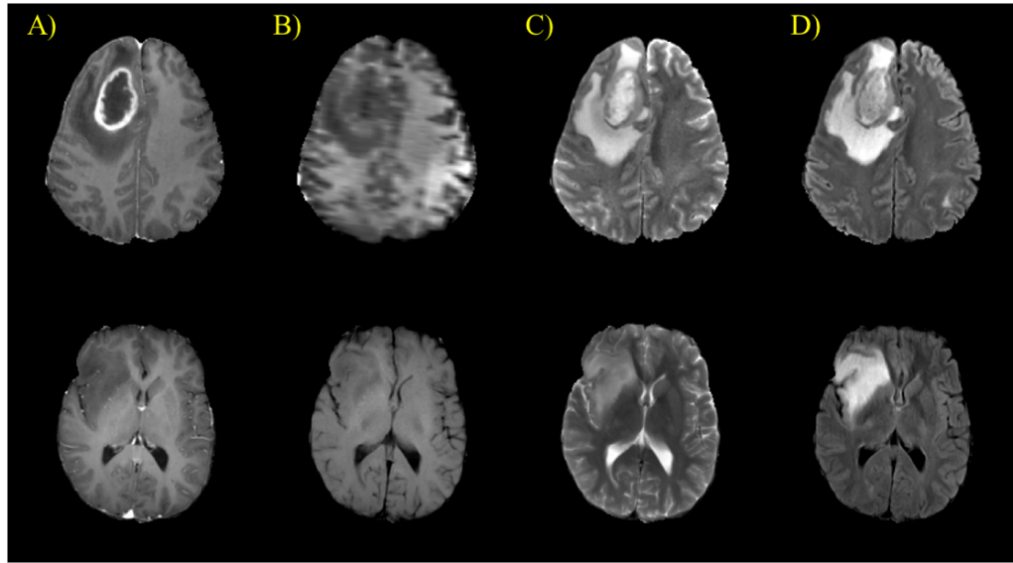


Figure 3.4. Different MRI sequences are examined to identify different brain tumor subregions: A) T1 gadolinium enhancement sequence; B) T1-sequence; C) T2-sequence; and D) FLAIR sequence. High-grade gliomas (first row) and low-grade gliomas (second row) appear with significantly different characteristics.

Nuclear medicine imaging modality. These imaging modalities use small amounts of radiopharmaceutical materials that are typically injected into the bloodstream. In this category, Positron Emission Tomography (PET) is a radiopharmaceutical-based imaging modality that can quantify in vivo cellular and molecular alterations. Whether as a single modal or a multimodal (e.g., PET-CT) imaging technique, it has exerted a significant impact on clinical practice and translational cancer research (see Figure 3.5). PET modality is capable of characterizing the tumor properties based on biochemical changes at the molecular levels. The most common type of radiotracer used with PET imaging is fluorodeoxyglucose (FDG) because it accumulates mostly in cancer cells to a greater degree than the non-cancerous cells. It should be emphasized that the rate of false positives on FDG-PET examinations is relatively high due to the fact that the increased level of glycolysis is not limited only to the cancer cell. The FDG-PET scan has been successfully employed for a variety of cancers, including lung, breast, head and neck, colorectal, lymphoma, melanoma, thyroid, cervical, pancreatic, and esophageal. Besides accurate cancer detection, another important application of the FDG-PET scan is to monitor the impact of the applied treatment and assess the response of the tumor cells to the therapies. It has been shown that comparing the Standardized Uptake Value (SUV) [49] among the pretreatment and follow-up scans can act as a potential biomarker. SUV is a mathematically derived ratio of tissue radioactivity concentrations at a certain time and the injected radiopharmaceutical dose, normalized by the patient's weight. This measure is a semi-quantitative value that can be influenced by different factors such as low image resolution, image noise, and the bias

imposed by the users in defining the Region Of Interest (ROI) [50]. In the domain of SUV applications, many studies in lymphoma, lung, breast, and other types of cancers show that changes in this value can signify an early response to the treatment, which in many cases correlates with clinical outcomes as well. In recent years, several studies have focused on other applications of PET imaging modalities in cancer screening, including hypoxia detection, imaging cell proliferation, and imaging apoptosis [51].

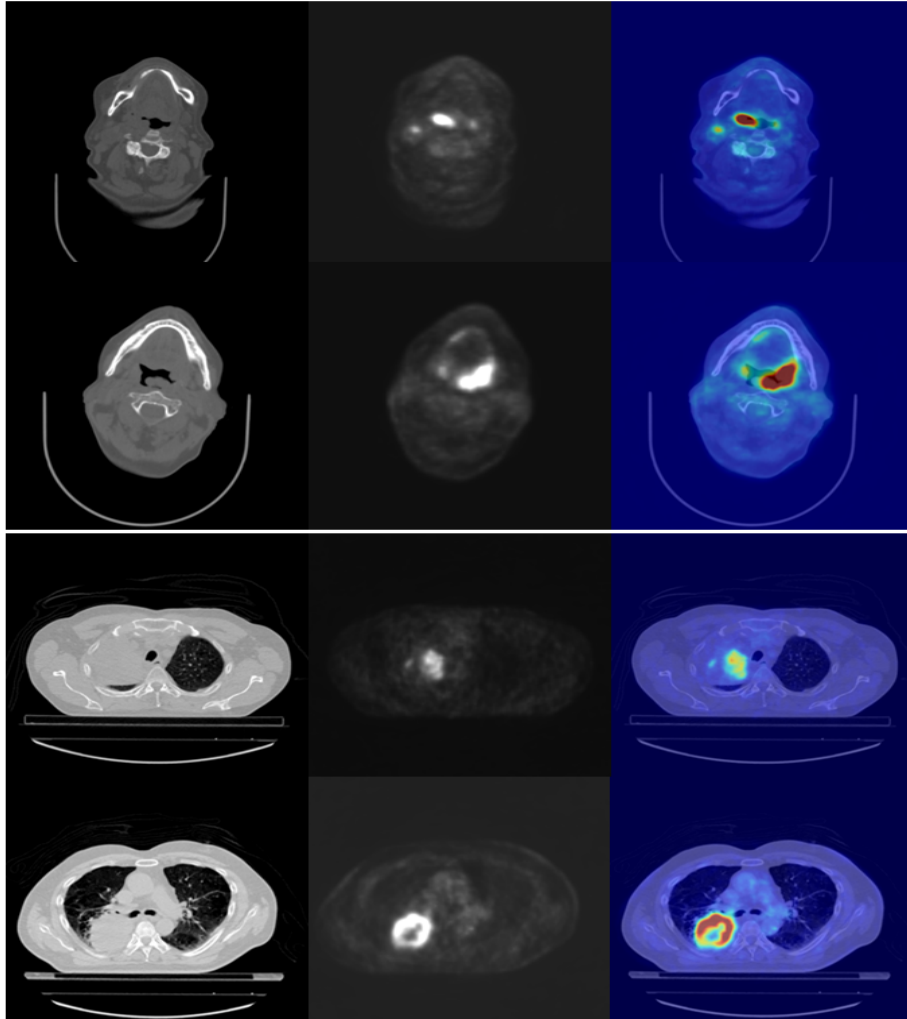


Figure 3.5. The important role of multimodal imaging in cancer screening. Examples of head-neck tumors (first two rows) and lung tumors (second two rows). The first column shows the anatomical CT images, the second column represents the molecular PET images, and the last column visualizes the fused images.

3.3.2 Cancer Treatment Methods

Cancer treatment is the process of using surgery, radiotherapy, chemical drugs, or other novel treatment options to cure, shrink, or stop the progression of cancer. Depending on the type of cancer and clinical conditions, one treatment or a combination of treatment methods

would be prescribed. The goal of all cancer treatment methods is to achieve a cure for the disease; however, helping patients live a normal lifespan free from cancer symptoms might or might not be possible. In other words, if the cure option is not possible, then treatments may be employed to shrink cancer or decelerate the cancer growth so that the patient would live with fewer cancer symptoms for as long as possible. Accordingly, the more specific goals of cancer treatment methods can be categorized as discussed below [52]–[54].

Primary treatment. The goal of primary treatment is to completely cure cancer and eliminate the cancer cells from the patient's body. Although all the therapy methods mentioned can be used as primary treatment, surgery is the most common. Moreover, if the diagnosed cancer is recognized as sensitive to radiotherapy or chemotherapy, one of these treatments or their combinations would be applied as well.

Adjuvant treatment. This kind of treatment refers to killing any cancerous cells that would remain after primary treatment to reduce the risk of cancer recurrence in the future. In some cases, *neoadjuvant* therapy, which is similar to adjuvant therapy, is prescribed to facilitate the process of primary treatment; however, it is used prior to the primary treatments. Similar to primary treatments, any kind of cancer treatment method can also be used as adjuvant treatment, with the most common ones including radiotherapy, chemotherapy, and hormone therapy.

Palliative treatment. This kind of treatment is used to relieve symptoms and improve the quality of life. It also can be used to reduce or control the side effects of cancer treatments and can be applied at any stage of an illness. Palliative treatment is helpful in prolonging the patient's life in cases of advanced cancers that cannot be cured. All kinds of treatment methods, as well as additional medications, can be used to relieve the symptoms.

Choosing the proper treatment method depends on several factors, including type and stage of cancer, as well as the general health conditions of the patient. These methods can be broadly categorized as described below.

Surgery. The goal of surgery in cancer therapy is to remove the tumor and nearby tissues during a surgical operation. Although it is the oldest type of treatment, it is still effective for many kinds of cancer.

Chemotherapy. This is an aggressive form of chemical drug-based intervention that aims at killing the rapidly growing cells inside the body. There are a variety of chemical drugs with different levels of side effects that are used to treat many types of cancer. These drugs

can remain inside the body for a week after treatment. Chemotherapy is often used in combination with other therapy techniques.

Radiotherapy. Radiation treatment is a type of cancer treatment that employs high doses of ionizing radiation to kill cancer cells, shrink the tumor size, and/or decelerate their growth. The radiation can be exposed from a machine outside the body (external beam radiation), or the source can be placed within the patient's body (brachytherapy).

Bone marrow transplant. Also known as stem cell transplant, this is a treatment for some particular types of cancer such as leukemia. This kind of treatment can be performed by using the patient's own bone marrow stem cells or those of a donor. It will allow oncologists to deliver higher doses of chemotherapy.

Immunotherapy. Also known as biologic therapy, immunotherapy is another type of treatment that uses the patient's own immune system to fight against cancer. This can be done either by stimulating the patient's immune system to work harder and smarter to attack cancer cells or by adding immune system components such as man-made immune system proteins.

Hormone therapy. This type of treatment blocks or lowers the number of specific hormones inside the body to stop or decelerate the growth of cancer cells. Some types of cancer, such as breast and prostate cancers, are fueled by hormones; therefore, eliminating these hormones would be helpful for treating the disease.

Targeted drug therapy. As with chemotherapy treatment, targeted therapy is considered a drug-based intervention method. However, in contrast to traditional chemotherapy drugs, targeted drugs are designed to interfere mostly with specific molecules involved in tumor growth while leaving most of the healthy cells unscathed.

Radiofrequency ablation. This is a minimally invasive procedure that employs heat as well as electrical energy to destroy the cancer cells. With the help of an image-guided procedure, a thin needle is directed through the skin into the cancer tissues. High-frequency energy then passes through the needle and heats the surrounding tissues to kill the nearby cells.

3.3.3 Treatment Response Evaluation Methods

Radiological imaging of cancerous sites plays a principal role in the management of cancer patients. It has been used to assess the changes in anatomical tumor burden as an integral part of the clinical evaluations of the therapeutic methods. There has been considerable interest in surrogate metrics for survival after the applied treatments, such as time-to-tumor progression, response rate, and progression-free survival. Changes in tumor size after treatment are often, but not necessarily, related to the overall survival time. To provide a unique and standard framework, WHO proposed a set of criteria in 1979 to assess the overall tumor response to treatment, mainly for use in trials where the tumor response was the primary endpoint. The idea was to compute the sum of products of bidimensional lesion measurements and determine the response to treatment by evaluating the changes in the same values during the ongoing treatment sessions. However, the original document proposed by WHO underwent several modifications to adapt it to new technologies and to address the cases not already clarified in the initial document. These alterations resulted in confusion about how to interpret the trial results; consequently, as a response to these problems, an international working party was formed to simplify and standardize the response criteria, which finally was published in 2000 as Response Evaluation Criteria in Solid Tumors (RECIST). The main contributions of the initial RECIST for an overall evaluation of the anatomical tumor burden contain instructions about the number of lesions for follow-ups (up to 10 and a maximum of five per organ), employing unidimensional instead of bidimensional measurements, and definitions of the minimum size of measurable lesions.

More specifically, a baseline imaging scan (CT or MRI) must be acquired within four weeks before the treatment starts. After accurately assessing the imaging data for the measurable diseases, target lesions should be carefully defined based on their size and their suitability for reproducible and repetitive measurements. During the follow-up scans, the longest diameter of each lesion should be measured on the axial image plane in a direction that best reflects their size. If a solid mass is split into multiple lesions, the sum of the longest diameters of each of the lesions should be calculated. However, if two nearby lesions fused and formed a single lesion, the longest diameter of the combined lesions will be considered. Finally, if a target lesion becomes too small to be measured precisely, the diameter should be set to 5 mm. The measured values during the follow-ups will be compared against the baseline measures to assess the objective response of the target tumor to the treatment. It should be emphasized that baseline measures are not necessarily those of the first imaging

scan but rather the measurements of the imaging scan where the smallest sum of diameters was observed. It is recommended to perform the follow-up scans at the end of each chemotherapy cycle—typically every six to eight weeks. However, to evaluate the overall response to the treatment, the measures calculated from the last sessions of therapy should be considered.

Based on these measurements, treatment response is categorized into three groups: Complete Response (CR), in which all the target lesions have disappeared; Partial Response (PR), in which the baseline sum of diameters diminishes more than 30%; and Progressive Disease (PD), in which the same metric expands at least 20%. Although RECIST has become the de facto standard for response evaluation, novel treatment methods as well as modern imaging scans entail continuous re-evaluation of the RECIST [55]–[60]. Figure 3.6 depicts slices of three separate subjects for measuring changes in the size of lung tumors as a result of applied radiation therapy.

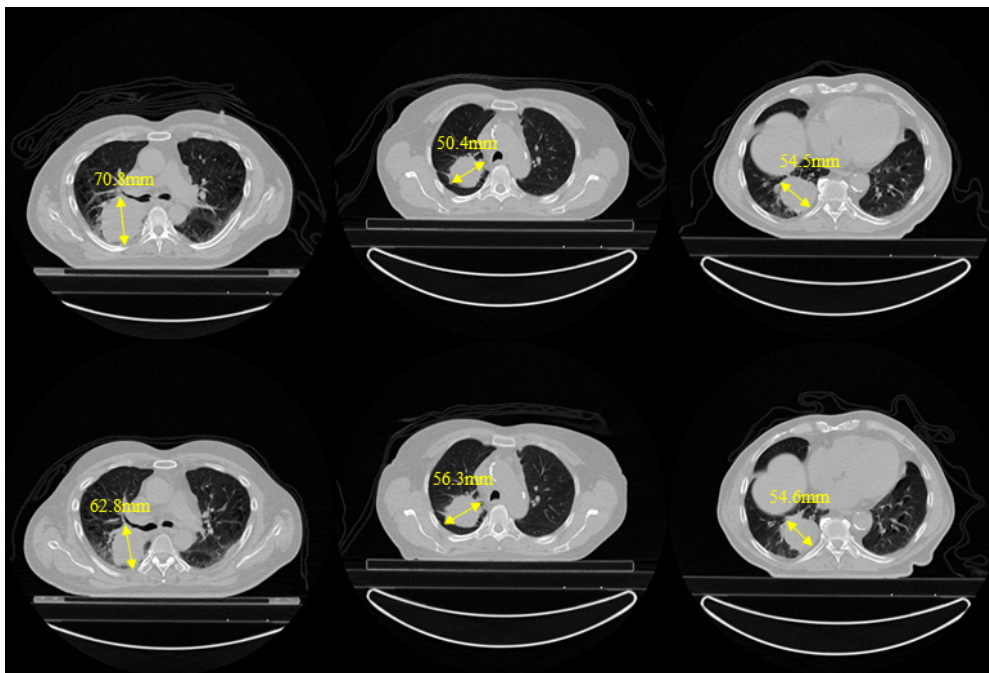


Figure 3.6. Measuring the changes in the size of lung tumors in computed tomography images. The axial slices in which the tumors appeared with the largest size were chosen, from which the diameter of the tumor in the most representative direction was measured. This manual procedure was examined for the pretreatment scan (first row) and the scans acquired from the second week of radiation therapy (second row).

With novel treatment methods, lack of tumor progression may be associated with improvements in outcome, even though a remarkable shrinkage in tumor size could not be observed. 18F-FDG PET has been employed for early monitoring to assess whether the tumor responds to the therapy or not. As this modality has been one of the most reliable

biomarkers introduced so far, comparing the SUV between a baseline scan and another scan after one or two cycles of therapy could be considered a reliable quantitative measurement to determine whether the treatment method was indeed effective or not. Therefore, tumor response to therapy at the level of metabolic activities has been introduced as a set of qualitative and quantitative approaches that are referred to as PET Response Criteria in Solid Tumors (PERCIST) [61].

3.4 Imaging Biomarkers

3.4.1 Precision Medicine in Medical Imaging

The term “precision medicine,” which has become widely used in recent years, refers to the view that factoring in variables such as human genomes, environment, and personal lifestyle, among others, when predicting, diagnosing, and treating human diseases will lead to remarkable improvements in human health [62]. Medical imaging plays an important role in precision medicine as it can provide supplementary information to stratify patients who are diagnosed with similar disease characteristics and share similar treatment responses [63].

The visual inspection and subjective interpretation of medical images by medical experts has been a traditional method over the past few decades for both diagnostic and therapeutic purposes. This procedure will be preserved as a practical routine by clinical experts, at least for the near future. In fact, the alternative approach of automatically interpreting the imaging data with ML-based or rule-based algorithms has not achieved an acceptable level of performance yet to replace the traditional approach—at least for many applications such as response evaluation. However, visual measurement is not only subjective but is also an error-prone and demanding method; it is also a time-consuming task, which leads to inter/intra-observer variability, which undoubtedly can degrade the expected clinical outcomes. One way to lessen the uncertainty surrounding interpretation is to translate the imaging data into quantitative characteristics from which quantitative results can be deduced.

Rapid advances in the field of image acquisition and hardware development over the past two decades have resulted in the development of modern medical imaging modalities that can capture high-resolution anatomical, physiological, functional, and metabolic quantitative information from the human body.

The CT scan reconstructs high-resolution images in which the voxel intensity represents the tissue density. Structural information—including tumor location or dimension, as well as quantitative physiological information of perfusion or necrosis—can be measured from CT images [64]–[66]. The PET scan yields high sensitivity images from the radioactive decays resulting from the injected radionuclides. It can quantify metabolic activities in the tumor tissues [67] or other organs. In contrast to CT and PET modalities, where the imaging signals are linearly proportional to the anatomical or metabolic features of the organs, and MRI signals are nonlinearly related to T1 and T2 magnetic relaxation phenomena as well as proton density distribution and/or other functionals such as diffusion, permeability, and cell density [68]. MR images can be employed in different ways to quantify the latent characteristics of tumors such as glioblastoma [69]–[71]. Such quantitative aspects of the modern imaging modalities can provide image-based characteristics that would be comparable or complementary to the laboratory biomarkers.

As a general definition, a biomarker refers to an objectively measurable indicator of some biological condition that is often measured to assess either normal biological processes or biological responses to a therapeutic intervention. Thus, it can be inferred that measuring anatomical, physiological, metabolic, or functional features of the organs and tissues through medical imaging is recognized as an imaging biomarker (see Figure 3.7). Such features can be used for a wide range of applications, such as the assessment of the functionality of normal tissues, quantifying the severity or degree of a change, and specifying the stage of the disease.

To make the radiology field more quantitatively rigorous, the Radiological Society of North America (RSNA) established the Quantitative Imaging Biomarker Alliance (QIBA) in 2007. QIBA's objective has been to unite researchers, health care professionals, and industrial collaborators to achieve accurate and reproducible quantitative evaluations from imaging modalities. A significant contribution of QIBA has been to prepare guidelines for the description of imaging biomarkers, which aims at characterizing and minimizing the bias-variance problem in the image-based measurements. These definitions are prepared for a broad range of targets, including but not limited to scanner and software manufacturers, pharmaceutical and diagnostic agent manufacturers, physicians, researchers, and so on. These detailed descriptions have been established as a set of norms and standards for both academic and industry investigators and are being implemented in clinical trials [68].

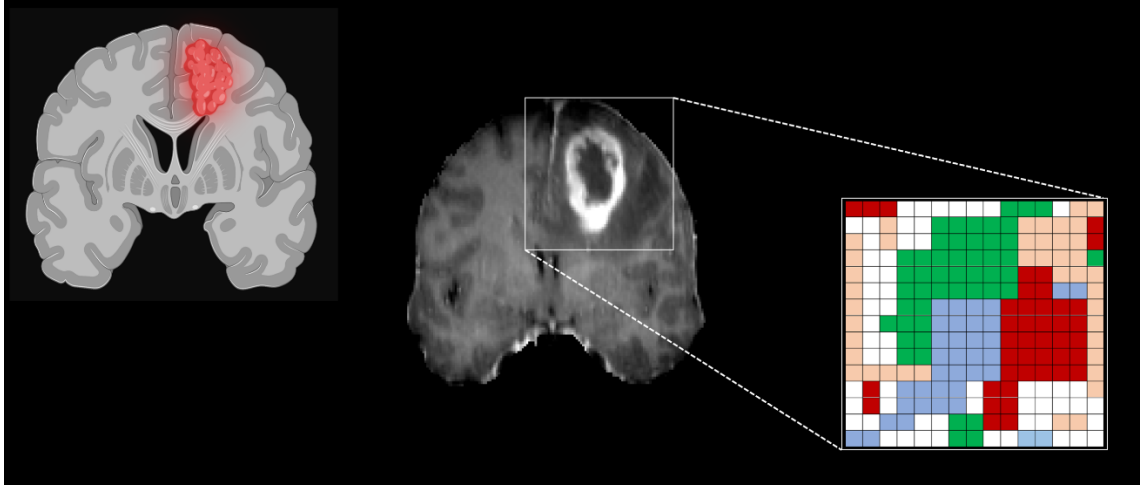


Figure 3.7. Imaging biomarkers in oncological scans aim to translate the tumor masses into relevant and meaningful numerical values that may represent the tumor phenotypes. (A part of the figure was created with Biorender.com.)

As previously described, precision medicine aims at employing novel diagnostics and therapeutic techniques specified to the individual needs of a patient, based on the patient's genetic, phenotypic, or psychosocial characteristics as well as biomarkers [72]. Since the goals of precision medicine are consistent with the initiatives of QIBA, imaging biomarkers provide an efficient source of information that is helpful for decision-making procedures to achieve an accurate diagnosis, prognosis, treatment planning, and response evaluation.

3.4.2 Development of Imaging Biomarkers

Medical images play a particular role in personalized medicine through the concept of QIBs. Although medical images have traditionally been investigated qualitatively, it should be noted that a clinical image is inherently acquired quantitatively, as it is a matrix/tensor of integer values. Each voxel in the image domain corresponds to a mapped region of anatomical organs. In other words, these voxel intensities can be considered virtual biopsies representing various anatomical and pathological characteristics. Therefore, a patient's specific physiological and anatomical attributes can be derived from such quantitative images. Although such deductions would lead to strong correlations with different underlying processes, they may not reveal the exact cause-effect association [68]. Biomarkers, in general, are categorized as (a) *predictive* if they give information on the effects of therapeutic intervention, (b) *prognostic* if they improve the accuracy of patient diagnosis/prognosis, (c) *response* if they show advantageous outcomes after treatment, and (d) *monitoring* if they detect relapse/recurrence or toxicity [68], [73]. In order to employ an imaging biomarker in clinical settings, multiple steps should be implemented and evaluated as it needs conceptual consistency, technical reproducibility, and acceptable accuracy. For

instance, there should be a well-defined relationship between the extracted quantified imaging features and the studied disease. Therefore, justifications are essential for the mechanism of defining and extracting the imaging biomarkers. Moreover, as the new measures should be compared against well-established ground truth, defining a proper reference method as a baseline standard is necessary. The choice of this reference method has a great impact on the final validation process. For instance, in many studies, performing a biopsy procedure followed by histopathology analysis is used as the gold standard. However, histopathology studies intrinsically entail measurement biases. More importantly, the biopsy is an invasive procedure, and therefore it cannot be repeated in every follow-up session. Moreover, sampling a tiny piece of an extremely inhomogeneous tumor will not represent the complex distribution of tumor cells; thus, biopsy has sampling bias. These are some of the main reasons for a potentially inaccurate estimation of tumor aggressiveness by the biopsy procedure that would further lead to ineffective treatment planning [68].

Other objectives such as the patient's overall survival, distant metastasis, time to recurrence, or time to progression of cancers are only a few examples of suitable endpoints in cancer research as prognostic outcomes [74]–[76].

3.4.3 Imaging Biomarkers and Quantitative Imaging

The first step to developing robust QIBs is to make sure the images are acquired under standard settings. Specifically, imaging modalities should be set to acquire and reconstruct images with high spatial and temporal resolutions, high signal-to-noise ratio, and high contrast-to-noise ratio as well. In addition, the acquisition settings should be replicable and, more importantly, they should be able to produce reproducible images with a low level of artifacts. It is also necessary to use the same acquisition parameters in the case of longitudinal and follow-up scans. Consequently, there have been efforts to set standard protocols for image acquisition and reconstruction [77], [78].

CT is the most common modality for quantitative imaging biomarker studies. CT scanners provides the potential to assess the tissue density, morphological and textural characteristics of the tumors and lymph nodes, and to capture changes in the size of the tumors from the follow-up scans [79]–[81]. In some tumors, certain CT characteristics such as necrosis and calcifications correspond to gene mutation, which is related to advanced-stage/grade of cancer and possibly unpromising overall survival [82].

PET has also been used to detect and stratify different types of cancers. A variety of radiotracers are employed in PET imaging for various purposes. The most common type is ^{18}F -FDG (radiolabeled glucose) as the vast majority of malignant cancers are characterized by high glucose consumption [83]. It has been verified that the SUV and PET-based textural features of tumoral regions are correlated to overall survival in lung cancer [67], [84] and to the development of metastatic diseases in cervical cancer [85].

MRI has been used to reconstruct high contrast functional and structural information to characterize soft tissues. QIBs have been extracted from structural MR images to discriminate prostate cancer from healthy tissues [86]. Dynamic contrast-enhanced MRI reflects the concentration of injected gadolinium agents over time and has been used to analyze brain tumor imaging biomarkers [87] as well as breast cancer [88], [89]. DW MRI represents the changes in the architecture of tissue cells based on the local differences in the movements of the water protons and has been hypothesized to depict cell death after the treatment [90]. The hyperintensity signal in T2-Weighted Fluid Attenuated Inversion Recovery (FLAIR) indicates the presence of infiltrative tumor cells in the edema part of brain tumors. It has been analyzed for the prediction of tumor recurrence and treatment outcome [91].

3.4.4 Imaging Biomarkers and Medical Image Analysis

Image analysis methods often start with basic *image preprocessing* steps. Noise reduction and image enhancement methods are crucial steps to improve the quality of the image and signal modeling processes. Moreover, artifacts introduce non-real data and cause biases in the analysis, so they should be eliminated if possible. Most of the time, there is a remarkable discrepancy between the acquisition settings of the same imaging modalities, which means the images are acquired with different spatial resolutions. Since the extraction of QIBs may be sensitive to spatial resolution, resampling the images into a fixed size is a necessary preprocessing step. In this context, super-resolution techniques are sometimes employed to increase the resolution synthetically [67, 68].

The images can be acquired from the same organ or subject but under different conditions such as time, angle, or even different modalities. Medical *image registration* is the task of spatially aligning two or more images to form a spatial correspondence of their common contents. Some of the applications of medical *image registration* are aligning the sequence of images acquired over time, detecting changes before and after the treatment, combining corresponding information from different modalities as well as image-guided

interventions [92]. For instance, extracting QIBs from PET-CT images for lung cancer [93], [94], as well as multimodal brain MR images for glioblastoma cancer, is done after registering the imaging modalities [95].

QIBs are often extracted from an ROI that should be accurately distinguished from the adjacent structures. This task is called *segmentation* and has been recognized as one of the most challenging steps in medical image analysis studies. In fact, partitioning the oncological images into “background” and “foreground” is complicated because target tumors and the surrounding healthy structures often share a high level of similarity in intensity and textural distributions. In many imaging biomarker studies, delineation of the target volume is done manually by at least one radiologist or radiation oncologist expert [96], [97]. However, a wide range of algorithms has been developed for automatic and semi-automatic segmentation, which can be categorized as either *data-driven* or *rule-based* methods. In *data-driven methods*, learning algorithms are trained in order to learn to extract relevant features with respect to the target labels from the image data. The advantage of such a method is that the algorithms can be trained in an end-to-end fashion; that is, within such models, independent segmentation, feature extraction, feature engineering, and classification steps are integrated into a single pipeline. However, the performance of these models depends heavily on the quality and quantity of the annotated data. Moreover, while the computational time for the prediction phase is usually negligible, their training procedure is computationally expensive. By contrast, *rule-based models* are often established on the basis of mathematical modeling, estimation theory, and probabilistic theories. Unlike the data-driven methods, establishing the rules with these methods does not require large-scale datasets, as knowledge about the concept is encoded in the explicit regulations. In other words, the direction of developing the rule-based models goes from general to specific. Of course, for these models, data are used to calibrate the model parameters and to test their generalization power. As stated, these methods do not require large-scale labeled datasets, which is an important advantage. Interpretability is considered another privilege of such models as well. However, it is not always possible to quantify the complicated physical concepts and encode them with mathematical rules. Accordingly, the application of such segmentation models is not as widespread as with data-driven models.

Semi-automatic segmentation methods involve some interactions between an operator and computational algorithms. Region growing is an algorithm that starts from an initial point or region named as the seed that needs to be defined either manually by a user or through another algorithm. Then, the neighboring voxels are added to the foreground in an

iterative approach, provided they satisfy a similarity measure [98]. In a similar approach, an ROI can be manually selected to surround the target volume on a single slice to initiate the evolution process of algorithms such as watershed or active contours by minimizing an energy functional [99]. The segmentation task in PET images is often adversely affected by the high level of intrinsic noise of the images. To address this issue, semi-automatic thresholding-based methods followed by variational algorithms are employed to successfully segment the lung tumors in FDG-PET images [100].

In contrast to semi-automatic methods, *automatic approaches* do not require any input from the users. These approaches often belong to data-driven models or combinations of rule-based and data-driven models. QIBs have been extracted after accurate breast tumor segmentation in MRI images with a fully automatic approach that has been performed on a large-scale dataset where the results were comparable to a gold standard delineated by experts [101]. Longitudinal brain tumors were segmented automatically in multimodal MRI sequences by employing a hidden Markov random field and a Bayesian approach on a challenging dataset, and yet this resulted in relatively high segmentation accuracy [102]. Finally, in the past few years, rapid developments in the field of DL methods have resulted in a variety of fully automated segmentation and detection models applied to different kinds of tumors (e.g., lung, brain, prostate, head-and-neck, and liver) with excellent performance [8], [9], [103]–[105].

After the image preparation and target segmentation steps, tumor volume will be translated into quantitative values by extracting QIBs. In general, the *extracted features* represent static anatomical attributes and/or dynamic biological ones. While static anatomical attributes estimate properties associated with the volume, topology, and morphology of the tissues, as well as the textural characteristics, dynamic biological features quantify the different physical, chemical, and biological processes. For instance, morphological features and textural inhomogeneities of the tumors are biomarkers extracted by static anatomical modeling of the acquired data. On the other hand, fat and iron measurements, or ADC quantification within the pancreas, belong to dynamic biological modeling.

3.4.5 Validation of Imaging Biomarkers

The reliability of QIBs can be realized if they are accurate, stable, and reproducible. To investigate such reliability factors, phantom-based experimental studies can be conducted. In other words, extracted quantified values from the biomarkers sometimes need to be

calibrated with phantoms to make sure that the extracted features are, to some extent, highly correlated with the real values. However, dealing with phantoms can be problematic because they cannot comprise all the properties of a human being's living tissues. In addition, there is always a potential lack of realism compared to in vivo measurements [68].

An important step in imaging biomarker development is to statistically determine the reproducibility and biological range of individual features. The most *reproducible features* are plausible candidates for distinguishing the subtle variations with respect to time changes, pathophysiology changes, or in response to the treatment. Moreover, this reproducibility must be assessed in the entire biological range, measured across the studied subjects. Therefore, a larger range tends to refer to the ability to capture a wide range of biological changes and represents the high level of usefulness of a feature. Additionally, *redundant* features increase the dimension of the feature vectors and may adversely affect the decision-making process; thus, the redundant features should be identified and eliminated from the feature space. Another source of variability is the different types of changes that can occur between the imaging scans. In fact, the inter-scan reproducibility of the features would be affected by the differences in the subject variables such as positioning, motion artifacts such as respiration phase, changes in the acquisition parameters such as image reconstruction and slice thickness, as well as image analysis algorithms such as initialization in semi-automatic segmentation methods. The most reproducible and informative features can be identified by analyzing a test-retest dataset that is acquired under certain conditions such as fixed acquisition parameters [106]–[108].

To be useful in clinical trials and real-world practices, QIBs research should be conducted and validated by following well-established standardized criteria. Specifically, qualification, validation, and standardization are critical parts of the development of imaging biomarkers. *Qualification* refers to the usage of a certain biomarker in specific content and relates to clinical approval, whereas *validation* relates to the performance of the biomarker in the test phase. One of the main challenges in the validation phase is that disease-related variations of the tissues that appeared in the images are often indirectly associated with multiple anatomical changes such as necrosis, vascular architectures, and fibrosis and are impacted by other factors such as inflammation, perfusion, and interstitial pressure as well [68], [109].

While *accuracy* refers to how close the measurements are to the real values, *precision* demonstrates how consistent the results are when measurements are repeated. The quality of diagnosis is related to the accuracy of the biomarker; however, the quality of disease

monitoring or treatment response effects is based on the precision of the biomarker. Therefore, the concepts of reproducibility and repeatability can be considered as two types of precision. Minimizing the effects of variations, especially in multi-center studies, should be performed by unifying the acquisition as well as processing protocols. Repeatability research is often performed in a single center with a specific imaging modality. After single-center experiments, the test phase should be completed at different centers under various conditions. On the other hand, reproducibility research is designed to assess the effects of variables such as different centers and different acquisition protocols on the biomarker measurement's precision. To improve the reproducibility by knowing the impact of variations, test-retest experiments should be performed. More importantly, in multicenter studies, the effects of different implementation settings on both the measurements, as well as the correlation with the reference standards, should be carefully investigated [68].

Biological validations are related to the correlation among the preclinical models or human studies and other reference points such as histopathology, invasive measurements, genomics, or proteomics. Such validations are conducted as experimental studies to select the proper references as gold standards relevant to the biological changes intended for assessment. In these experimental studies, the effects of epidemiological data and the biological variations on the measurements will be examined.

Clinical validations start with single-center observational studies on patients who participate in the imaging biomarker experiments to find out how well it works in a limited controlled condition. If the achieved results prove to be accurate, multicenter observational cohort studies and even clinical trials—both retrospective and prospective—may be started [68].

3.5 Radiomics

Since 2010, the field of *radiomics* has emerged as a new way to develop imaging biomarkers. This term is derived from the word “radio” as adopted from the term “radiology” and the suffix “omics,” which was first used in the term “genomics,” the study of human genomes. In general, “omics” refers to the study of biological phenomena such as transcriptomics (RNA), proteomics (proteins), and metabolomics (metabolites) [110]. As radiomics deals with quantifying medical images and using statistical/ML methods, it could be mistakenly considered to be a new application of Computer-Aided Diagnosis/Detection (CAD) methods. However, radiomics is not a branch of CAD systems for two reasons. First, the number of image-based features extracted in radiomics is in the order of hundreds or even thousands,

which is much more than the limited number of features in CAD studies. Second, radiomics aims at developing predictive or prognostic models, which means the extracted features are associated with biological or clinical endpoints [77], while CAD-based models are mainly focused on diagnostic purposes.

It has been verified that solid tumors consist of several sub-clones of cancer cells with different characteristics that represent both spatial and temporal inhomogeneities within the tumor mass. *Quantifying these heterogeneities* through medical images would represent an expression of genomic heterogeneity, which is a sign of an unpromising prognosis. In fact, tumors with a high level of genomic heterogeneity are more resistant to the treatment and are more likely to develop metastasis [76]. The goal of the rapidly growing radiomics field is to capture valuable information regarding the shape, size, intensity, and texture of tumor phenotype that would be consistent with or complement other sources of information such as clinical reports, laboratory, genomics, or proteomics analyses [111]. Radiomic feature extraction can be applied to different imaging modalities such as CT, MRI, PET, and ultrasound. The feature vector, which is a set of all extracted imaging features from one single patient, is called “radiome” and a combination of different features that hold the most prognostic or predictive values are recognized as “radiomic signature” or sometimes QIBs [87].

In general, radiomic features can be divided into three main classes: shape (*morphological*), first-order histograms (*global statistics*), and second-order histograms (*textural*). All textural features are both rotation and translation invariant. Moreover, illumination invariance of textural features can be achieved by applying certain preprocessing steps such as histogram matching. However, none of the textures is scale-invariant. The mentioned features are extracted from the original images (after preprocessing) and the transformed images after applying filters such as wavelet, Laplacian of Gaussian (LoG), or Gabor. In addition to the general preprocessing steps, some of the feature families require additional processing steps before feature extraction. For example, first-order histograms and textural features need prior intensity discretization into gray level bins.

Looked at from another perspective, radiomic features can be divided into time-invariant (static) and time-variant (dynamic) features. This stratification is defined based on the nature of the image acquisition settings at the time of scanning. Examples of *static features* include statistical features extracted from the images after conversion of raw intensity values into SUV in PET images or Hounsfield Unit (HU) in CT images, Intensity

Volume Histogram (IVH) that is analogous to Dose Volume Histogram (DVH) and is used widely in radiotherapy treatment planning, as well as geometrical and textural features. On the other hand, *dynamic features* are extracted from time-varying modalities such as certain MRI sequences or dynamic PET. Dynamic radiomic features are defined on the basis of kinetic analysis using tissue compartment models [112]. It should be emphasized that some radiomic features belong only to certain applications. For instance, SUV-based features are limited only to PET studies, while fusion or fractal features are applicable to multimodal studies.

Morphological features describe geometric aspects of the segmented region and are based on the ROI voxel coordinates representing the volume. In this class, features such as surface area, volume, 2D, and 3D maximal diameter characterize the size of the segmented target. The eccentricity, compactness, sphericity, and surface-to-volume ratio of the tumor determine how much the segmented tumor looks like a sphere. The solidity of the tumor is defined as the ratio of the number of voxels inside the segmented tumor to the number of voxels in the smallest polyhedron containing the tumor region. Major, minor, and least axis lengths, respectively, indicate the main orientation, extension of the tumor toward the second direction, and the least extended direction of the tumor. From these lengths, then, tumor flatness is calculated as the ratio of major and least axis lengths [96], [111], [113].

In contrast to morphological features, *global statistical radiomics features* determine the characteristics of the intensity distribution of the voxels inside the ROI and are calculated from the histogram, such as minimum, maximum, mean, and standard deviation. While skewness and kurtosis demonstrate the degree of distribution asymmetry and sharpness, uniformity and entropy represent distribution inhomogeneities. To capture the local intensity variations, the mentioned features can be extracted from local neighboring voxels [114], [115]. SUV-based features are another type of global statistics that are calculated from PET images. Parameters such as SUV_{max} , SUV_{peak} , SUV_{mean} , respectively, quantify the maximum uptake within the ROI, average uptake in a local neighborhood of maximum uptake, and average uptake inside the ROI. Average intensity inside the ROI can also be calculated at different time points when dealing with time-varying signals such as contrast injection in MRI, in which the focus will be on enhancement kinetics. Features such as maximum uptake (peak value of the measurements), peak location (time index in which the maximum uptake occurs), and uptake rate (ratio of maximum uptake to peak location) can be extracted from the measured values [114], [116].

In image analysis, *textures or second-order statistics* were originally designed to assess the surface texture in 2D images; however, they have been extended to volumetric data. In general, textural radiomic features describe the spatial distribution of voxel intensities. The term “image texture” indicates the perceived spatial variation of the pixel/voxel intensity that can be visualized as a gray-level scale. In the Gray-Level Co-occurrence Matrix (GLCM), the number of rows and columns indicates the intensity values, and the value of each matrix element represents the number of times that corresponding intensity values appeared in a certain spatial order, such as distance or angle. For instance, in Figure 3.8, the voxel values of a subpart of the tumoral region are presented in a 4×4 matrix. This matrix contains only four values; therefore, the GLCM will form a 4×4 matrix, and its values are determined based on the number of times the voxel values are repeated within a certain direction. GLCM-based features include second-order entropy and energy, both representing intensity inhomogeneities; contrast (quantifies local intensity variations); cluster shade (sensitive to heterogeneity); as well as homogeneity, correlation, and dissimilarity [77], [114].

Gray Level Run Length Matrix (GLRLM), like GLCM, assesses the distribution of discretized intensity values in an image. While GLCM examines the co-occurrence of intensity values with neighboring voxels, GLRLM assesses run lengths. A run length refers to the length of a consecutive sequence of voxels with the same intensity values along a certain direction. Therefore, GLRLM incorporates the occurrence of runs with certain lengths for specific discretized intensity values. Radiomic features such as short/long runs emphasis, low/high gray level run emphasis, gray level non-uniformity, run-length non-uniformity, run percentage, and gray level variance can be calculated from this matrix [114], [117].

Gray Level Size Zone Matrix (GLSZM) counts the number of zones of linked voxels. Linked voxels refer to the neighboring voxels with identical intensity values. In 2D images, 8-connectedness is considered for the neighboring pixels; however, in a volumetric image, voxels can be neighbors in 26 different directions. A wide and flat GLSZM indicates a heterogeneous ROI, while a narrow one represents a homogeneous ROI. Some of the features that can be extracted from this matrix include small/large zone emphasis, low/high gray-level emphasis, and small/large low gray-level emphasis [113].

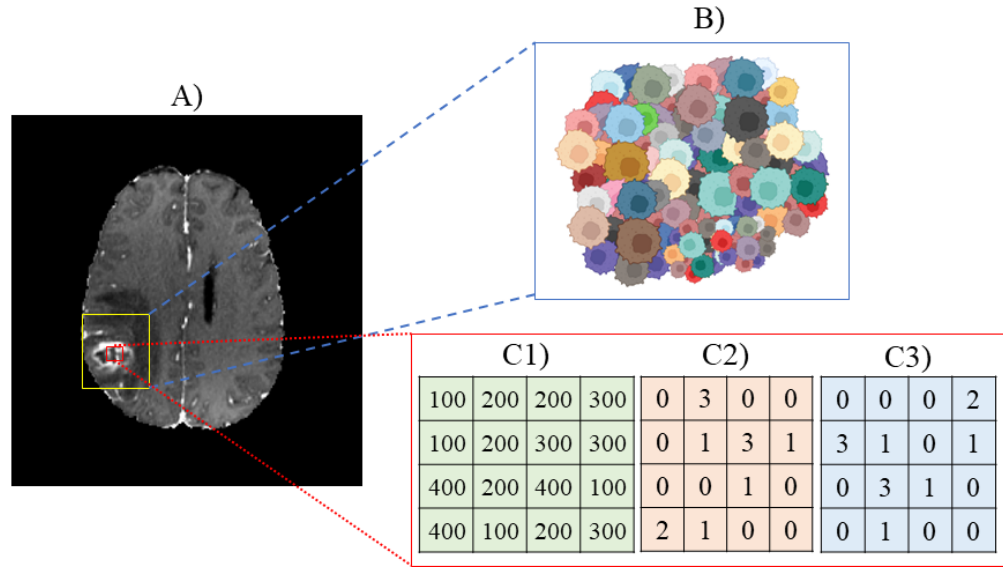


Figure 3.8. Quantification of the tumor heterogeneities with textural radiomics. A) T1w contrast enhancement sequence of high-grade glioma; B) graphical illustration of cancer cell heterogeneities; C1) raw voxel intensities of a small tumor subregion; C2) gray level co-occurrence matrix at 0 degrees; and C3) gray level co-occurrence matrix at 180 degrees. (A part of the figure was created with Biorender.com.)

Neighboring Gray Tone Difference Matrix (NGTDM) contains the sum of intensity differences of voxels with discretized intensity values and the average discretized intensity of neighboring voxels within a Chebyshev distance [113]. Radiomic features from this matrix include coarseness, busyness, complexity, texture strength, and contrast. For instance, in images with a neighborhood of relatively uniform voxel intensities, the value of the coarseness feature will be high. In contrast, images with a high spatial frequency lead to a high value of busyness [118].

The Image Biomarker Standard Initiative (IBSI) is an independent international collaboration that aims at standardizing the definition and extraction of QIBs as the lack of reproducibility and external validations have been major challenges in this field. IBSI, therefore, seeks to provide guidelines for radiomics analysis, including feature nomenclature and definition, feature calculation, benchmark dataset, and reporting regulations [119].

3.5.1 Radiomic Feature Analysis

The goal of radiomic feature analysis is to develop a mathematical model or function to stratify patients based on the predicted outcome using extracted radiomic features. From the ML perspective, this task is equivalent to training a learning algorithm to predict the class labels of given data points (see Figure 3.9). In other words, the learning algorithm analyzes the training data and learns the underlying representative characteristics of the

data to infer a hypothesis in order to predict the class labels of unseen data. In radiomics analysis, class labels could be patient outcomes or tumor phenotypes [120], [121].

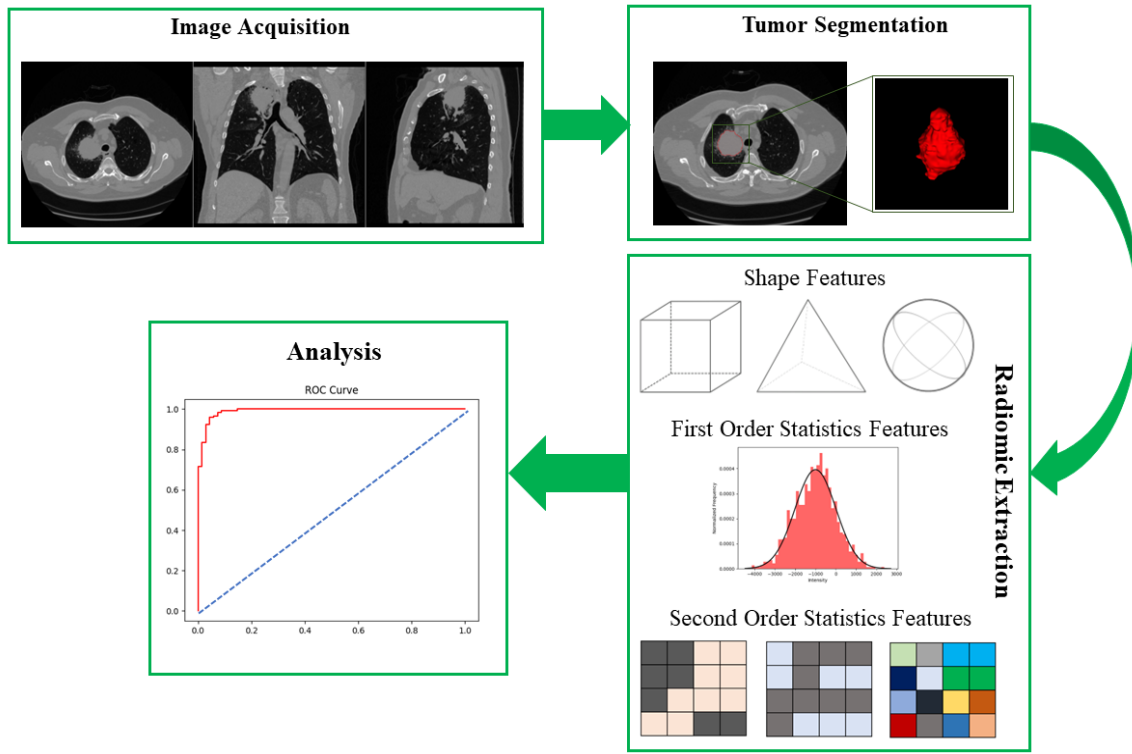


Figure 3.9. Radiomic analysis pipeline consists of three main steps, including tumor mass segmentation, radiomic descriptor extraction, and feature analysis.

3.5.1.1 Feature Selection Methods

The number of extracted radiomic features from the ROIs can be large, especially if the transformed images (wavelet and/or LoG) are also analyzed. In such cases, the number of extracted features lies in the range of a few hundred to even a few thousand; nevertheless, they usually exceed the number of subjects [122], [123]. However, not all these features are necessarily informative with respect to the class labels. In other words, some of them might be redundant or correlated to each other. Therefore, employing them all together to train a learning algorithm would not be a wise choice because irrelevant or partially relevant radiomic features can adversely impact the predictive performance. In other words, eliminating the irrelevant and redundant features leads to reducing the computational time, which may increase the generalization competence and accuracy of the model. Therefore, it is necessary to identify the most informative features and eliminate the least representative ones. Feature selection is the process of selecting a subset of features that contribute most to the prediction variable.

The simplest way to perform a feature selection algorithm is to set a scoring criterion for the features and measure the stability or correlation of the features with respect to the criteria, rank the results, and eliminate those features with the minor ranks. This plain feature selection approach is used in univariate methods. The main drawback of such approaches is the lack of dependency calculation between the features [124].

One way to *identify redundant features* is to calculate the correlation between the extracted radiomic features. Therefore, removing the highly correlated features assures us that the remaining features have the least degree of linear dependency on each other. Such an approach can be implemented by the computation of the correlation matrix of all features. The same approach can be employed to identify nonlinearly associated features by calculating the Mutual Information (MI) metric [122]. Moreover, to compute the nonlinear dependency between the features, Minimum Redundancy Maximum Relevance (MRMR) calculates the MI metric between a set of features and the class labels. Therefore, this approach preserves the features with minimum dependency on each other (*minimum redundancy*) while posing a high degree of relation to the class labels (*maximum relevance*). MRMR is one of the most common feature selection algorithms in radiomics analysis [125]–[128].

By maximizing a penalized log-likelihood function, the *Least Absolute Shrinkage and Selection Operator (LASSO) regression* model is able to apply multiplicative coefficients to the features and shrink some of them to zero to eliminate their contribution to the objective function and thus their effect in the model. This type of radiomic feature selection has been mostly used for regression tasks with a high dimensional feature set [129], [130].

RElevance in Estimating Features (RELIEF) is another feature selection algorithm that leads to promising results in some radiomics studies. The core idea of RELIEF is to estimate the quality of the features based on how well the features can distinguish between the subjects that are near to each other. The main advantage of the RELIEF algorithm is that it does not assume any kind of dependencies among the features; in other words, it is a context-aware method for feature selection [131], [132]. The downside of this algorithm, however, is its incapability to identify redundant features that tend to have a similar score for evaluation.

Class imbalance is one of the most common issues in ML studies. This issue occurs when each of the classes does not make up an equal portion of the entire dataset. Therefore, it is essential to properly adjust the employed metrics and methods toward the desired goals.

Feature Assessment by Sliding Threshold (FAST) is a feature ranking method designed to address the issues with small datasets and imbalanced class labels [131].

One of the most efficient feature selection techniques with respect to model performance is *Forward Feature Selection (FFS)*. This technique evaluates the performance of the learning algorithms in different combinations of feature subsets. The model starts to train the learning algorithm with every single feature; the one that leads to the highest predictive power will be held; then, the combinations of the retained feature and all the other single features will be examined one by one to find out the best subset of two features with the highest predictive power. This process continues iteratively until no improvement in model performance is observed, or a predefined number of features is selected. In contrast to FFS, *Backward Feature Selection (BFS)* starts with the total number of features and iteratively eliminates the feature subsets [133].

3.5.1.2 Learning Algorithms

Generally speaking, the majority of the learning algorithms employed for radiomics belong to supervised learning methods. *Logistic regression* has been one of the most widely used statistical methods in radiomic studies, perhaps because of its simplicity [134]–[137]. *Support Vector Machine (SVM)* is another powerful tool in radiomics analysis. It is a discriminative supervised technique that solves an optimization problem to infer a decision boundary (hyperplane) in order to separate radiomic features from two categories, such as responders or non-responders [88], [131], [138], [139]. Many efforts have been made to develop learning algorithms in the field of medicine that would be capable of resembling human reasoning. Among them, Decision Trees (DTs) can represent hypotheses as sequential “if-then” statements that can, to some extent, resemble medical expert reasoning. The Random Forest (RF), which is an ensemble of DTs, incorporates two levels of randomization—first in subject sampling and second in feature subset sampling. Thus, it can also be considered as a learning algorithm with an embedded feature selection method that might be robust against overfitting problems and often leads to high prediction values in radiomics analysis [140]–[142]. Although classic Artificial Neural Networks (ANNs) have been used in some radiomics analyses, the development of advanced deep networks over the past few years has resulted in a variety of powerful learning algorithms that will be discussed, in detail, in the next section.

Regardless of the type of learning algorithms or feature selections, it should be emphasized that in almost all the radiomics studies, the number of extracted features is

much greater than the number of subjects. Therefore, the risk of model memorization, also known as overfitting, instead of generalization is quite high, which requires some strategies to examine the potential of this risk and ideally reduce it. One conventional approach, not limited only to radiomics but to all the ML tasks, is to use the *K-fold Cross-Validation (CV)* resampling procedure. This procedure has a single parameter, K , that indicates the number of groups that the subjects be split into. As a result, the learning model will be trained with $K-1$ subgroups of the subjects and will be validated with the remaining subgroup of subjects. The most extreme extension of CV is called the *Leave One Out (LOO)* algorithm in which, for each subject on the dataset, a learning model will be trained using all the other subjects and validated on the selected subject (the parameter K is equal to the number of subjects) [143].

Finally, it is worth mentioning that radiomics analysis has been successfully used in various aspects of the oncology field and has yielded promising results. Prediction of early response to chemoradiotherapy [144] or chemotherapy [145], [146], the occurrence of distant metastasis [147], [148] or lymph node involvements [149]–[151], prediction of overall survival [14], [152], [153], or tumor characteristics as histology [154] as well as genetic footprints [155] are some examples in which reliable and stable radiomic signatures have been introduced as potential prognostic and predictive QIBs in different types of cancers such as lung [67], [156], [157], glioblastoma [158], prostate [159], [160], breast [15], peritumoral edema [161], and head and neck [162], [163].

3.6 Deep Learning Methods for Medical Image Analysis

With continuous development in medical imaging modality systems, a variety of medical imaging techniques have emerged that allow for qualitative and quantitative assessments of anatomical details. They acquire high-resolution images from vital parts of the body to examine the structural, functional, or metabolic characteristics of healthy organs or specify the attributes of diseases. However, evaluating the reconstructed images requires the expertise of an experienced radiologist to visually assess the images. This standard approach, however, faces several limitations. First, compared to the great demand for manually analyzing medical images, the number of expert radiologists is limited. Second, careful examination of the volumetric scans on a slice-by-slice basis is a burdensome and time-consuming task. Third, such a subjective assessment leads to inter-/intraobserver variability issues. Accordingly, developing accurate and robust computerized tools to assist radiologists in medical image analysis tasks has been considered as a vital contribution. In

this context, although CAD-based systems have a long history, it was not until the advent of DL methods that automatic medical image analysis methods dramatically gained importance.

3.6.1 Deep Neural Networks

The human brain, which consists of about 100 billion structurally and functionally simple units—that is, neurons—is considered the most well-organized system [164]. This complex system, called a biological neural network, can simultaneously and efficiently process information from different sources such as hearing, vision, touch, etc. One of the underlying reasons for the high processing performance of this system is the collaboration between the neurons. In fact, each neuron can be connected to up to 10,000 other neurons. As a result, the probable processing power of a human brain is roughly equal to 100-1000 petaflops of information processing [165]. ANNs are a group of computational models that are inspired by the structure and function of biological neural networks [166], [167]. However, it should be emphasized that they are far from the real biological neural networks. The *perceptron* is the simplest learnable artificial neural model and incorporates a set of trainable weights $\{w_i\}_{i=1}^D$ and a bias w_0 as well as the inputs $\{v_i\}_{i=1}^D$ (see Figure 3.10.A). Given D-dimensional input data, the output will be calculated from an activation function by taking the weighted sum of the inputs:

$$y(v, \theta) = f\left(\sum_{i=1}^D v_i w_i + w_0\right) = f(w^T v + w_0)$$

where f is a nonlinear function known as activation functions, $\theta = \{w, w_0\}$ presents a parameter set, $w = [w_i]_{i=1}^D \in R^D$ is a connection weight vector and, and w_0 indicates the bias. There are a variety of nonlinear activation functions (f) and, for instance, logistic sigmoid $f(z) = 1/(1 + \exp(-z))$ is a common choice for a binary classification task. Despite the use of the nonlinear function, one of the major limitations of a single-layer neural network is that the model can only perform linear classification tasks. Such limitations can be addressed by adding different layers on top of each other, which is also known as a *Multi-Layer Perceptron (MLP)* (see Figure 3.10.B):

$$y_k = f^{(L)} \left(\sum_l W_{kl}^{(L)} f^{(L-1)} \left(\sum_m W_{lm} f^{(L-2)} \left(\dots f^{(1)} \left(\sum_i W_{ji}^{(1)} x_i \right) \right) \right) \right)$$

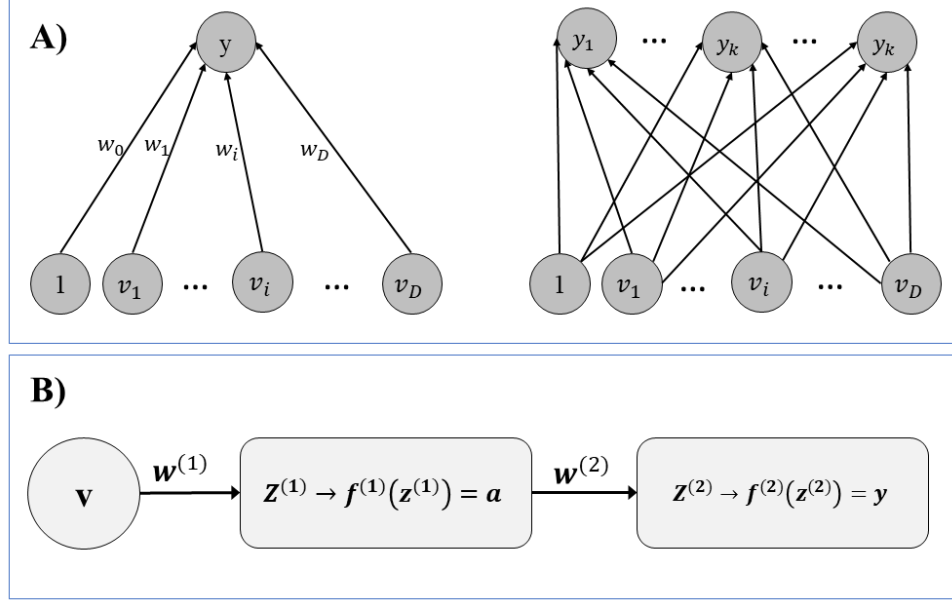


Figure 3.10. The schematic architecture of basic neural networks. A) Single-layer network with single-output (left) and multiple-output (right); B) A two-layer neural network.

It should be noticed that while it is possible, in theory, to use a different type of nonlinear activation function for different layers, the choice of the activation function of the last layer heavily depends on the type of problem.

Network hyperparameters learning refers to the concept of two fundamental problems: network architecture learning and network parameters learning. While the problem of network architecture learning remains an open challenge, there is a well-established efficient algorithm for network parameters optimization called backpropagation. Simply speaking, this optimization procedure calculates the gradient of the error function with respect to the model parameters, including weights and biases, by applying the derivation chain rule to update such parameters into their optimum values during the training phase.

Theoretically, a two-layer neural network with a limited number of neurons is a universal approximator, which means it can approximate any arbitrary continuous function [168]. However, by increasing the network capacity, i.e., the number of layers as well as the neurons within each layer, a deep network can approximate complex functions, theoretically, in a more accurate way. In contrast to shallow networks that require well-

established extracted features and even feature engineering steps, deep models can discover the informative features automatically in a hierarchical fashion. The central problem in DL is to meaningfully map the input data to targets (label or ground truth). To do so, the model should learn latent representations of the input data and transfer the learned characteristics to the desired outputs. Simply speaking, to represent the data means to encode the data. As these representations are captured in successive layers, the more appropriate term would be “learning of hierarchical representations (across layers).” Figure 3.11 illustrates how a network transforms an input image into representations that are increasingly become more abstract and differ from the original image but informative with respect to the expected output. One advantage of deep models is that they can generate high-level feature representations directly from raw images. With the availability of a large-scale dataset as well as powerful Graphics Processing Units (GPUs), DL techniques have shown outstanding success over the past few years in different fields of medical image analysis such as classification, target detection/segmentation, and image generation [167], [169]. In the following subsections, some of the major components of deep neural networks as well as some important applications of DL models in the field of medical image analysis are outlined.

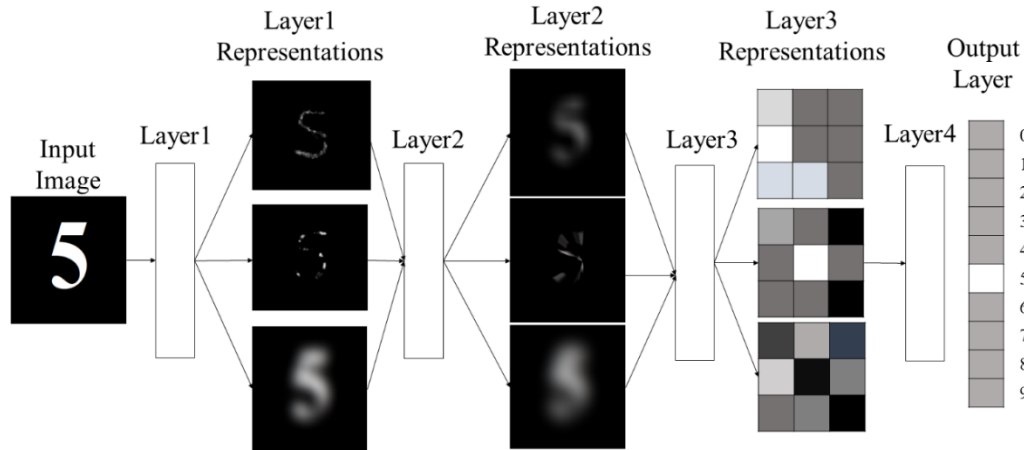


Figure 3.11. Learning hierarchical representations from an input digit image with a shallow classification network.

3.6.2 Convolutional Neural Networks

Employing the traditional ANNs- in which all the neurons in consecutive layers are connected to each other, also known as fully-connected networks- for real-world problems with images as input is challenging for the following reasons. Imagine a 2D image with the size of 100×100 pixels as an input to a network containing 1024 neurons in the first layer.

Therefore, the matrix of the weights just for the first layer consists of $100 \times 100 \times 1024 = 1,0240,000$ elements. By increasing the number of layers and neurons, the number of network components will increase dramatically, which leads to more computational complexity. Aside from computational burdens and hardware limitations, reordering the image matrix/tensor into a vector will lead to the loss of spatial information. In the following paragraphs, the basic components of *Convolutional Neural Networks (CNNs)* will be described briefly [170]–[172].

Convolution. One way to address the problems of matrix vectorization, and computational complexities, is to replace the matrix multiplication with the convolution operator. In other words, learning a set of convolution kernels with the size of $m \times m$ (a typical choice is $m = 3$) is much more tractable than learning the elements of a large matrix (e.g., 10000×1024). Convolution is a well-established mathematical operator that can be implemented with two constraints in the context of regular neural networks. The first constraint is *locality*. The size of the convolution kernels is much smaller than the size of the image (or feature maps) it is applied to. This is in contrast to the global approach typically incorporated in fully-connected layers. The second constraint is *weight sharing*. Performing a convolution operator means that the same kernel will be applied across the image (or feature maps). In other words, the same local kernel is used on different locations of the image. Although in the classical convolution operator, the kernel shifts one position to the right after each multiplication, shifting in the right direction in CNNs would be applied with a larger step. In CNN nomenclature, this parameter is recognized as stride, which means the image will be downsampled after applying the convolutional operator with a stride parameter larger than one.

Max-pooling. This operator functions as a maximum filtering unit in which each $m \times m$ region is replaced by its maximum value. The Max-pooling operator serves two purposes. First, holding the maximum value is equivalent to holding the highest activation in a local region. Therefore, it makes the feature maps more robust against the changes in the position of objects in the image. This operation, to some extent, resembles the way complex biological cells work. The second purpose is that retaining one value instead of a matrix with the size of $m \times m$ reduces the size of the activation for the next successive layer by a factor of m^2 . As a result, the number of learnable hyperparameters in the successive layers can be reduced. It is worth mentioning that other types of pooling layers have been introduced, such as average pooling, winner-takes-all pooling, and stochastic pooling as well.

Nonlinear activation functions. Convolution is a linear operation, and the stack of convolutional layers on top of each other will form a cascade of linear operators. Accordingly, applying *nonlinearity* is quite necessary to increase the expressive power of the model. In CNNs, a nonlinear function usually follows each convolutional layer. The nonlinear function which is often employed in traditional neural networks has been restricted to a sigmoid or hyperbolic tangent. However, the CNN models are trained with other types of nonlinear functions such as *Rectified Linear Unit (ReLU)*, $ReLU(x) = \max(0, x)$. It has been verified that using ReLU leads to faster training. However, the downside of being zero for all the negative values can be problematic, especially if the learning rate is too high or there is a large negative bias. *Leaky ReLU* applies a small slope for negative values instead of setting them to zero. Specifically, it is defined as $Leaky_{ReLU}(x) = \max(0, x) + \alpha \min(0, x)$ where α is a predetermined parameter. This strategy can fix the issue with zero values and avoid zero gradients (Figure 3.12) [173]. More variations of activation functions include Exponential Linear Unit (ELU) [174], Concatenated ReLU (CReLU) [175], and ReLU-6 [176].

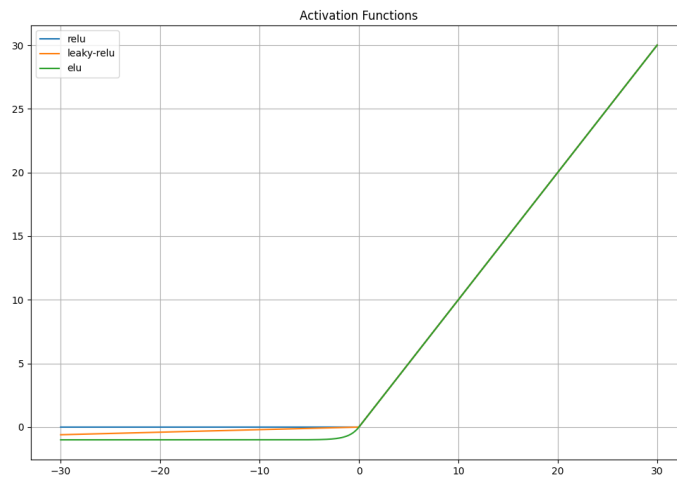


Figure 3.12. Comparison between ReLU, Leaky ReLU, and ELU activation functions. Please note that the positive values of the curves are overlapped for all three functions.

Dense layers. The classical matrix multiplications of MLPs are called dense (fully connected) layers since each neuron in the previous layer is connected to every neuron in the next layer. Such layers greatly increase the number of learnable parameters, and therefore in modern networks, using many dense layers is often avoided.

Network depth. According to the Universal Approximation theorem, a neural network with only one single hidden layer but a large number of neurons is enough to approximate any arbitrary continuous function. However, it has been verified that networks with more layers but a much smaller number of neurons can make the same approximation [177]. The

motivation for developing deeper networks has been clear; however, properly training networks with more than three layers was not feasible before the introduction of the greedy layer-wise pre-training model [178]. Relatively recent architectures such as VGGNet [179], GoogleNet [180], and ResNet [181] incorporate, respectively, 19, 22, and even 152 layers without encountering any training problems such as vanishing gradient.

Optimization algorithms. Regardless of the network depth, neural networks are trained using a conventional backpropagation algorithm that employs the chain rule to speed up the gradient calculations in the Gradient Descent (GD) algorithm. Unfortunately, it is not feasible to employ classical GD over a large dataset in one step; instead, an algorithm called Stochastic Gradient Descent (SGD) is often used to apply several steps of GD (i.e., updates) computed over small, random samples of the dataset, so-called mini-batches, in a consecutive fashion. It is worth mentioning that the number of samples per mini-batches depends on the computational resources. Although the SGD algorithm leads to better generalization power compared to GD, it is a relatively slow optimizer. Therefore, it often uses a small portion of the data instead of single data (mini-batch). Other types of optimizers include Adagrad, Adadelata, and Adam [182].

Dropout. Similar to other kinds of optimization procedures, adding a *regularization* term is important for reducing the risk of overfitting. In the context of ANNs, *regularization* refers to a set of techniques aimed at lowering the complexity of a model during the training to prevent overfitting. In this context, applying conventional regularization techniques such as L1 and L2 norms on the weights and biases of the neurons may be considered. The idea behind a *dropout algorithm* is to train only a random subset of neurons at each iteration of SGD. Thus, for each mini-batches during the training, a predefined portion of neurons with a probability of p will be dropped randomly and do not contribute to training. However, all the neurons will be activated in the test phase and multiplied with p to account for the scaling. Other kinds of regularization techniques in the context of the dropout algorithm include DropConnect [183] and Fast DropOut [184].

Batch normalization: Another efficient regularization technique is Batch Normalization (BN) which is helpful in speeding up the convergence. During the training procedure, as a result of the previous layer's changing weights, the distribution of each layer's input changes continuously. These changes decelerate the convergence and require careful initialization. To address this issue, BN normalizes the output of a layer by using the running average of the mean-variance statistics of each mini-batch. This module consists of two parameters—

shift and scale— which, along other weights and biases, are optimized during the training procedure.

By adding the described basic elements of a CNN model in a different order, a large number of architectures have been introduced and applied to different tasks with outstanding results. Applications of DL in medical image analysis have rapidly grown so that more advanced functionalities are derived from more sophisticated network architectures [19], [185]–[187].

3.6.3 Deep Classification

Medical image classification is an intricate process that can be affected by many different factors. Because the classification results are the basis for many further interventions such as treatment and surgery planning, great efforts have been made to develop robust and accurate advanced classification methods. Deep classification models, similar to other ML models, can generally be categorized into supervised and unsupervised models. While in supervised classification approaches the deep networks are trained in an end-to-end fashion to learn the relevant features with respect to the provided class labels, unsupervised networks are trained to capture the representative features of the provided image data in the latent space, which will then be used for further clustering steps.

LeNet [188] and AlexNet [189], introduced over a decade ago, were among the first deep classifiers designed to classify image data in an end-to-end fashion. Both networks were shallow and designed with relatively similar network architecture. Specifically, both consisted of two and five convolutional layers, respectively, and employed kernels with large receptive fields in the earlier layers, while convolutions with smaller kernels were added at the top of the networks. Taking into account the impressive results reported by AlexNet in natural image classification tasks, it can be considered a fundamental model on which many of the advanced deep classifiers have been built. Accordingly, the exploration of novel architectures began in 2012 with a focus on deeper and wider networks.

The plain architecture of AlexNet was further elaborated by adding a stack of kernels of smaller size instead of using a single kernel with a large receptive field. This effective strategy led to conducting the same functions but with fewer parameters. Following a similar strategy, a deeper network of up to 19 layers, VGG19, was introduced, having small kernels with fixed sizes employed in each layer [190]. This model won the ImageNet challenge in 2014. Figure 3.13 shows the schematic architectures of AlexNet and VGG19. As can be seen,

the general design of the networks shared some similarities. In particular, the model architecture starts with a set of consecutive convolutional layers. As the model goes deeper, the depth of each layer increases while the width and heights of the feature maps decrease. The last convolutional layer is, finally, followed by a few dense layers at the top of the networks.

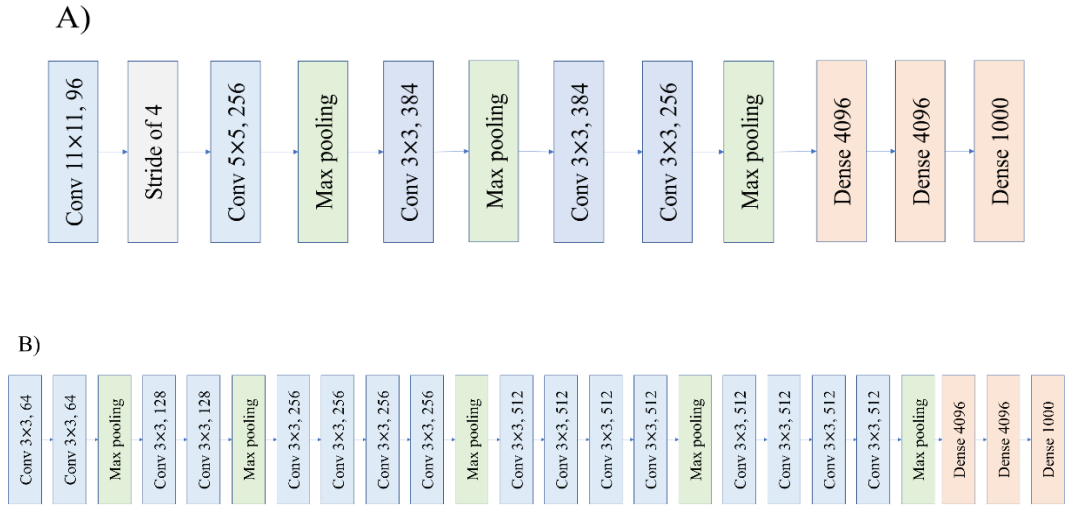


Figure 3.13. The network architecture of two famous deep classifiers: A) AlexNet model and B) VGG19 model. “Conv $m \times m, n$ ” refers to a convolutional layer with n feature maps, each of which has a kernel size of m . Similarly, “Dense x ” indicates a fully-connected layer with x number of neurons.

Of course, the advances of the deep classifiers are not limited only to increasing the depth of the networks, but more complex building blocks have been developed to improve the performance of the models while reducing the number of learnable parameters. Specifically, inception blocks were introduced in the GoogLeNet model and were designed in such a way that they need fewer parameters and less computational complexity than a single 3×3 or 5×5 convolutional layer by adding a set of convolutions of different sizes in one single module [191]. Increasing the network depths by stacking the layers on top of each other does not necessarily guarantee improving the model performance. Additionally, such deeper networks are much harder to train because of the notorious vanishing gradient issues. In particular, as the gradient is back propagated to earlier layers, a large number of sequentially repeated multiplications may make the values of the computed gradients very small. As a result, with increasing depth, network performance saturates or degrades rapidly. The ResNet architecture, the winner of the ImageNet challenge in 2015, introduced the residual block. Instead of directly learning the mapping from input data to output data, a residual block learns an additive term to add to the input data. Thereby, this approach is pre-conditioned towards learning a mapping close to the identity function in each layer.

With this strategy, they developed much deeper networks of up to 101 layers that could be trained effectively [192].

Since 2014, the classification performance on the ImageNet challenge has become saturated, which makes it difficult to evaluate whether the small increases in performance of the novel proposed models can really be attributed to more complicated learning procedures or network architectures. The privilege of the lower memory footprint these models provide is typically not as important for medical image applications as the classification performance in some challenging tasks is still poor and improving the model performance, nowadays, is more important than the memory efficiency issue. Consequently, AlexNet or other simple models such as VGG are still popular for medical image classification tasks [193].

The ordinary CNN architecture can easily employ multiple sources of information or representations from the input in the form of additional channels or additional paths presented to the input layer. This multi-channel/path input information can be efficiently integrated and merged at any point in the network. This strategy is beneficial for fusing different sources of information in different ways to improve classification performance by integrating domain-specific knowledge. As a result, multi-stream architectures have been developed for different tasks in the medical imaging domain [194][195].

One of the challenges of applying DL techniques to the medical imaging domain is often related to the difficulties of adapting the network architectures that were initially designed for computer vision tasks. Simply replacing the 2D convolutional modules with 3D convolutions will not only lead to a dramatic increase in the number of model parameters but also raise concerns over the functionality of the extended 3D model with respect to the original 2D model. As an alternative solution, orthogonal 2D slices from axial, coronal, and sagittal views have been extracted from the volumetric data to train so-called 2.5D multi-stream network architectures [196], [197].

3.6.4 Deep Segmentation

Segmentation of a target region within medical images is a challenging task due to the large variations in shape, size, and appearance of the anatomical targets as well as the lack of sharp boundaries between the target and surrounding tissues. This task can include the segmentation of organs, substructures, and lesions as a preprocessing step for further analyses and classifications. One of the common approaches in DL-based medical image

segmentation methods is to classify the individual voxels into foreground and background by extracting small image patches from the input image. This approach has been applied in a variety of applications, such as brain tumor segmentation [198], white matter segmentation in multiple sclerosis [199], and rectal cancer [200]. In this approach, well-established classification networks are often employed; however, such approaches analyze redundant data and are computationally inefficient because of the overlapping zones of the sub-images. In addition, classifying each voxel into foreground or background by considering only the local information ignores the global context, which endangers the model performance.

Fully Convolutional Neural Networks (FCNNs) can process the entire image or a large portion of the image at the same time and generate a segmentation map in which different labels will be assigned to the foreground(s) and background. Hence, these models do not analyze every single pixel one by one and therefore do not assign labels for individual pixels. Accordingly, FCNNs can set labels for all the pixel/voxel in the image simultaneously so that, computationally, they are much more efficient than the classification-based segmentation networks. Among the many developed segmentation models, the U-Net model is considered to be a breakthrough as it was primarily designed for biomedical image segmentation. Since the introduction of the U-Net model, it has been adopted extensively for a variety of medical image segmentation models [17]. U-Net is an FCNN model, and its structure is built upon an *encoder-decoder network*. In fact, such segmentation networks incorporate one FCNN as an encoder and another FCNN as a decoder [201]. The U-Net architecture incorporates symmetric encoder and decoder paths with four levels of resolution. In the encoder path, each resolution includes two consecutive convolutional layers with a kernel size of 3×3 and a conventional max-pooling with the size of 2×2 . Each of the convolutional layers is followed by a ReLU activation function. In the decoder path, with the help of upsampling operators, feature maps with lower resolutions are rescaled to higher resolutions. Each of these layers is followed by two convolutional layers with the kernel size of 3×3 and the same nonlinear function. To recover the spatial information lost in the downsampling encoder path, skip connections were introduced that add the feature maps from layers of equal resolution in the encoder and decoder paths together before proceeding with the next upsampling layer. Finally, the last convolutional layer used a pixel-wise softmax activation function to generate a probability map as the segmentation output. Figure 3.14 shows the original architecture of the U-Net.

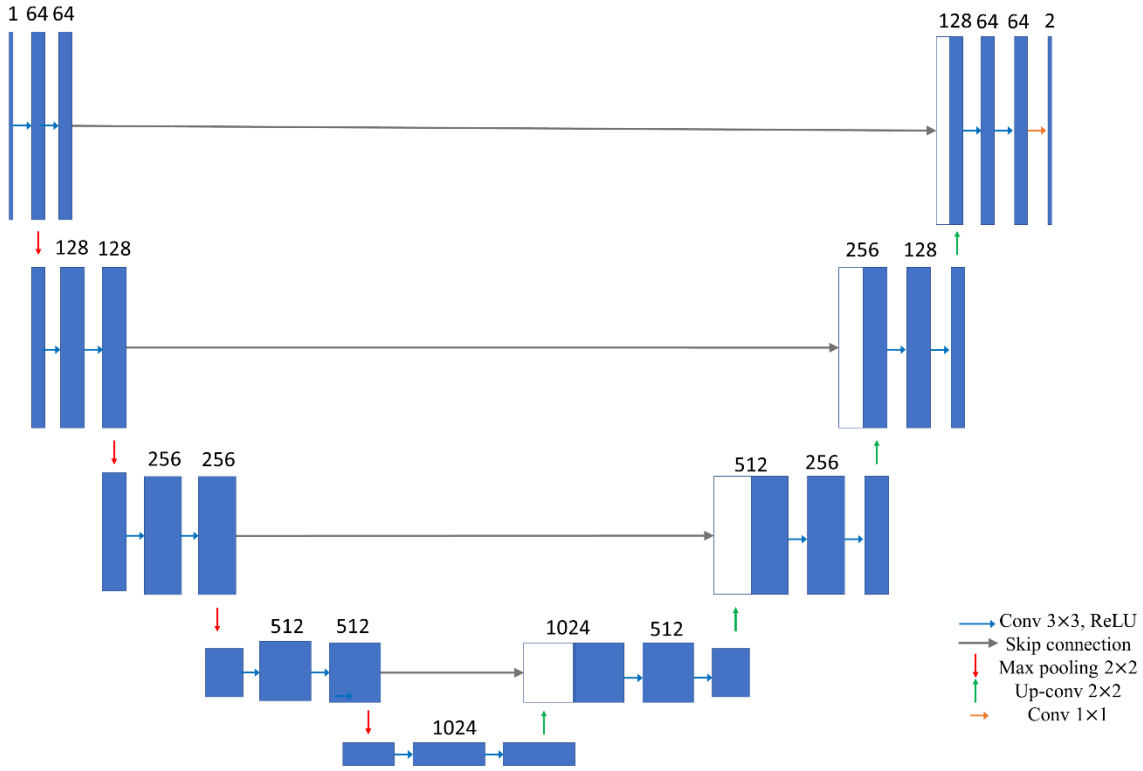


Figure 3.14. The architecture of the fully convolutional U-Net model consists of an encoder and a decoder which were linked together with skip connections. Each layer is presented with an integer value indicating the number of feature maps in the layer.

Another efficient model originally developed for medical image segmentation is V-Net. Like U-Net, this model follows an encoder-decoder architecture. The main difference between the V-Net and U-Net is that residual connections were employed instead of conventional convolutional layers in the V-Net. Another notable difference between these two networks is the strategy used in V-Net to reduce the resolution between the consecutive stages. In fact, in the V-Net model, the outputs of nonlinear functions go through a convolutional layer with a kernel size of $2 \times 2 \times 2$ and stride of 2. The authors showed that this strategy is memory efficient, especially in the case of volumetric medical images [202]. Progressive advances in the segmentation networks adopted from the original U-Net have led to the development of modern models with outstanding performance in a variety of challenging segmentation tasks, such as attention U-Net [203], inception U-Net [204], dense U-Net [205], recurrent U-Net [206], nested U-Net [207], adversarial U-Net [208], ensemble U-Net(s) [209], and nnU-Net [210].

Employing *prior information* along with the original image has been recognized as a classic strategy in medical image segmentation to deal with issues such as organ motion,

lack of sharp boundaries, or even missing boundaries. In the classical rule-based models, algorithms such as the active shape model [211], active appearance model [212], active appearance motion model [213], and multi-energy-based level sets [214]–[216] have been developed to improve the segmentation accuracy. The prior knowledge has also been integrated into the deep segmentation networks in different ways. In [217], the authors proposed a segmentation network containing two sub-networks, namely analysis and synthesis, in which a pre-trained analysis network was used to extract the image features that are then fed into an FCNN to obtain a global segmentation mask. The global masks, then, were refined by using the weights from the low-level layers of the analysis network. In [218] the authors proposed a cascade of two networks. The first network is a U-Net-like segmentation model. The output of the first model that often carries an under/over or incomplete segmentation mask was fed into an *AutoEncoder (AE)* network, called the shape regularization model. The reconstructed mask of the AE along with the U-Net output was incorporated into a custom loss function to update the segmentation network. Such an approach improved the segmentation accuracy of kidneys in ultrasound images by 5% in terms of the Dice metric. The authors in [219] used a pre-trained AE to incorporate the shape prior to the segmentation network. However, instead of only correcting the initial segmentation output, they regularize the weight adaptation process of a generic segmentation network by using the encoding component of the AE. To add the shape priors, unsupervised learning approaches such as Variational Auto Encoders (VAEs) (see Section 3.6.6) have been successfully employed [220]. As different imaging modalities represent different characteristics of anatomical organs, it has been verified that training a segmentation model with more than one modality would lead to better segmentation accuracy on the condition that the input images were already co-registered. This approach has been used in different tasks, such as brain tumor segmentation with different MR sequences [221].

Although all the mentioned segmentation models are trained with supervised signals, developing unsupervised segmentation models for medical images is an active field of research. In fact, one limitation of the supervised methods is their dependency on the large-scale annotated datasets, which are often scarce and costly to obtain. Therefore, developing segmentation models capable of learning the attributes of target regions directly from the unlabeled image data is of great importance for medical imaging communities. Specifically for the task of pathology segmentation, this task is recognized as Unsupervised Anomaly Detection (UAD), which will be discussed in Section 3.6.6.

3.6.5 Deep Inpainting

Image inpainting refers to the task of restoring the “damaged” part of the pixels in an incomplete image with semantically meaningful textures in order to reconstruct a realistic-looking image. In other words, the aim of such models is to replace the invalid pixels of an image with the learned features from the valid regions and synthesize a smooth image with contextual consistency. Some of the image inpainting applications include but are not limited to text removal, object removal, scratch removal, and repairing block masks. With the rise of DL methods, novel inpainting techniques have been developed primarily for computer vision tasks. However, conventional CNN models could not be directly employed as robust inpainting models. One of the main reasons behind such failures is the instability of the extracted convolutional features in both the learning and inference phases of the model. Specifically, convolutional kernels cannot constantly capture the information from the corrupted input images because the characteristics of the corrupted regions vary remarkably in different images. Therefore, the same feature values of the convolutional maps of an inpainting model might represent completely different textures which would not be compatible with all the corrupted images [222]. Another important reason why ordinary CNN models are unable to synthesize semantically correct images is that the spatially shared convolutional filters process all input pixels or features from both the valid and invalid regions as the same valid ones. In fact, vanilla convolutions apply the same filters on all the valid and invalid pixels, which leads to filling the holes without preserving the context, color, and textural consistency.

Several novel approaches have been proposed to address the limitations of ordinary CNN models for inpainting purposes. To fill the large continuous holes in the image by progressively imposing constraints for the hole centers, a Recurrent Feature Reasoning (RFP) network was developed to recurrently infer and gather the hole boundaries for the encoded feature map [223]. Specifically, the RFP module utilizes the correlation between the surrounding pixels and strengthens the constraints for estimating deeper pixels. The RFR module, which is a plug-and-play module with a recurrent inference design, can be added to any part of the network. This module is built on top of the three submodules: (a) the area identification module, which is used to identify the area to be inferred in this recurrence; (b) a feature reasoning module for inferring the content in the identified area; and (c) a feature merging module that merges the intermediate feature maps. To recover features with high fidelity, the RFP module was fortified by adding a certain attention module that can adaptively combine the scores from different recurrences. This strategy ensures the

consistency between patch-swapping processes among recurrences and results in synthesizing high-quality images that preserve fine-grained details. Yu et al. proposed a GAN-based model that functions in two stages [224]. In the first stage, a simple dilated convolutional network is trained and optimized with a reconstruction loss to roughly fill the invalid pixels. A contextual attention mechanism was developed to use the features of valid regions as convolutional filters in order to process the generated patches. This attention mechanism, which is integrated into the second stage, is designed for the following tasks: (a) to match the generated patches with valid contextual patches; (b) to weigh relevant patches with channel-wise softmax; and (c) to reconstruct the generated patches with contextual patches with deconvolution. This attention module also entails a spatial propagation layer to encourage spatial coherency of attention. The whole network is trained end-to-end with reconstruction losses and two GAN losses, whereby one critic looks at the global image while the other looks at the local patch of the missing regions.

Even though promising results have been achieved by the many recent methods, they still suffer from a number of limitations. First, conditioning the output on the invalid pixels results in reconstructed images with visual artifacts, especially for the large-sized images that were corrupted with large holes. A second limitation is that they often focus on filling the holes with regular shapes such as rectangles or squares. Nevertheless, such restrictions can potentially increase the risk of overfitting, resulting in the learned model properly inpainting only the holes with regular learned shapes. To efficiently address these limitations, a Partial Convolutional (PConv) neural network was developed to replace the invalid irregular-shaped pixels with high-fidelity patterns. Assume that $I_{(y,x)}$ stands for the input image with c -channels (or a feature map at the current layer), and W represents a set of convolutional filters. The conventional convolutional operator filters the input image and returns a c' -channel output, $O_{(y,x)}$:

$$O_{(y,x)} = I_{(y,x)} * W_{(y,x)} = \sum_{i=-k'_m}^{k'_m} \sum_{j=-k'_n}^{k'_n} W_{k_m+i, k_n+j} \cdot I_{y+i, x+j}$$

where k_m and k_n are the size of the kernel, $k'_m = \frac{k_m-1}{2}$ and $k'_n = \frac{k_n-1}{2}$. In this notation, for simplicity, the bias term is not presented. As was previously explained, this type of convolutional operator is not suitable for image inpainting applications as it ignores the presence of holes within the subregions and simultaneously takes into account both the valid and invalid pixels. As a result, the inpainted holes do not fully match the nearby textures,

and the generated images contain textural/color inconsistencies. The Pconv operator uses a binary mask M with the same size as the input image to condition the convolutional computations on the valid pixels. The partial convolution at every spatial location for the current sliding window can be defined as:

$$O_{(y,x)} = \begin{cases} W_{(y,x)}^T (I_{(y,x)} \odot M_{(y,x)}) \frac{\text{sum}(1)}{\text{sum}(M)} & \text{if } \text{sum}(M) > 0 \\ 0 & \text{otherwise} \end{cases}$$

where \odot denotes element-wise multiplication, 1 is an all-one matrix with the same size as M . Compared to the ordinary convolution operator, one can understand that the output values of Pconv depend only on the valid areas defined by the binary mask (M). The term $(\frac{\text{sum}(1)}{\text{sum}(M)})$ functions as a scaling factor to adjust for varying amounts of valid inputs. After each Pconv operator, the binary mask will be updated by the following rule: If Pconv could condition its output on at least one valid pixel, that spatial location will be updated to become a valid pixel. This strategy was implemented in a U-Net-like network (see Figure 3.15) in which the ordinary convolutions were replaced by the described Pconv modules [225]. To fill the holes with meaningful semantic patterns, the proposed model is optimized with a multi-term objective function. Let the input image with holes be I_{in} ; I_{gt} represents the original image without holes (ground truth), I_{out} indicates the predicted image, and M denotes the binary mask used for corrupting the image; the first two terms in the loss function are pixel-wise errors that can be calculated separately for the valid and invalid zones with L1-norm. These two terms aim to minimize the intensity differences between the predicted and ground-truth images inside and outside the hole regions, respectively:

$$\mathcal{L}_{valid} = \frac{1}{N_{I_{gt}}} \|M \odot (I_{out} - I_{gt})\|_1$$

$$\mathcal{L}_{hole} = \frac{1}{N_{I_{gt}}} \|(1 - M) \odot (I_{out} - I_{gt})\|_1$$

where $N_{I_{gt}}$ shows the number of pixels in the I_{gt} .

Perceptual loss is the third employed objective function that aims to minimize the differences between the high-level feature representations extracted from the predicted and ground-truth images. It calculated the L1 norm between two sets of high-level features extracted from I_{out} and I_{comp} where I_{comp} is the composite output, which is similar to the

predicted image but with the intensity of valid pixels replaced by those of the ground truth. The high-level features are extracted from the 1st, 2nd, and 3rd layers of a VGG16 [190] network pre-trained on ImageNet:

$$\mathcal{L}_{perceptual} = \sum_{p=0}^{p-1} \frac{\|\Psi_p^{I_{out}} - \Psi_p^{I_{gt}}\|_1}{N_{\Psi_p^{I_{gt}}}} + \sum_{p=0}^{p-1} \frac{\|\Psi_p^{I_{comp}} - \Psi_p^{I_{gt}}\|_1}{N_{\Psi_p^{I_{gt}}}}$$

Here, $\Psi_p^{I_*}$ refers to the outputs of the activation function of the p th layer of the pre-trained network, given the input I_* .

In order to minimize the differences in style between the synthesized and ground-truth images, style loss was computed as well. To reconstruct images with a high level of style similarities inside and outside the hole regions, the style loss was calculated for predicted and composite images separately:

$$\begin{aligned} \mathcal{L}_{style_{out}} &= \sum_{p=0}^{p-1} \frac{1}{C_p^2} \left\| \frac{1}{K_p} ((\Psi_p^{I_{out}})^T (\Psi_p^{I_{out}}) - (\Psi_p^{I_{gt}})^T (\Psi_p^{I_{gt}})) \right\|_1 \\ \mathcal{L}_{style_{comp}} &= \sum_{p=0}^{p-1} \frac{1}{C_p^2} \left\| \frac{1}{K_p} ((\Psi_p^{I_{comp}})^T (\Psi_p^{I_{comp}}) - (\Psi_p^{I_{gt}})^T (\Psi_p^{I_{gt}})) \right\|_1 \end{aligned}$$

As can be seen, the style loss is similar to the perceptual loss; however, it first calculates the autocorrelation of extracted features and then computes the L1 norm. In this notation, C_p indicates the depth of channels in Ψ_p , and K_p refers to the number of elements in Ψ_p tensor.

The final loss used in this multi-objective function is Total Variation (TV) which is a common noise reduction method that can attenuate the impact of high-frequency noise. In fact, it functions as a smoothing term that makes the intensity values of the neighboring pixels closer to each other:

$$\mathcal{L}_{tv} = \sum_{(i,j) \in R^2} \frac{\|I_{comp}^{i,j+1} - I_{comp}^{i,j}\|_1}{N_{I_{comp}}} + \sum_{(i,j) \in R^2} \frac{\|I_{comp}^{i+1,j} - I_{comp}^{i,j}\|_1}{N_{I_{comp}}}$$

where $N_{I_{comp}}$ is the number of pixels in the composite image.

The final loss function, then, is:

$$\mathcal{L}_{total} = \mathcal{L}_{valid} + \mathcal{L}_{hole} + 0.05\mathcal{L}_{perceptual} + 120(\mathcal{L}_{style_{out}} + \mathcal{L}_{style_{comp}}) + 1\mathcal{L}_{tv}$$

Although the Pconv model synthesized seamless, high-quality images with a high level of fidelities in the inpainted regions, the rule-based mask updating procedure faces some limitations. First, all feature channels in each convolutional layer share the same mask regardless of their inconsistencies. This limitation will be problematic, especially for multi-channel input images such as multimodal PET-CT slices. A second limitation is that the binary mask will be updated constantly unless all the invalid pixels disappear, regardless of how many pixels were covered in the previous layers. The Gated Convolution (Gconv) operator [226] has been proposed to turn the problematic rule-based mask updating of Pconv into a dynamic learnable procedure. Specifically, gated convolutions learn soft mask automatically from the image/feature maps. This will enable the convolutional operators to learn the dynamic feature selection mechanism for each channel and each spatial location independently. This process can be formulated as:

$$Gating_{(y,x)} = \sum \sum W_g \cdot I$$

$$Feature_{(y,x)} = \sum \sum W_f \cdot I$$

$$O_{(y,x)} = \varphi(Feature_{(y,x)}) \odot \sigma(Gating_{(y,x)})$$

where σ refers to the sigmoid function that scales the output of the gating signal into the range of 0 to 1; φ can be any kind of nonlinear activation function; W_g and W_f are two separate convolutional filters.

Although the image inpainting models are primarily developed for computer vision tasks, they have been adopted for different applications in the medical imaging domain. Such applications include but are not limited to anomaly removal [227], image augmentation [228], artifact removal [229], and synthesizing fake healthy structures from deformed ones [230].

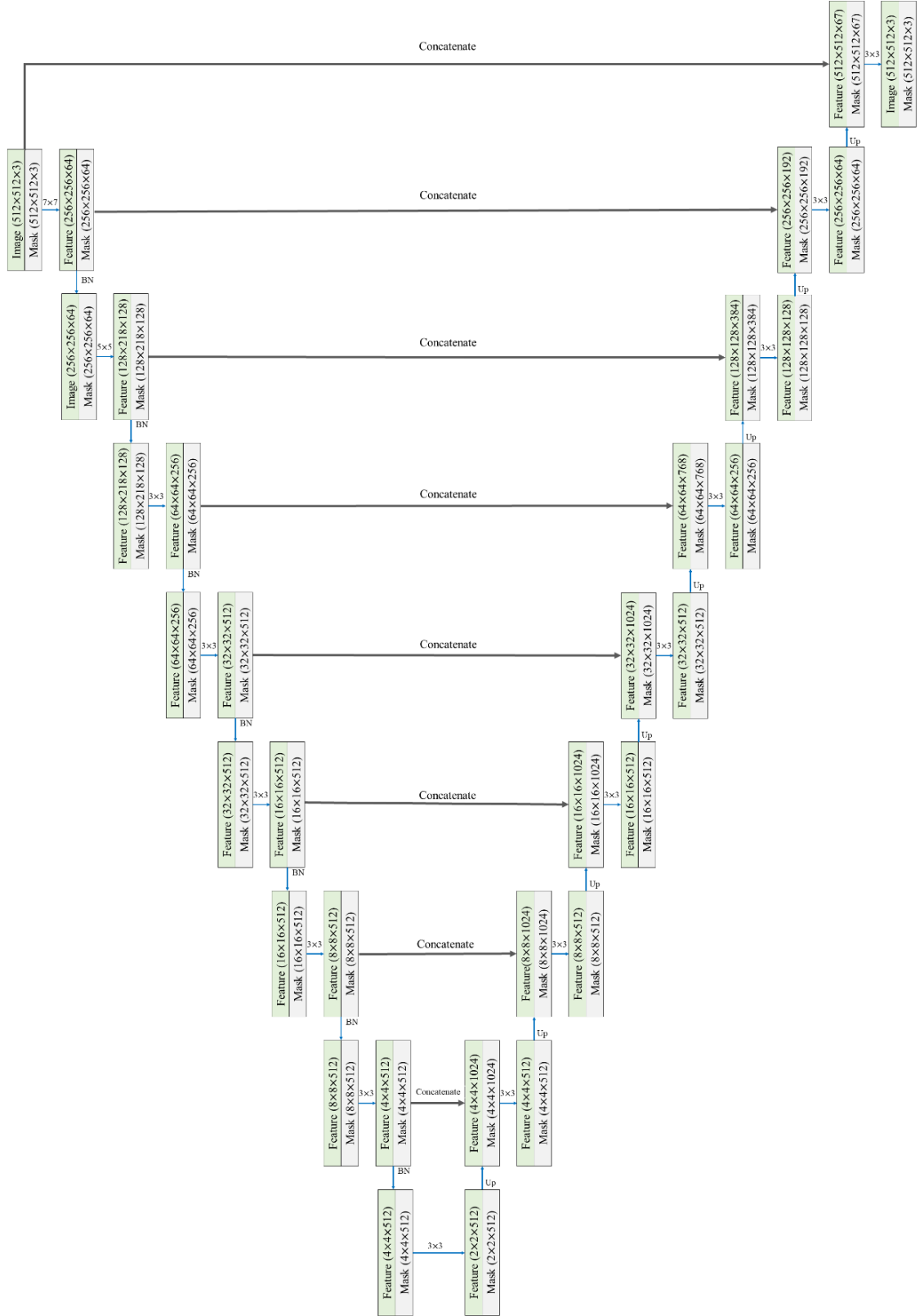


Figure 3.15. The network architecture of the PConv model. In this figure, BN refers to the batch normalization module, and Up indicates the upsampling operators.

3.6.6 Deep Representation Learning

Representation learning refers to the task of finding a discriminative feature representation from the studied data for a certain task. Depending on whether the training data is labeled or not, representation learning can be categorized into supervised and unsupervised methods. In the supervised representation learning approach, a model is trained to properly map the input data into the corresponding target labels by minimizing a loss function. On the other hand, unsupervised representation learning aims to learn the underlying semantic feature space directly from the input data without any supervised signals [231]. Learning effective data representation space has been a longstanding goal in ML communities. In general, representations are considered “effective” if they capture discriminative factors of the data, which are useful for the studied task. Learning effective data representations in medical imaging domains poses additional challenges. Such a representation must lend itself to a wide range of clinically useful tasks and, at the same time, should work properly across data acquired from different imaging vendors and modalities [231]. In this context, supervised and unsupervised representation learning with deep networks has achieved promising results in various tasks. For example, Liao et al. developed a DL framework to learn a discriminative representation for automatic prostate segmentation in MR images [232]. Ning et al. developed a deep cross-view co-regularized representation learning framework to learn latent view-specific representations based on cross-view images generated from MRI. The proposed model was tested on the challenging task of glioma subtype identification which resulted in outstanding performance [233]. Liu et al. addressed the problem of incomplete multi-modality images by learning the effective representation space of the available multi-modal data. In other words, they proposed an AE-based Multi-View missing data Completion framework (AEMVC) to learn common representations for Alzheimer’s disease by mapping the original complete view to a latent space. After this, the latent representations measuring statistical dependence learned from the complete view are used to complement the kernel matrix of the incomplete view [234].

One of the interesting fields of study within representation learning is unsupervised segmentation. Before detailing the procedure of this approach, it should be emphasized that accurate results in different tasks have already been achieved by supervised segmentation models. However, the performance of supervised models depends heavily on the quality and quantity of the data. Regardless of the difficulties in annotating large-scale datasets, the generalization power of the supervised segmentation models is restricted to the type of annotated data presented in the training set. However, there is always the risk that a certain

type of data would be under-presented or even not presented in the training set, such as some rare diseases with a limited amount of available data. In such cases, the model would fail to generate a meaningful segmentation mask. Among other methods, Unsupervised Anomaly Detection (UAD) aims to address the above-mentioned challenges. While the applications of UAD methods in the medical imaging domain have been studied for a long time, the ability of convolutional networks to learn sophisticated distributions of the high dimensional data makes the DL-based UAD method an active field of research. DL-based models such as AEs are capable of learning non-linear transformations of input data into a low-dimensional representation space from which cluster-based anomaly detection can be achieved. In recent years, there has been a plethora of studies based on the abilities of AEs and GANs to not only detect the anomalies within medical images but also to localize and even segment the boundaries of anomaly regions directly in the image space.

In this domain, the main hypothesis is that the distribution of the healthy images differs from images containing pathologies. Therefore, with the help of DL-based representation learnings, the networks are trained to model the distribution of healthy anatomies. The trained network, then, is used to detect the anomalies as outliers from the normative learned distribution. In other words, once the model is trained, it is supposed to reconstruct only healthy anatomical structures from the input data. Thus, in the test phase, if the network is fed with a healthy image, the model should reconstruct a healthy image, ideally without any anatomical distortion. On the other hand, if the input image contains some pathologies such as tumors or lesions, the model should, ideally, reconstruct only the healthy anatomical parts of the input and replace the pathological regions with the learned appearance of healthy anatomies. It is worth mentioning that the majority of the research within this field has been dedicated to segmenting the brain lesions such as Multiple Sclerosis (MS) and glioma within specific MR sequences in which the lesions tend to appear with hyperintensity signals [235]–[240].

Among the other models, VAEs [241] are efficient frameworks for approximating the distribution of complex high-dimensional data. In fact, capturing the complex nonlinear association between the dependent image elements is a challenging task. Having learned such an intricate distribution, the second challenge for a model is generating a realistic image. In other words, the well-trained model should first decide which image to synthesize before it assigns values to the pixels. This process is related to the concept of latent variables. It is called latent because, given just an image generated by the model, it is not known which settings of the latent variables produced that image [242]. VAEs are a special type of general

AE model that can regularize the encoded distribution during the training procedure so as to ensure that effective representations are learned in the latent space from which the generation of new relevant data would be achievable from the decoder part of the learned model. In contrast to the ordinary AE models, which encode the input data as a single point, VAEs encode the input data as a distribution over the latent space. The VAE models are trained with the following ordinal protocols: (a) input data is encoded into a distribution over the latent space; (b) a point is sampled from the latent distribution; (c) the sampled point passed through the decoder network to reconstruct a new output from which the per pixel reconstruction loss is computed; and (d) the whole model is optimized by backpropagating the reconstruction error. There are several advantages to regularizing the latent distributions, including that (a) it forces the learned distribution by the encoder part of the model to follow a standard distribution, and (b) continuity of the regularized learned distribution is beneficial to reconstructing two similar outputs if they were sampled from two close points. In practice, the encoded distributions can be chosen to follow a Gaussian distribution so that the encoder can be trained to return the two parameters of the Gaussian, namely as the mean and covariance matrix. Thus, the objective function of the VAEs should include a regularization term in addition to the ordinary reconstruction loss:

$$L_{VAE} = L_{rec} + L_{reg} = MSE(input, reconstructed) + KL\{z, N(0, I)\}$$

where MSE indicates the mean square error (L2) and KL stands for Kullback-Leibler, which is a measure to quantify how one PDF $p(x)$ diverges ($||$) from another PDF $q(x)$:

$$D_{KL}(p(x)||q(x)) = \int_x p(x) \log \frac{p(x)}{q(x)} dx$$

This measure converges to the minimum value, zero, if two distributions are similar. However, it should be noticed that KL divergence is an asymmetric measure. Specifically, if $p(x)$ is close to zero and $q(x)$ is far from zero, the effect of $q(x)$ might be disregarded. Jensen-Shannon (JS) divergence is a more smooth and symmetric measure that quantifies the similarities between two PDFs:

$$D_{JS}(p(x)||q(x)) = \frac{1}{2} D_{KL} \left(p(x) \left\| \frac{p(x) + q(x)}{2} \right\| \right) + \frac{1}{2} D_{KL} \left(q(x) \left\| \frac{p(x) + q(x)}{2} \right\| \right)$$

In brief, the objective of the reconstruction term is to make the encoding-decoding as performant as possible; moreover, the regularization term is applied to the latent layer to make the distribution captured by the encoder as similar as possible to the predefined

normal distribution ($N(\mathbf{0}, I)$). Accordingly, in a sense, the regularization term measures the similarity between two PDFs. As KL divergence between two Gaussian distributions has a closed-form and can be expressed in terms of mean (μ) and covariance (σ) matrices, it is conventionally used as the regularization term. Figure 3.16 shows a graphical illustration of the VAE architecture.

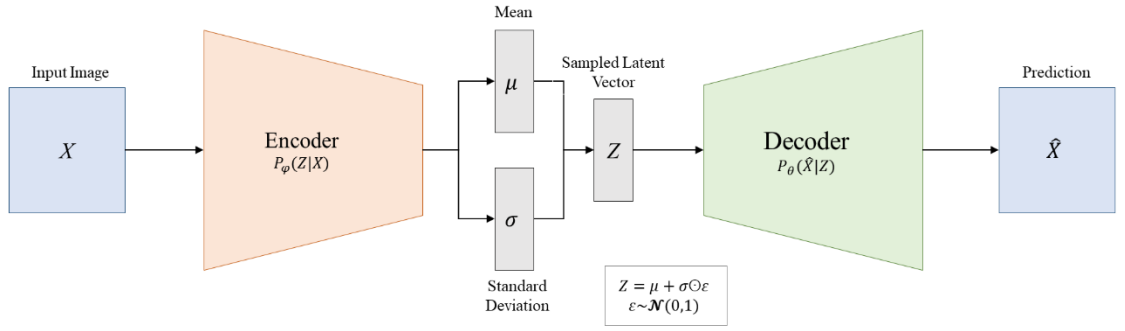


Figure 3.16. The general architecture of the variational autoencoder model consists of an encoder, a sampling layer, and a decoder to generate an output from the sampled latent vector.

Chapter 4

Methods

4.1 Deep Learning Methods for Tumor Segmentation (Papers I, II)

One of the central tasks in medical image analysis applications is to segment the anatomical organs and/or pathological regions [18]. In fact, accurate and reliable segmentation of the target regions has a critical impact on further diagnosis and intervention steps. The first two papers presented in this thesis focused on developing advanced instance segmentation models for the task of automatic tumor segmentation. Paper I aims to develop a supervised prior-aware segmentation model to improve the segmentation accuracy of lung pathologies by capturing the prior information about the pathological regions. In Paper II, an unsupervised anomaly detection method is proposed by developing an autoinpainting model to automatically segment the lung and head-neck tumors with a purely unsupervised approach.

4.1.1 Overview of the Studies

The main objective of Paper I is to develop a DL pipeline capable of learning the appearance of lung tumors in CT images. Therefore, the learned characteristics of the tumors can be encoded as prior information into a segmentation model to boost the segmentation accuracy of lung tumors. Integrating prior knowledge such as shape, size, and appearance of candidate regions into segmentation models has been shown to be a beneficial strategy to improve segmentation performance [18], [243]. In Paper I, to obtain prior knowledge about complicated lung tumors appearing with a wide range of characteristics, we proposed to first capture the prior knowledge about the normal lung anatomies instead of lung pathologies and then derive the prior knowledge of tumoral regions indirectly. Finally, the obtained prior knowledge was encoded into a supervised segmentation network. Hence, the proposed pipeline consists of three modules: (a) an inpainting model to synthesize faked healthy images, (b) a Normal Appearance Autoencoder (NAA) model to automatically reconstruct tumor-free images, and (c) a prior-aware segmentation network to delineate the tumor boundaries.

Paper II attempts to address one crucial challenge faced by supervised segmentation models: dependency on the large-scale labeled dataset. In fact, supervised training of data-greedy segmentation models suffers from two limitations. First, the number of training medical images is often limited because of the costly slice-by-slice data annotation procedure. Second, even if large-scale training data is available, the generalization power of the learned models over the unseen classes is poor, which necessarily requires the collection of annotated data from the new class followed by retraining of the model [19]. In the medical image analysis domain, UAD is an active field of research that aims to model the distribution of healthy anatomies by training deep representation learning models in order to detect anatomical abnormalities as outliers with respect to the learned distributions. In this context, a variety of AE-based and GAN-based models have been proposed; however, these models suffer from a number of limitations. First, to learn the distribution of healthy organs, these models are often trained with low-resolution images; therefore, they will not be able to learn the fine-grained details, which are extremely important for several applications. Second, they often focus on detecting anomalies that appear with hyperintensity signals, such as MS lesions in specific brain MR sequences. A third limitation is that they face difficulties in preserving the anatomical constraints within the generated images. As an attempt to tackle the limitations of conventional UAD models, a fully unsupervised autoinpainting model is proposed in Paper II that is capable of automatically replacing the tumoral regions with the appearance of healthy tissues in full-resolution images while preserving the anatomical constraints. Having compared the input tumoral images with the synthesized tumor-free images, the tumoral regions can be segmented by applying a set of post-processing steps.

4.1.2 Image Preparation and Preprocessing

Paper I. Contrast enhancement was the only main preprocessing step applied to the employed CT volumes. To enhance the contrast between the tissues inside the lung regions, the Hounsfield values of the CT volumes were clamped to the range of $[-1000, 500]$. Figure 4.1 shows the impact of different windowing levels on the appearance of a chest CT image. The intensity windowing operator can be simply presented with:

$$I'_{(x,y,z)} = \begin{cases} I_{(x,y,z)}, & \min < I < \max \\ \max, & I \geq \max \\ \min, & I \leq \min \end{cases}$$

where $I_{(x,y,z)}$ refers to the voxel intensity of coordinate (x,y,z) , and min and max show the lower and upper boundaries of the applied window.

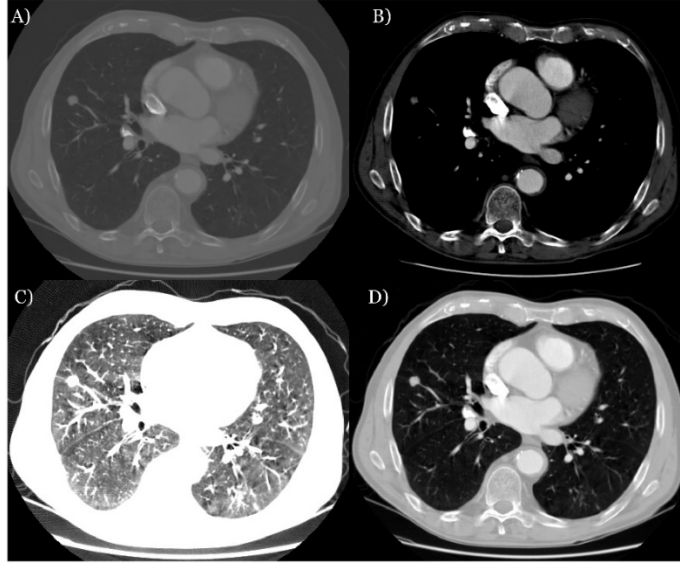


Figure 4.1. Modification of CT grayscale components by histogram windowing. Visualization of the same slices with intensity range of: A) the originally acquired range [-2000,3000]; B) [-100,500]; C) [-1000,-700]; and D) [-1000, 500].

Since the models were examined with 2D images, the axial slices were extracted from the 16-bit volumes and converted into 8-bit grayscale images. Finally, prior to feeding the images to the networks, their intensity ranges were normalized in the range of [0,1].

Paper II. The following preprocessing was applied to the employed multimodal Non-Small Cell Lung Cancer (NSCLC) and Head-Neck (HN) datasets. First, the voxel spacing of the PET data was resampled to those of the corresponding CT data by using a third-order Spline interpolation method. Second, the intensity values of PET images were converted into SUV units. Third, to enhance the contrast between the tissues within the target anatomies, the intensity values of CT and PET images were clamped. Specifically, the Hounsfield values of CT images were clamped into the range of [-1000,500] for NSCLC data and [-200,200] for the HN dataset. The SUV values of PET images were constrained in the range of [0,12] as well. Similar to Paper I, the axial slices from 16-bit volumes were extracted and converted into 8-bit gray-level images with the size of 512×512 pixels. Finally, the intensity range of images was rescaled into the range of 0 to 1.

4.1.3 Prior-Aware Tumor Segmentation

In Paper I, a pipeline was proposed to capture the prior knowledge of lung tumors, inspired by the concept of UAD methods but with a rather different approach. Unlike the

conventional UAD models that aim to remove pathologies with hyperintensity signals, lung tumors appear with characteristics highly similar to surrounding soft tissues. In practice, it is not possible to discern the healthy soft tissues from tumoral regions by relying only on the intensity difference patterns. Therefore, we turned the training procedure of representation learning models from an unsupervised into a supervised approach. In other words, we aimed to enforce the model to learn the appearance of healthy lung anatomy by explicitly setting healthy images as target labels for the corresponding input pathological data. However, it is rare in actual cases—if not impossible—to have both a healthy scan and tumoral scan of the same subject. To address this practical limitation, employing the concept of image inpainting allows paired images to be synthesized by carefully inpainting the tumoral regions from the pathological data. Hence, “faked” healthy images were synthetically generated to form paired images of real tumoral images and faked tumor-free images. Building on the paired images, we modified the conventional VAE model and turned it into the supervised NAA model to automatically remove the tumoral regions from the input tumoral slices and reconstruct normal-looking slices. Accordingly, prior knowledge of the tumors was then estimated by finding the differences between the original tumoral images and the corresponding tumor-free images generated by the NAA model. The captured prior knowledge was then encoded into a U-Net segmentation model as an additional input channel to guide the attention of the segmentation model to the regions estimated by prior knowledge. Furthermore, to quantify how well the proposed NAA model could estimate the tumor appearance, post-processing steps were applied to the residual images to segment the tumors directly. It is worth noting that in Paper I, in addition to the studied lung nodules and NSCLC datasets, the ability of the developed model was investigated for COVID-19 lesions segmentation as well because Covid lesions share similar characteristics to some extent with respect to the NSCLC pathologies in CT volumes. Figure 4.2 shows the general pipeline of the study.

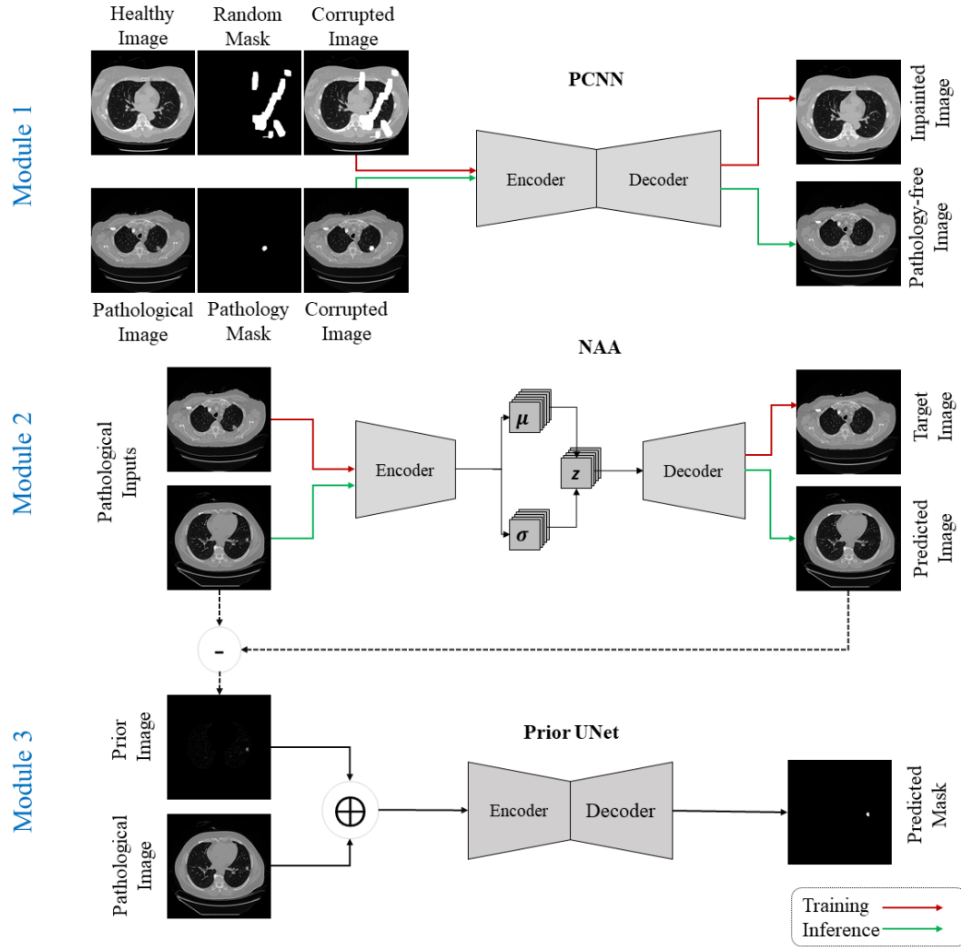


Figure 4.2. The pipeline of the proposed model of Paper I consists of three modules: an inpainting module, a normal appearance autoencoder module, and a prior-aware segmentation module. [From Astaraki et al. 2022, Reprinted with permission]

Generating healthy lung images. Synthesizing healthy images from pathological images can be achieved by inpainting the tumoral regions. In other words, the inpainting model considers the tumor region as a hole that should be filled with the characteristics of surrounding tissues. To inpaint the tumors, the Partial Convolutional Neural Network (PCNN) model, which was described in Section 3.6.5, was employed in two phases. First, 20,000 healthy image slices were extracted from the LIDC-IDRI dataset [244] with the full-resolution of 512×512 , which were corrupted with random holes. The PCNN model was trained with the corrupted images and was able to successfully learn to replace the holes with semantically meaningful anatomical patterns. Random holes were generated by combining regular geometric shapes such as circles and ellipses. The size of the random holes was designed to resemble the diverse size of lung pathologies which, on average, occupy 25 to 30% of the image size. The second phase is the test phase. The label masks of the tumoral slices were used to corrupt the tumoral locations from which the learned model

fills the holes with the appearance of healthy tissues. As a result, for each pathological input slice, a faked high-quality pathology-free image was synthesized to form a set of paired images.

Automatic pathology removal. As a powerful representation learning model, VAEs can transform the high dimensional input images into the probabilistic latent space to reconstruct new images with acceptable quality; accordingly, the model has been widely used for UAD tasks [236]. To further enforce the conventional VAE to compress the high-resolution CT slices into low-dimensional latent manifolds and reconstruct high-fidelity images while suppressing the pathologies, the following modifications were applied:

1. Building on the generated paired images from the PCNN step, the model training was changed from unsupervised to supervised. In other words, pathological images were fed as model inputs, and corresponding fake pathology-free images were set as target images.
2. To preserve the spatial context and geometric features in the latent manifolds, dense layers were replaced by convolutional layers.
3. To mitigate the inconsistency between the latent manifold of the input tumoral images and latent manifolds of the target images, a regularization term was embedded in the loss function. This regularization aims to minimize the differences between the latent variables of input pathological images and target pathology-free images.
4. Reconstruction loss was updated with higher weights within the lung regions by integrating lung field masks into the optimization process. This modification was done by adding a weight map into the reconstruction loss to assign higher weights to the regions inside the lungs and lower weights to the areas outside.

This model is named NAA. The described modifications lead to the following multi-loss function:

$$\mathcal{L} = \omega \odot \mathcal{L}_{rec} + \alpha \mathcal{L}_{reg} + \mathcal{L}_{prior} = \omega \odot MSE(y, y') + \alpha MSE(z(x), z(y)) + D_{KL}(z, N(0, I))$$

where x , y , y' , and z denote the input, target, model prediction, and latent variable, respectively. MSE denotes the element-wise mean square error and \odot indicates element-wise multiplication. The coefficient α balances the effect of the regularization term through

a scalar multiplication, and its value was experimentally chosen as 0.001. Finally, the symbol ω shows the weighting map of the lung fields.

Pathology segmentation as anomalies. To determine how well the NAA model could remove the tumors from the image contents while preserving the normal structures, differences between the tumoral images and the outputs of the NAA model were used to directly segment the pathologies. Accordingly, the following steps were examined:

1. Residual images were formed by finding the pixel-wise differences between the tumoral slices and the reconstructed slices.
2. For each subject, the residual slices were aggregated to create the residual volumes, which were multiplied by the lung field masks.
3. Otsu thresholding was applied to the residual volumes to remove the background noises.
4. Morphological operators were used to discard the discrete random tiny components of the thresholded images. The final binary volumes were then compared against the label masks.

Prior-aware U-Net model. The residual images, which were computed as differences between the input slices and the predictions of the NAA models, were employed without any further processing as the prior image to boost the segmentation accuracy. These residual slices, which may contain a highlighted region or regions that represent the attributes of pathologies such as shape, intensity/textural distribution, and location can enforce the attention of the segmentation network toward the candidate regions. On the other hand, if they do not entail informative knowledge, they should not degrade the functionality of the segmentation network. Therefore, the prior images were integrated into the U-Net model as a second channel along with the original slices. In other words, the segmentation network is fed with 2-channel input data in order to predict the binary label masks.

Lastly, since the calculation of prior knowledge is independent of the segmentation pipeline, the obtained prior image can be integrated with different approaches into different segmentation models. To benchmark the efficacy of the proposed pipeline, the standard nnU-Net model [210] was adopted and tested both as the baseline and the 2-channel input data.

4.1.4 Unsupervised Tumor Segmentation

In Paper II, an unsupervised segmentation model was built upon an autoinpainting method for the task of tumor segmentation in single/multimodal medical images. Specifically, a robust inpainting method was proposed that is capable of reconstructing realistic-looking multi-modal medical images from corrupted ones while preserving fine-grained details. The developed inpainting model was trained by carefully corrupting the healthy images with random irregular holes to simulate the morphological characteristics of heterogeneous tumors. The learned model was then used for automatic tumor removal in the test phase in an autoinpainting pipeline. Specifically, a set of subregions within the main target organ is defined through a sliding window approach to be inpainted. The autoinpainting procedure is followed by a post-processing strategy to detect the candidate region for the final tumor removal. The proposed inpainting model is optimized with a multi-term objective function in order to replace the holes with plausible image characteristics and preserve the anatomical constraints.

Image inpainting for learning the appearance of healthy anatomies. Inspired by the concept of the PConv and GConv operators, which were described in Section 3.6.5, a U-Net-like architecture was developed for the inpainting task. In general, the architecture framework is similar to the PCNN model; however, the PConv operators were replaced by the Gconv modules. Specifically, the encoder part of the model consists of 8 Gconv blocks, each of which includes a Gconv layer with the stride of 2, followed by an optional BN layer and a ReLu activation function. The decoder stage of the model, similarly, contains 8 Gconv blocks, each of which consists of a nearest neighbor upsampling layer, a Gconv layer, an optional BN layer, followed by a LeakyReLu activation function. Skip connections connected the corresponding blocks from the encoder path to the decoder path by concatenating the two feature maps and corresponding binary masks, respectively. The final output layer of the model is an ordinary convolutional layer activated with a sigmoid function, which is fed by a concatenation of the last Gconv block from the decoder path and the original input image with holes, along with the original binary mask (see Figure 4.3). Similar to the PCNN model, a multi-term objective function was employed to reconstruct images with both textural and conceptual similarities. However, in addition to the pixel-wise mean square error loss, perceptual loss, style loss, and total variation loss, another loss term was included to preserve the high-frequency patterns. Specifically, since the early layers of the model focus on learning the edge-based features, the loss terms previously mentioned by themselves cannot preserve the high-frequency attributes well. This issue will be problematic,

especially, when input data consists of different channels, each of which carries different contents such as multi-modal PET-CT images. Therefore, to preserve the edge information and synthesize images with high-frequency details as much as possible, the Laplacian (lap) pyramid loss was added to the objective function:

$$\mathcal{L}_{lap(I_{out}, I_{gt})} = \sum_j 2^{2j} \|L^j(I_{out}) - L^j(I_{gt})\|_1$$

where $L^j(x)$ indicates the j th level of the Laplacian pyramid representation of input x . In this study, the parameter j was set to 3, that is, three levels of pyramid representations were computed.

Therefore, the overall objective function is the combination of all the loss terms mentioned:

$$\mathcal{L}_{total} = 30\mathcal{L}_{valid} + 240\mathcal{L}_{hole} + 0.2\mathcal{L}_{perceptual} + 0.05(\mathcal{L}_{style_{out}} + \mathcal{L}_{style_{comp}}) + 250\mathcal{L}_{tv} + 20\mathcal{L}_{lap}$$

The proposed inpainting model, named Gconv_{Lap}, was used to learn the appearance of healthy anatomical structures by learning to replace the irregular holes with the attributes of healthy tissues. Having the corrupted images as input to the model on one side and the original images as the ground truth on the other side, the inpainting model is trained to smoothly replace the holes with semantically meaningful patterns in order to synthesize realistic-looking images while preserving fine-grained details and anatomical constraints. Considering that tumors appear with irregular shapes and different sizes at different locations, the holes should be generated in a way to imitate the visual attributes of the tumors. Therefore, irregular holes were synthesized by carefully combining the ordinary regular geometric shapes, including circles, ellipses, and lines. Thus, the simulated holes were distributed randomly over different spatial coordinates of the image space to occupy, on average per batch, 25 to 30% of the image size. With this approach, two models were trained separately for NSCLC and HN datasets.

Autoinpainting for unsupervised tumor segmentation. The trained Gconv_{Lap} model learns to synthesize semantically correct and contextually smooth contents in the predefined missing regions. The strategy to train the model only with healthy slices reinforces the model to replace the missing healthy tissues with the appearance of healthy tissues. This strategy enables the inpainting network to model the distribution of healthy anatomical structures that can be further utilized to detect the anomalies as outliers from the learned normative distribution. In other words, filling the tumor regions with the appearance of already learned

healthy tissues leads to synthesizing tumor-free images from which the tumoral regions can be detected by calculating the differences between the original and synthesized images. Hence, the learned inpainting network can function as a UAD model, given that no segmentation label is required to localize the tumor location. That being the case, a pipeline is proposed to turn the manual inpainting network into an autoinpainting model in order to segment the tumors in a purely unsupervised fashion.

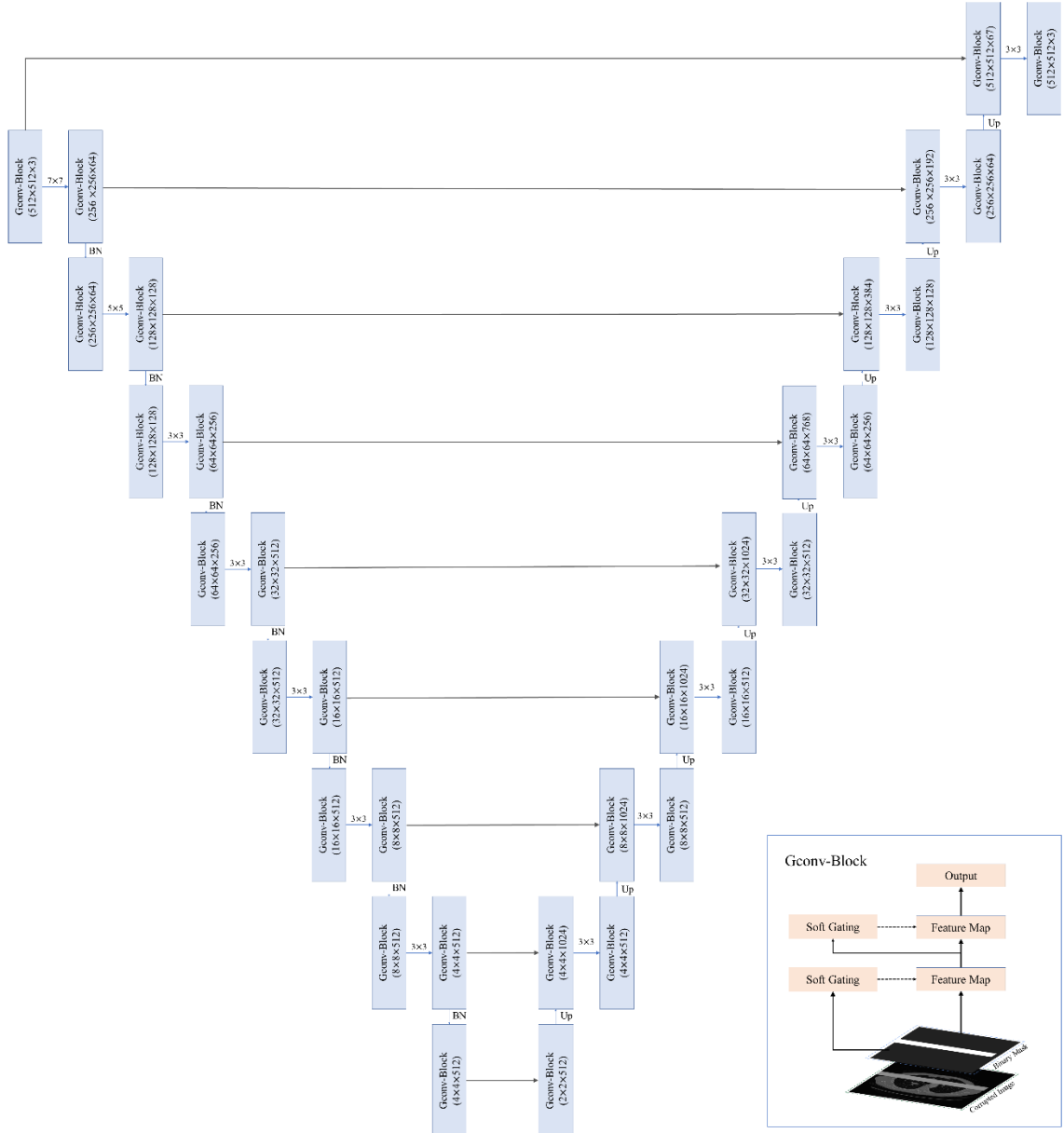


Figure 4.3. The architecture of the U-Net-like inpainting network is constructed based on the Gconv modules.

The underlying idea of unsupervised tumor segmentation here is to replace the random holes with a sliding window to sweep different anatomical regions for the inpainting process.

Therefore, if the sliding window covers healthy regions, the inpainting network will replace the appearance of healthy structures with already learned healthy structures; thus, the newly generated images remain intact. On the other hand, if the sliding window encounters tumoral regions, it can substitute the textures of the tumors with the appearance of learned healthy tissues. Accordingly, for each original tumoral slice, a fake tumor-free image can be generated without needing any kind of supervised signal. Hence, a pipeline is proposed to efficiently inpaint the tumoral regions while preserving the appearance of healthy tissues with anatomical constraints. This pipeline functions in the following four folds:

1. Preparing the input slices: extracting the axial slices from the organ of interest.
2. Detecting the candidate regions: identifying which moving windows cover the tumoral regions.
3. Determining the target region: estimating the location of the tumor(s) for removal.
4. Segmenting the target tumor: post-processing the residual volumes to segment the tumors.

To compare the efficacy of the proposed pipeline with other standard models, we evaluated the performance of two types of methods. First, the nnU-Net model [210] was employed as a robust supervised segmentation model to find out the optimal performance that can be achieved over the investigated tasks. Second, a set of recently developed deep UAD models was analyzed to objectively compare the segmentation accuracy of the proposed unsupervised model against the relevant UAD references. In this context, the following models were examined [240]: dense AE (dAE), spatial AE (sAE), constrained AE (cAE), VAE, context Variational AE (cVAE), Gaussian Mixture VAE (GMVAE), Fast-AnomalyGAN (F-AnoGAN), and Adversarial AE (AAE).

4.2 Imaging Biomarkers for Cancer Diagnosis and Prognosis (Papers III, IV, V, and VI)

The last four papers presented in this thesis focus on developing models for cancer diagnosis and prognosis. Papers III and IV concentrate on benign-malignancy prediction of early-stage tumors, while Papers V and VI aim to predict the survival status of cancer patients.

4.2.1 Overview of the Studies

Paper III aims to automatically distinguish the benign pulmonary nodules from the malignant ones by analyzing the Low Dose Computed Tomography (LDCT) scans of 968 subjects. Most lung cancers emerge from small-sized malignant lung nodules. However, tiny malignant nodules share highly similar visual characteristics with benign nodules. Therefore, distinguishing the malignant nodules from the benign ones is considered one of the most challenging tasks in lung cancer screening. In this study, a dual-pathway deep classification framework was developed to predict the binary class labels of the input nodule data.

Paper IV, as a continuation of Paper III, seeks to compare the performance of a series of carefully selected conventional radiomics models and end-to-end DL models for the task of benign-malignancy prediction of lung nodules in LDCT images. In addition, several fine-tuning and processing steps were integrated into the baseline methods, and the potential of combining radiomic features with deep features was also investigated.

In Paper V, longitudinal PET-CT scans of 30 patients diagnosed with NSCLC are analyzed to predict the survival status of the patients. For each subject, one PET-CT scan was acquired before the beginning of the treatment, and the second scan was acquired during the second week of radiotherapy. The objective of this study is to predict the overall survival status of the patients two years after the last session of radiation treatment by analyzing the longitudinal imaging data of the first two weeks of the treatment.

While Paper V focuses on a binary survival prediction task, Paper VI investigates Survival Rate Prediction (SRP)—that is, estimating the survival function in patients diagnosed with NSCLC and HN cancer by investigating the potential of different types of features. The studied features include conventional CNN-based, radiomics, as well as rotational invariant features learned by Sphercial CNN (SphCNN) models.

Figure 4.4 presents a schematic illustration of the general pipeline used in these studies.

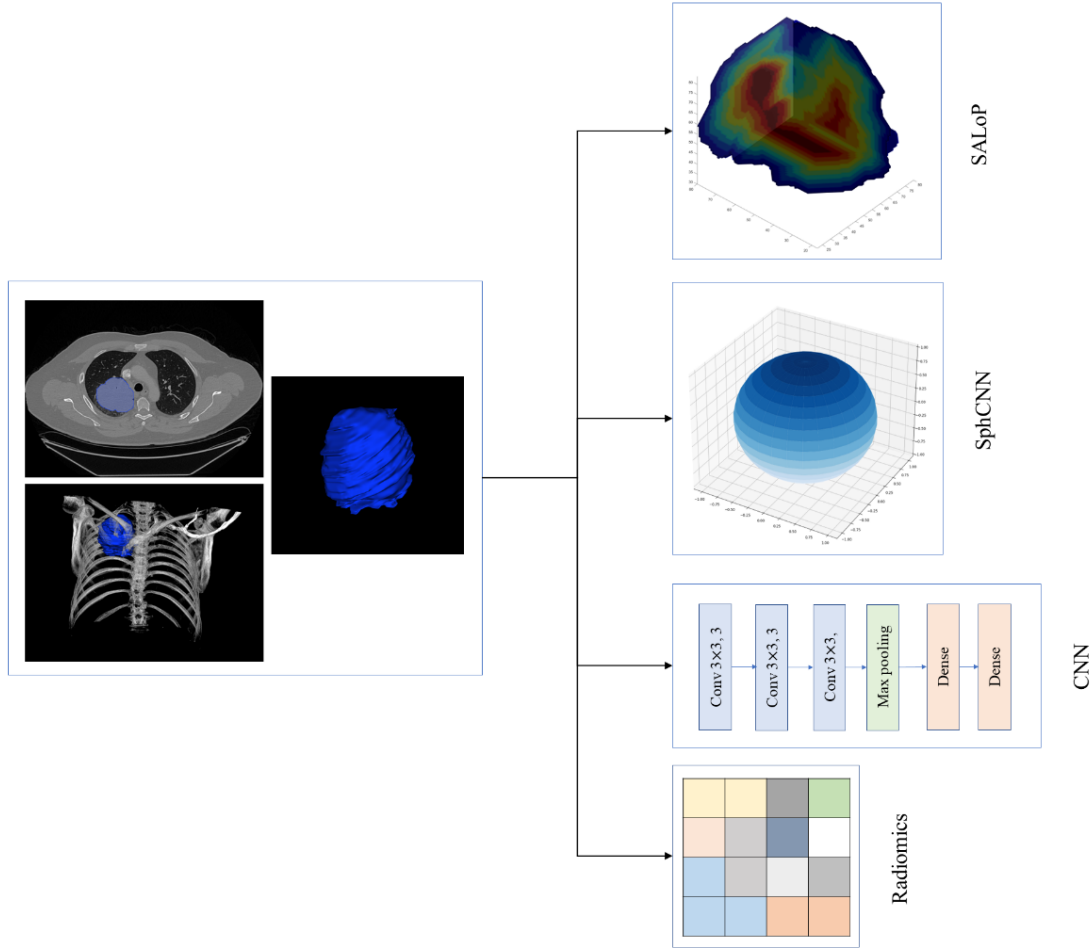


Figure 4.4. Schematic illustration of the methods used in the last four studies. The segmented tumors were analyzed with CNN models in Paper III, radiomics and hybrid pipelines in Paper IV, the SALoP feature set in Paper V, and the SphCNN model in Paper VI.

4.2.2 Image Preprocessing

To prepare the image data for feature extraction and analyses, a set of preprocessing steps were first applied to the intensity and geometry of the image volumes. These preprocessing techniques include intensity conversion and normalization, contrast enhancement, image registration, and voxel-space resampling.

Contrast enhancement. To enhance the appearance of lung textures in CT images, an intensity windowing was applied to the image histogram to clamp the voxel intensities in the range of $[-1000, 500]$.

Intensity normalization. In Papers III, IV, and VI, in which the image data were used to train deep classifier networks, the range of voxel intensities for each volume, after contrast enhancement, was rescaled to the range of $[0, 1]$.

Voxel-space Resampling. Considering that the voxel size of the volumes within each dataset varied significantly, the spatial resolution of the volumes within each dataset was resampled into a fixed size. Bicubic and nearest-neighbor interpolation functions were employed to resample the spatial resolutions of the image volume and segmentation masks, respectively.

Image registration. In Paper V, prior to any preprocessing steps, the PET volumes were resampled to the corresponding CT volumes in order to bring them into a fixed geometric space. Image registration was then applied to the longitudinal PET-CT scans to align the two volumes for the proposed feature extraction steps. The second CT scans were registered to the pretreatment CT volume in two folds. First, a rigid transformation was applied to the lung fields by manually defining anatomical landmarks on the corresponding volumes. This step was achieved by employing MeVisLab software [245]. Second, to further align the tumoral regions, an automatic rigid registration was applied only to the tumoral masks. The obtained deformation fields were then applied to the second PET volumes to wrap them into the pretreatment PET volumes as well.

Intensity conversion. For the employed PET dataset, the raw intensity values of the voxels were converted into SUV units by using the following equation [246]:

$$SUV = \frac{r}{(\alpha'/w)}$$

where r is the radioactivity concentration [kBq/ml] measured within a ROI, α' refers to the decay-corrected amount of injected radiopharmaceuticals [kBq], and w represents the patient's weight [g].

4.2.3 Tumor Segmentation

The main objective of Papers III, IV, and V is to develop models in order to quantify the tumor characteristics for further prognosis or prediction tasks. Accordingly, in these studies, the tumoral regions were segmented manually. In Papers III and IV a dataset including 968 LDCT volumes were visually examined by an expert radiologist from which a total number of 1297 pulmonary nodules were manually delineated. Since the identification of exact boundaries around the tiny pulmonary nodules is a challenging task even for radiologists, a region of interest was first cropped and zoomed around the candidate region, and then manual delineation was done over the upsampled cropped image. The final contours were

then downsampled to match the original image. In Paper V, the tumoral regions were delineated for each of the CT and PET images separately and for each of the pretreatment and second scans independently. A semi-automatic 3D level set-based segmentation tool was used to segment the target regions by selecting the initial seed over the candid region to instantiate the intensity-based surface evolution algorithm [247]. The delineations were done by an experienced user, and later an expert radiologist visually examined and refined the masks.

In contrast to Papers III, IV, and V, Paper VI aims to conduct a fully automatic analysis for the survival rate prediction task. Therefore, it was necessary to segment the tumors automatically. In this study, two types of tumors were examined: NSCLC and HN tumors. The NSCLC tumors were investigated from two datasets, named Lung1 [248] and Lung3 [249]. The segmentation of the lung tumors was done by using the NAA model. This model was developed as the main objective of Paper I. The NAA model was trained with a 5-fold cross-validation fashion over the Lung1 CT volumes. For each fold, the predicted masks of the test set were saved as the model output. The learned model was then used to infer the segmentation masks of the Lung3 dataset; however, the HN tumors were analyzed directly by using the original ground-truth masks. In fact, the segmentation of the HN tumors in CT images is known as an arduous task due to the lack of intensity/textural differences between the tumors and surrounding soft tissues. Since such poor segmentation results will negatively impact further analyses, the original label masks were used instead.

4.2.4 Feature Set Development/Extraction

Radiomic Features

In Papers IV, V, and VI, radiomic features were extracted and analyzed to benchmark the performance of the main developed model of each study. The employed protocols for extracting the radiomic features in these three studies are briefly described in this section.

Papers IV and VI. In these studies, the standard PyRadiomics package [250] was employed to extract the radiomic pools from the ROIs. In particular, in Paper IV, radiomic descriptors were not only extracted from the manually defined target regions, but the same features were also extracted from the image patches surrounding the tumors. Surrounding regions that best cover the context information around each tumor were cropped, and the contextual radiomics were characterized from the image patches. On the other hand, in Paper VI, the automatically generated segmentation masks for the NSCLC datasets and the

ground-truth masks of the HN dataset were used to define the target tumor area for feature extraction. FOS, SOS, and Morph-based features were extracted directly from the preprocessed volumes. Moreover, multiscale features were extracted from the transformed images filtered by Wavelet and Laplacian of Gaussian filters.

Paper V. In this study, a set of radiomic features were extracted from the pretreatment and second PET and CT scans separately. The radiomic pools include 13 First-Order Statistics (FOS), 18 Morphological (Morph) descriptors, 53 Second-Order Statistics (SOS), and 130 multi-scale features extracted from filtered images. In other words, 214 3D descriptors were used to translate the tumoral volumes into quantitative values. These features were extracted by employing a standard MATLAB-based package for radiomics analysis [251]. In addition to these features, three SUV-based features— SUV_{max} , SUV_{peak} , and SUV_{mean} —were added to the PET feature pools as well. Moreover, 20 other features representing the effective radiosensitivity [252], which account for both the changes in the PET uptake and the accumulated dose delivered by the time of the second PET image, were also included in the feature set. In summary, 451 quantitative features represent each tumor volume from the PET-CT scans. It should be noted that except for the Morph-based features, which were extracted from the independent tumoral regions of the pretreatment and second scans, other types of features were calculated from a union mask defined between the two longitudinal segmentation masks. RIDER dataset includes 32 patients diagnosed with NSCLC, each of whom underwent two chest CT scans within 15 minutes by using the same imaging protocols. Such a unique dataset is invaluable for investigating the reproducibility of the developed methods. Therefore, the mentioned radiomic features were extracted from the longitudinal RIDER tumor volumes as well. A Concordance Correlation Coefficient (CCC) metric was employed to quantify the reproducibility of the extracted features. Hence, the CCC value of 0.8 was set as a hard threshold by which only the CT radiomics with $CCC \geq 0.8$ were considered reproducible and preserved for further analysis steps. With the explained protocols, two feature sets were formed, known as F_{pre} and F_{sec} . Lastly, in order to capture the effect of applied treatment on the tumor attributes, delta features were calculated by finding the differences between the two feature sets, that is, $\Delta F = F_{sec} - F_{pre}$. All the feature sets, including pretreatment, second scan, and delta features, were analyzed accordingly.

Deep Features

Unlike study V, which was primarily focused on hand-crafted features, in Papers III, IV, and VI, DL-based features were investigated as well.

Paper III. In this study, DL-based classifiers were employed for the benign-malignancy prediction of pulmonary nodules in LDCT images. The term “lung nodules” refers to the moderately well-marginated round opacities with the largest diameter of less than 30 mm. Different subtypes of lung nodules—solid, part-solid, and ground-glass opacities—have varying probabilities of cancer [253]. Unfortunately, benign and malignant pulmonary nodules share highly similar visual characteristics, making their distinguishment a challenging task. In clinical practice, radiologists visually examine the chest LDCT scans slice by slice by following the guidelines such as LungRADs [254] to estimate the malignancy likelihood. Such guidelines are based mainly on the morphological characteristics of the nodules, such as size. In addition to the nodule size, it has been shown that the intensity distribution and relative position of the nodules are strongly associated with the malignancy [23]. Accordingly, in Paper III, to effectively capture both intra-nodule heterogeneities and contextual attributes, a dual-pathway CNN framework was proposed. The proposed approach aims to simultaneously learn the characteristics from inside the nodule regions as well as their relative position with respect to the surrounding structures. Therefore, both the target nodule regions and cropped patches around the nodules were fed into two parallel convolutional paths within one single network for simultaneous feature learning. The examined deep networks were used as either end-to-end classifiers or feature extractors trained with both supervised and unsupervised fashions. Figure 4.5 depicts a graphical illustration of the described framework.

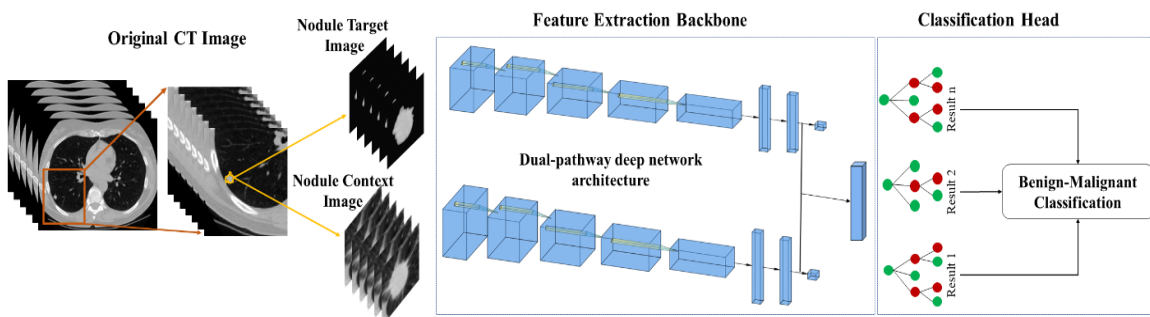


Figure 4.5. The graphical illustration of the dual-pathway model. The target nodules images and context nodule images were used together to feed two parallel paths of an end-to-end deep classifier. In addition, the learned deep features were augmented to balance the class labels for a separate final benign-malignancy classification with random forest. [From Astaraki et al. 2021, Reprinted with permission]

The supervised setup of the classification framework is independent of the network architectures; therefore, several different classification models were adopted. While the nodule target pathway is assumed to mainly learn the association between the intra-nodule characteristics and class labels, the nodule context pathway aims primarily to learn the correlations between the context information and class labels. Thus, by concatenating the learned features from each of the parallel pathways in a last shared layer, the model would learn to predict the class labels by adaptively learning the disentangled intra- and contextual features simultaneously. This setup was applied to the following models: VGG-like [190], ResNet [192], DenseNet [255], and EfficientNet [256]. Regardless of the backbone architectures, the last but one layer of each pathway was concatenated with each other to form the last shared final layer. The last layer of each pathway in the VGG-like model consists of a fully connected layer. Therefore, the final shared layer of the whole pipeline was formed by concatenating the last dense layers of each network. For the ResNet model, the output of the last residual modules at each pathway was concatenated before feeding the global average pooling layer to form the shared layer. For the case of the DenseNet model, the output of the convolutional filters of the last transition layers of the parallel networks was concatenated to form the shared feature layer. Finally, for the EfficientNet setting, the outputs of the last inverted residual blocks of the parallel networks were concatenated to form the shared feature set. It should be noted that the mentioned models were trained in an end-to-end fashion to predict the binary class labels of the input images. In addition to that, the outputs of the last shared layer, which consists of the learned features from the target and context pathways, were extracted for independent analyses with conventional ML algorithms.

As an unsupervised feature-learning approach, VAE was fitted to the dual pathway framework to learn the rich attributes of the input nodule data. The extracted features from the encoder part can provide high-level semantic attributes not only to capture the semantic nodule features such as lobulation and spiculation but also to learn the association between them.

To benchmark the efficacy of the proposed dual pathway models, conventional single pathway networks were examined as well. These single pathway networks were trained separately with target and context nodule images to predict the malignancy status either in an end-to-end fashion or by extracting deep features for further steps. It is worth mentioning that the extracted deep features of Paper III were directly used in Paper IV for comparing and combining radiomics features.

Paper VI. In this study, both SphCNNs and conventional CNNs were examined to perform deep-feature extractions for the task of SRP. While conventional CNN models function on the Cartesian domain, SphCNNs are designed to work on signals defined on the sphere domain. The architecture of SphCNNs consists of a series of spherical filters that are applied to the spherical activation signals through the spherical convolution. This convolutional operation is often carried out as a multiplication in the spherical harmonics domain. The ordinary CNNs are naturally translational invariants, which means they can learn the characteristics of the image contents regardless of their positional shift in the image domain. On the other hand, SphCNNs can be used for solving problems where rotational equivariance (i.e., the output rotates when the input is rotated) or rotational invariance (i.e., the output is always the same even if the input is rotated) is required [257]. In the context of the survival analysis task, we can assume that SRP can be a rotational invariant problem because the patient’s survival might be the same regardless of the orientation of the tumor in the lung regions. Therefore, both SphCNNs and ordinary CNNs were investigated to extract two different sets of features.

Unlike the CNNs that are often fed with the image slices/volumes, image data should be first mapped into the unit sphere S^2 before being fed to the SphCNNs. To map the target volume regions, three types of projections were examined: (a) Depth-based Projection (DbP) of the segmentation mask; (b) Extended Gaussian Image (EGI) of the tumor mask, which is the orientation distribution function of the normal vectors from the tumor surface, and (c) Intensity-based Projection (IbP) of the tumor. The DbP was calculated in two steps. An enclosing sphere was first centered on the tumor’s center of mass; then, a ray was cast from each sampling point onto the surface of the enclosing sphere to the centroid. The distance of the first intersectional point defined the value of the spherical signal on that specific orientation. The EGI mappings were directly produced from the normal vectors generated from the segmentation masks. The IbP was computed by accumulating the voxel intensity values of the tumoral regions along with every ray. Figure 4.6 shows a graphical depiction of the three explained spherical projections. In order to find out whether segmentation masks alone carry predictive power or whether the addition of the image contents is beneficial, two sets of the network were configured. In particular, SphCNN-1 was fed only with the DbP projection, while SphCNN-2 was fed with both EGI and IbP projections. The architectures of both configurations were similar and consisted of a convolutional layer to lift the input signal from the spherical domain into a manifold, followed by another convolutional layer and a final dense layer to encode 40 features. It should be noted that the network architecture was determined experimentally.

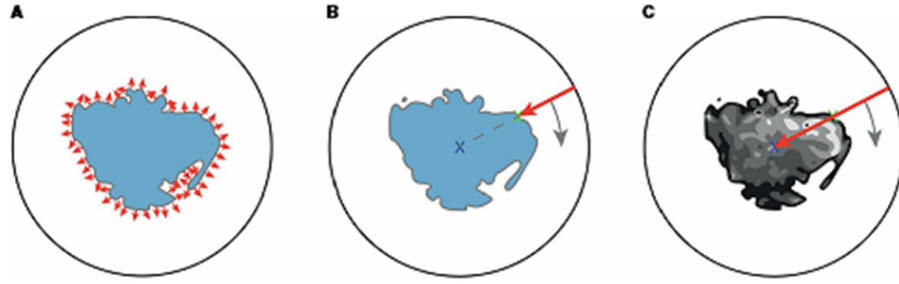


Figure 4.6. Illustrative of the three employed spherical mappings. Note that the tumor volumes are depicted as 2D image slices; therefore, the spheres are visualized as circles. (A) The extended Gaussian image (EGI) can be viewed as an accumulation of the gradient vectors (small red arrows) at the surface of the tumoral boundary. (B) Depth-based projection of the solid segmentation mask. A ray (red arrow) is cast from a projecting sphere to the surface of the segmentation. The distance from the sphere to the surface determines the value of the spherical signal at the respective position. (C) Intensity-based projection of the voxel image content. A ray (red arrow) is cast from the surrounding sphere through the segmented tumor image toward the centroid. The value of the spherical signal is the sum of all intensities of the voxels that the ray traversed. [Caption and figure From Sinzinger et al. 2022. Reprinted with permission.]

As a baseline CNN-based method in Paper VI, a pre-trained ResNet50 model was employed as a transfer learning approach for feature extraction. From each subject, the axial slice in which the tumor appears with the largest size was selected to be fed to the pre-trained network. This results in extracting 1,000 deep features for each tumor.

Size-Aware Longitudinal Pattern (SALoP)

It is a well-established idea that cancer is a dynamic disease. During the course of the disease, cancer cells generally become more heterogeneous and form a heterogeneous bulk tumor that might consist of a diverse collection of cells with a wide range of characteristics. Thus, it would be possible to explain the partial tumor response to the treatment based on the existence of subregions within the tumor volume with different sensitivity to the treatment. Therefore, SALoP was designed in an attempt to address the challenge of quantifying the intra-tumor heterogeneity and its impact on the treatment response. We hypothesized that the tumor microenvironment leading to the formation of sub-volumes with different vasculature and interstitial fluid pressure conditions could be characterized by dividing the tumor volume into a given number of separate zones based on the distance from the tumor boundaries. In other words, it was hypothesized that the sensitivity of tumor cells to treatment varies as a function of the distance from the tumor borders; therefore, by partitioning the tumor volume into separate subregions, metabolic activity and characteristics of tumor cells might be captured from more homogeneous regions.

To implement this idea, the Euclidean distance of each voxel within the Gross Target Volume (GTV) from the tumor border was measured. Then, for every 5 mm of distance from the tumor border, one subregion was added. Thus, the smallest studied tumor with a radius of 5 mm was constructed as only one subregion, while the largest studied tumor with a radius of 50 mm consisted of 10 subregions counting outward, with the tumor core being the most inner layer and tumor borders as the outermost layers. With this strategy, all the studied tumors were divided into 1 to 10 concentric zones (see Figure 4.7). From each subregion, average intensity values were then calculated as a quantitative feature. In other words, each tumor for each of the PET and CT images was presented with 1 to 10 quantitative values. Finally, by subtracting the average intensity within one zone of the second scan from the corresponding zone of the first scan, the delta SALoP feature was formed for each of the PET and CT volumes separately. This delta feature can potentially characterize the changes in the tumor characteristics as a result of the applied treatment. Technically speaking, while the delta PET features can capture the changes in the level of metabolic activities, the delta CT features quantify the changes in the densities of the tissues, both from more homogeneous subregions. Accordingly, one of the advantages of the proposed SALoP feature set is the fact that it can be interpreted physiologically.

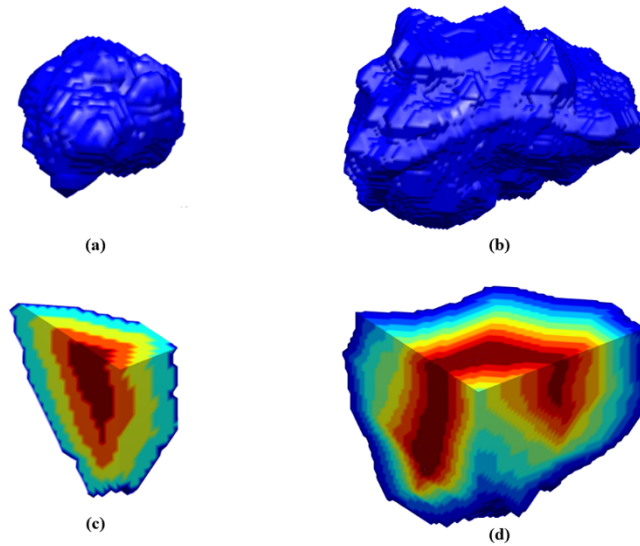


Figure 4.7. Partitioning the tumors into concentric zones based on the distance to the tumor borders. The first row shows 3D surface plots of a middle-sized tumor (a) and the largest studied tumor (b). The second row depicts the cross-sectional iso-surface of the tumor volumes partitioned into 5 and 10 concentric zones, respectively. [From Astaraki et al. 2019, Reprinted with permission]

4.2.5 Benign-Malignancy Prediction

Paper III. In this study, the developed dual pathway models were used for classifying the malignant nodules from the benign ones. This classification was done in two ways—through end-to-end learning and deep feature extraction for further analysis. In the end-to-end prediction approach, all the four employed network architectures were trained with conventional binary cross-entropy as the objective function in a 5-fold cross-validation fashion. In addition to the dual pathway models, ordinary single pathway networks were trained with the same setting for objective comparisons. In other words, 1,297 target nodule images along with 1,297 context nodule patches were used either simultaneously or independently to train the networks. To reduce the risk of overfitting, conventional image augmentation techniques such as random flipping along one axis and affine transformations were applied to the image volumes.

As for the deep feature extraction approach, the outputs of the last but one layer of each pathway of the already trained networks were extracted. More specifically, the target and context nodule features were extracted either as a concatenated feature vector from the dual pathway networks or independently from the ordinary single pathway models. Feature engineering steps were then applied to the extracted deep features for the final class label prediction. From the 1,297 pulmonary nodules studied, 867 cases belong to the benign group, and the remaining 421 nodules were labeled as malignant. This unequal distribution of the class labels leads to a high bias toward the majority class, which in turn degrades the model performance that often leads to poor predictive power over the minority class. To avoid such negative impacts, the Synthetic Minority Oversampling Technique (SMOTE) [258] was employed to balance the class distribution by generating synthetic samples from the minority class. Technically speaking, SMOTE fits a hypercube among instances of the minority class to generate new samples along the fitted hypercube. As a result of this augmentation technique, 455 new samples were synthesized belonging to the malignant class. The RF model was experimentally chosen as the learning algorithm. Therefore, RF was trained with a 5-fold cross-validation over the balanced dataset. In addition, to quantify the effect of the feature augmentation step, the RF model was trained with the same setting but over the unbalanced feature sets. Different experiments were conducted to compare the performance of the context nodule images, target nodule images, and their combinations for malignancy prediction.

Paper IV. Both DL and radiomics methods have shown comparable performance in identifying pulmonary nodule malignancy in CT volumes. However, there is a lack of

conclusive evidence to identify one of these methods as more accurate than the other. Therefore, Paper IV aims to present a comparison between the performance of a series of carefully selected conventional radiomics, DL models, and their combinations for the task of pulmonary nodule malignancy prediction. More specifically, the predictive power of three independent modules was examined on the same dataset as Paper III. These modules include radiomics module, end-to-end DL module, and hybrid module, which were built upon the combination of the first two modules. It should be noted that the end-to-end module was directly adopted from Paper III; therefore, deep features were extracted from the trained models of Paper III as well. In the hybrid modules, feature sets were built by combining the radiomics features with the extracted deep features. In each module, the performance of the target nodule images and context nodule images were investigated separately. In addition, to examine how context and target nodule images would complement each other, the predictive power of their combined feature sets was assessed as well.

The predictive power of the radiomics features was evaluated by integrating seven feature selection/dimensionality reduction methods into eight learning algorithms. To reduce the effect of irrelevant or partially relevant features, five filter-based Feature Selection (FS) and a wrapper FS method were employed. Filter-based FS algorithms include the following: (a) Constant: to remove constant features, (b) Correlation: to remove linearly associated features, (c) Mutual Information: to remove nonlinearly associated features, (d) RELIEF: to estimate the quality of the features based on how well the features can classify the subjects that are close to each other; and (e) LASSO: to apply coefficients to the features and shrink to zero the coefficients whose features are less predictive [259]. On the other hand, as a wrapper method, FFS was used to examine the performance of the learning algorithms with different combinations of feature subsets. Finally, Principal Component Analysis (PCA) was employed as well to project the high-dimensional feature space onto only the first few principal components [260]. The predictive power of the original feature sets as well as selected features—after being augmented by the SMOTE algorithm—were evaluated by the following learning algorithms: Adaptive Boosting (Adab), DT, RF, K-Nearest Neighbor (KNN), SVM, Linear/Quadratic Discriminant Analysis (LDA/QDA), as well as Naïve Bayesian. It is worth noting that all the analyses were executed with a 5-fold cross-validation resampling technique.

From the 64 different settings executed over the radiomics module, it was understood that integrating the FFS method into the Adab learning algorithm leads to the highest

predictive power. Accordingly, feature sets within the hybrid module were examined only with the FFS-Adab model. Extensive experiments were executed to objectively compare the performance of different modules with different fine-tuning settings, such as effects of feature selection, effects of class balancing, effects of learning algorithms, effects of target/context images, and effects of the complementary roles of different feature types. Figure 4.8 presents a graphical depiction of the studied pipeline.

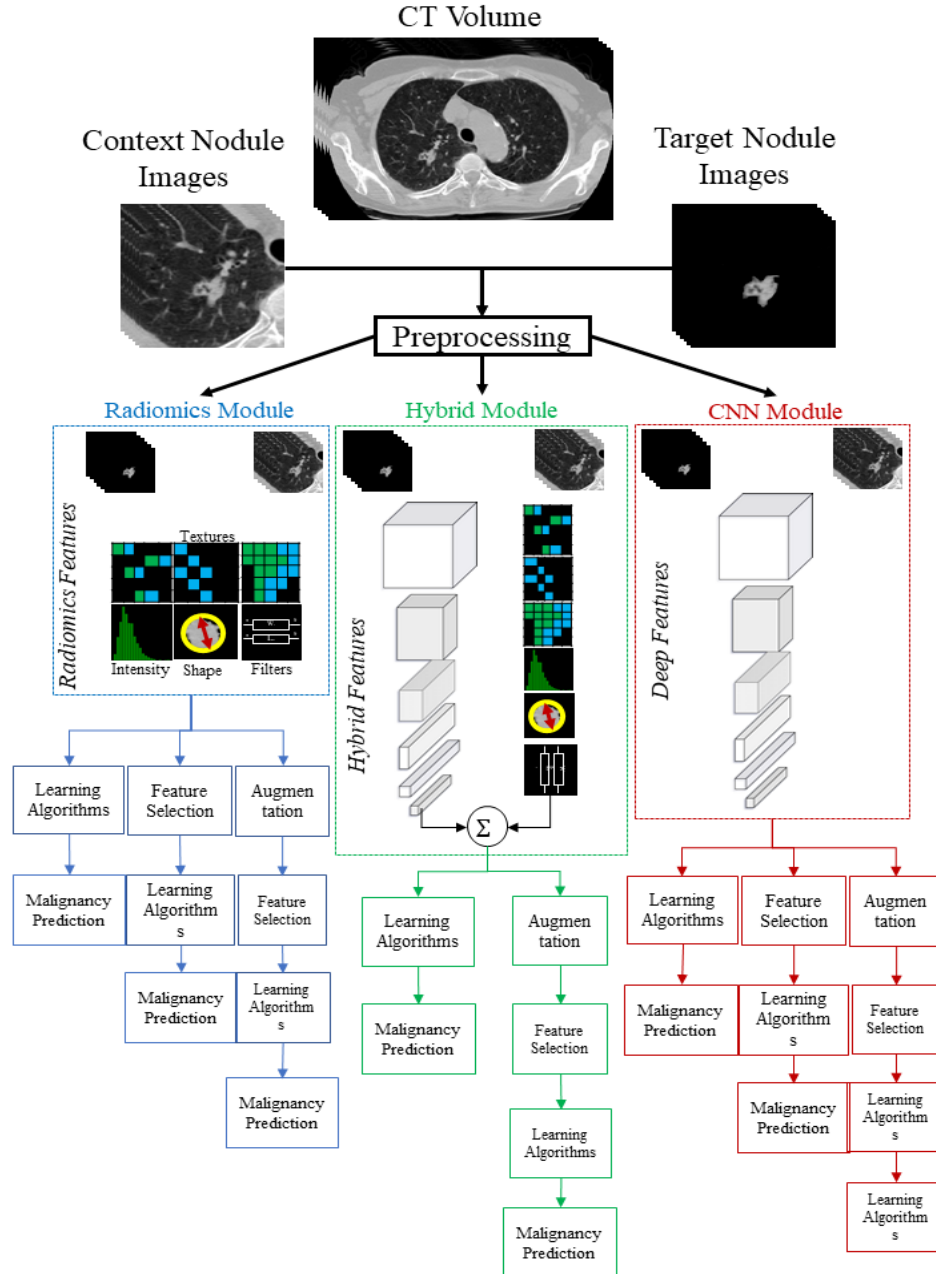


Figure 4.8. Graphical depiction of the pipeline of Paper IV. The performance of the radiomics, deep learning, and hybrid modules was examined with target nodules, context nodules, and their combinations. [From Astaraki et al. 2021. Reprinted with permission.]

4.2.6 Survival Prediction

Paper V. In this project, the patients were followed up for two years after the last session of radiation treatment. If the patients remained alive by the time of the last follow-up session, their study endpoint was considered positive; otherwise, their survival status was set as negative. Accordingly, the Overall Survival Status (OST) was defined as a binary classification problem. Therefore, in Paper V, the OST of the NSCLC patients was predicted by using the developed SALoP feature set, the described radiomics descriptors, and their combination. Differences between the mentioned features from the pretreatment and second scans, delta features, were examined as well. In particular, the prognostic power of three types of feature sets was assessed by training a SVM as the learning algorithm. However, due to the small number of subjects and large number of extracted features, the dimensionality of the feature sets was reduced by applying a FFS algorithm to preserve the most informative features with respect to the class labels. More importantly, to minimize the risk of overfitting, the learning process was evaluated through 10-fold cross-validation and Leave One Out (LOO) resampling techniques. Several experiments were executed to compare the predictive performance of the studied feature sets.

Paper VI. In this study, the specific objective of survival analysis is the task of SRP, which refers to estimating the time until a terminal event that could occur in some cases but not necessarily in all subjects. Therefore, unlike in Paper V, which focuses on a binary classification problem, SRP can be framed as a regression-type problem. However, a conventional difficulty in this task is the presence of censored data, which refers to those subjects for whom survival times are not available. This issue can occur if (a) subjects do not experience the event when the study is over, (b) subjects were lost to follow-up during the study period, or (c) subjects were withdrawn from the study. The most common type of censored data is known as right-censored data, which means the starting point of the observations is known for all the cases, but the definitive endpoint of the observations might be missed for some subjects. In this study, the SRP task was conducted by SVM with Ranking (SVM-K), SVM Regression (SVM-R), Cox's Proportional Hazard (CoxPH), and Ensemble Gradient Boosting (EGB) models. These models were trained with four different types of features, including radiomic features, CNN-based features, and the SphCNN-based features extracted from two different settings, SphCNN-1 and SphCNN-2. One thousand five hundred extracted radiomic features and 1,000 CNN-based features are much larger than 40 SphCNN-based features. Therefore, by ranking each regressor based on its cross-

correlation value with the target, a subset of 32 features was selected for radiomics and CNN-based features prior to model training.

Chapter 5

Results

5.1 Deep Learning Methods for Tumor Segmentation

5.1.1 Paper I

Prior-aware Autoencoders for Lung Pathology Segmentation

In this study, three separate models have been sequentially analyzed to segment four different types of lung pathologies. Therefore, the performance of the proposed pipeline can be quantified in three folds: pathology inpainting, automatic pathology segmentation, and prior-aware pathology segmentation. Each of these are discussed in the following paragraphs.

Pathology inpainting. Figure 1a in Supplementary Materials of the original paper shows how the inpainting network could successfully synthesize realistic-looking images by filling the random holes with semantic lung textures. In fact, the size of the random holes was intentionally varied from small regions to extremely large regions to enhance the ability of the model to deal with pathologies of different sizes that appear at various locations. As can be seen from the figure, the model seamlessly inpainted the images regardless of the size of the random holes. In other words, the learned model could inpaint different pathologies from peripheral tiny pulmonary nodules to a large solid NSCLC tumor that was connected to the healthy soft tissues. More importantly, the employed Pconv network inpainted only the masked regions and did not have any impact on other parts of the images. As a result, in the test phase, the model inpainted only the pathological regions without any effect on the structures of the other anatomical regions.

Automatic pathology segmentation. The performance of the proposed NAA model to automatically remove the pathologies from input slices was evaluated by quantifying the ability of this model for indirect segmentation of pathological regions. In fact, a set of post-processing steps was applied to the residual volumes, which were calculated as differences between the original slices and outcomes of the NAA model. Table 2 in the original paper

presents the segmentation accuracy in terms of numerical metrics. The segmentation accuracy in terms of Dice metric for challenging lung nodules, NSCLC-Decathlon, NSCLC-Radiomic, Covid19-Coronacases, and Covid19-Challenge datasets are 0.26, 0.52, 0.54, 0.72, and 0.64, respectively. As was expected, applying the non-learnable post-processing steps such as hard thresholding is not effective in preserving all the pathological regions in the residual volumes; thus, the presence of many false positive and false negative pixels adversely impacts the segmentation accuracies. For instance, the intensity and textural patterns of Covid pathologies are more homogeneously distributed in different subjects compared to lung nodules; therefore, higher values of the Dice coefficient were obtained for Covid cases. In addition, the size of pulmonary nodules is often much smaller than those of the NSCLC and Covid pathologies, so only a few misalignment pixels are enough to drop the accuracy.

Prior-aware pathology segmentation. The residual images obtained from the NAA model, without any post-processing, were integrated into a U-Net segmentation model as a prior channel to guide the learning procedure of the network. To benchmark the efficacy of the proposed idea of integrating the prior image into the segmentation network, the self-configuring nnU-Net model was tested as a baseline and a prior-aware model as well. Figure 5.1 shows quantitative comparisons between the segmentation performance of the baseline model and prior-aware model for both our implementation and the standard nnU-Net model. By comparing the results of the baseline models and prior models, it can be observed that integrating the prior images into the networks could successfully improve the segmentation accuracy for all the employed datasets. The Dice values of the prior-aware models outperformed those of the baseline networks by a relatively large margin. For example, the Dice score of lung nodule segmentation was improved from 0.712 to 0.750 with our U-Net model. For the same dataset, the Dice score of the nnU-Net model was improved from 0.779 as the baseline model to 0.795 as the prior-aware model. In addition, the results of the nnU-Net model, both as a baseline and as a prior-aware model, outperformed our implementation of the U-Net model. Finally, performing the Wilcoxon signed rank test between the Dice scores of the baseline and prior-aware models indicates significant differences for all the datasets, excluding the CoronaCases dataset.

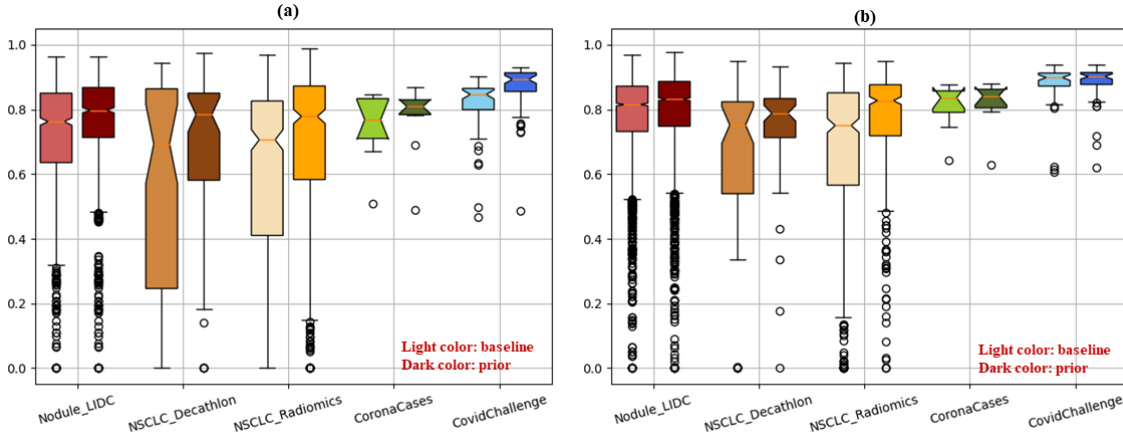


Figure 5.1. Comparing the segmentation performance of the baseline models and prior-aware models in (a) our implementation and (b) nnU-Net model. [From Astaraki et al. 2022. Reprinted with permission.]

5.1.2 Paper II

Unsupervised Tumor Segmentation

The performance of the proposed autoinpainting pipeline for tumor segmentation can be evaluated quantitatively in three folds: inpainting quality, autoinpainting for tumor segmentation, and external validation.

Inpainting quality. The performance of the proposed $Gconv_{Lap}$ model to inpaint the images is quantified by measuring three quantitative metrics, including MSE, PSNR, and SSIM. Furthermore, to benchmark the efficacy of the proposed $Gconv_{Lap}$ model, Pconv and Gconv models were examined on the same dataset as well. Tables 1 and 2 in the original paper indicate the numerical comparisons between the three models for NSCLC and HN datasets, respectively. From these comparisons, it can be observed that the proposed $Gconv_{Lap}$ model could inpaint the corrupted images more accurately than the other two models for both datasets, regardless of the image modalities (see Figure 5.2). In particular, the numerical metrics obtained from the proposed $Gconv_{Lap}$ indicate fewer reconstruction errors in terms of MSE metric and higher similarity in terms of PSNR and SSIM for all the experiments. Since noisy PET images do not carry anatomical details and consist of substantially fewer textures than CT images, quantitative values of the PET modality denote higher image similarities. In addition to evaluating the performance of the three models with multimodal images, the same models were trained and tested using ordinary single modality images. Nevertheless, the superiority of the proposed $Gconv_{Lap}$ model was not affected by the type of input images. Moreover, the impact of leveraging the multi-term objective function with the Laplacian (lap) pyramid loss function is qualitatively visualized in Figure

5 in the original paper. While the proposed $Gconv_{Lap}$ model could preserve the edges between the soft and hard tissues, the ordinary $Gconv$ model could hardly keep the edges, which leads to losing the edge information in the inpainted images, thereby increasing the reconstruction errors.

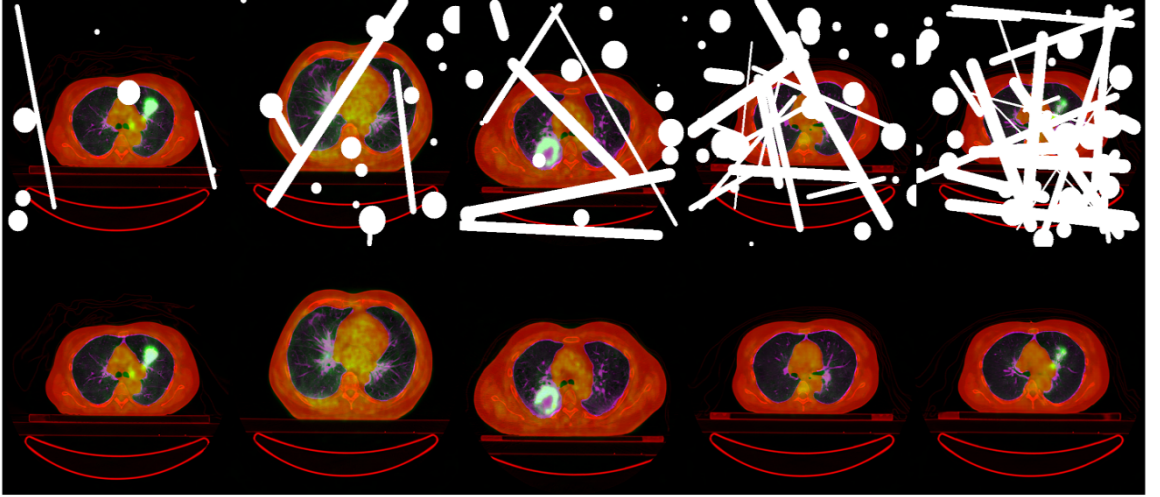


Figure 5.2. The inpainting model could seamlessly inpaint the corrupted images. The first row shows the multimodal slices corrupted with random irregular holes, and the second row indicates the inpainted slices by the proposed $Gconv_{Lap}$ model.

Autoinpainting for tumor segmentation. The efficacy of the proposed autoinpainting pipeline to segment the tumors is quantified by measuring the agreement between the segmented volumes and the label masks. The same autoinpainting pipeline was applied to all three inpainting models, and their segmentation performances were compared. Tables 5.1 and 5.2 present the segmentation accuracy of the proposed autoinpainting pipeline for NSCLC and HN tumors, respectively.

Table 5.1 – Numerical results of NSCLC tumor segmentation with autoinpainting pipeline

| Model-Data | Quantitative Metrics ($\mu \pm \sigma$) | | | <i>Dice</i> |
|----------------------|---|----------------------|-------------------|-------------------|
| | [<i>Dice</i>] | [<i>Precision</i>] | [<i>Recall</i>] | |
| Pconv-CT | 0.382 ± 0.157 | 0.408 ± 0.186 | 0.389 ± 0.151 | 0.353 ± 0.111 |
| Gconv-CT | 0.423 ± 0.180 | 0.463 ± 0.199 | 0.411 ± 0.178 | 0.398 ± 0.124 |
| $Gconv_{Lap}$ -CT | 0.442 ± 0.176 | 0.482 ± 0.192 | 0.426 ± 0.176 | 0.410 ± 0.134 |
| Pconv-PET | 0.709 ± 0.215 | 0.793 ± 0.196 | 0.669 ± 0.221 | 0.654 ± 0.132 |
| Gconv-PET | 0.750 ± 0.176 | 0.792 ± 0.192 | 0.747 ± 0.189 | 0.690 ± 0.184 |
| $Gconv_{Lap}$ -PET | 0.746 ± 0.196 | 0.822 ± 0.169 | 0.706 ± 0.217 | 0.686 ± 0.121 |
| Pconv-Multi | 0.673 ± 0.245 | 0.771 ± 0.219 | 0.622 ± 0.252 | 0.625 ± 0.122 |
| Gconv-Multi | 0.747 ± 0.172 | 0.799 ± 0.178 | 0.718 ± 0.183 | 0.692 ± 0.136 |
| $Gconv_{Lap}$ -Multi | 0.766 ± 0.171 | 0.832 ± 0.158 | 0.726 ± 0.184 | 0.708 ± 0.118 |

Table 5.1 indicates that the segmentation accuracy achieved by the proposed Gconv_{Lap} model is higher than that of the PConv model, regardless of the employed imaging modality. The same trend can be seen when comparing the Gconv_{Lap} model with the ordinary Gconv model for the CT and multimodal images though the Gconv model slightly performs better on the PET images.

Table 5.2 – Numerical results of HN tumor segmentation with autoinpainting pipeline

| Model-Data | Quantitative Metrics ($\mu \pm \sigma$) | | | <i>Dice</i> |
|---|---|---|---|---|
| | [<i>Dice</i>] | [<i>Precision</i>] | [<i>Recall</i>] | |
| Pconv-CT Gconv-CT Gconv _{Lap} -CT | NA | NA | NA | NA |
| Pconv-PET Gconv-PET Gconv _{Lap} -PET | 0.412 \pm 0.190 0.445 \pm 0.188 0.453 \pm 0.196 | 0.541 \pm 0.230 0.557 \pm 0.407 0.550 \pm 0.236 | 0.462 \pm 0.172 0.408 \pm 0.181 0.413 \pm 0.181 | 0.389 \pm 0.132 0.407 \pm 0.130 0.405 \pm 0.130 |
| Pconv-Multi Gconv-Multi Gconv _{Lap} -Multi | 0.408 \pm 0.241 0.462 \pm 0.202 0.464 \pm 0.198 | 0.511 \pm 0.274 0.539 \pm 0.233 0.541 \pm 0.233 | 0.360 \pm 0.224 0.443 \pm 0.189 0.435 \pm 0.188 | 0.344 \pm 0.100 0.418 \pm 0.133 0.422 \pm 0.135 |

In the HN experiments, the segmentation accuracy of the proposed Gconv_{Lap} model outperformed the Pconv model with a relatively large margin and performed slightly better than the ordinary Gconv model on the PET and multimodal images. On the other hand, the characteristics of the HN tumors are similar to those of the nearby soft tissues. Hence, the appearance of tumors in CT images cannot be distinguished from the surrounding structures due to the lack of intensity and textural contrast. Accordingly, the inpainting methods are not able to detect the HN tumoral tissues in CT images.

External validation. Tables 5, 6, and 7 in the original paper show the segmentation accuracy achieved by other competing methods. In particular, the performance of the supervised nnU-Net model for segmenting NSCLC and HN tumors was examined in the numerical results reported in Table 5. As was expected, the segmentation accuracy of the supervised models outperformed the unsupervised pipeline. Nonetheless, after carefully comparing the results, one can observe that the performance of the unsupervised autoinpainting models is relatively close to that of the supervised nnU-Net models, especially in the cases of multimodal and PET images. For example, the calculated Dice metric from the proposed Gconv_{Lap} model for NSCLC tumors in multimodal images is 0.708, which is around 0.1 lower than that of the nnU-Net model ($Dice = 0.802$). For the case of HN tumors, the supervised nnU-Net model outperformed the unsupervised approach by a substantial margin ($Dice_{Gconv-lap_{multi}} = 0.422$ vs. $Dice_{nnU-Net_{multi}} = 0.660$).

Finally, to examine the performance of the proposed autoinpainting model against the unsupervised methods, several UAD models were examined on the same NSCLC and HN datasets. Tables 6 and 7 in the original manuscript show the segmentation accuracy of the models on the multimodal images.

5.2 Imaging Biomarkers for Cancer Diagnosis and Prognosis

5.2.1 Paper III

Benign-malignant pulmonary nodule classification in low-dose CT with convolutional features

In this work, two main approaches were implemented to investigate their capabilities in the benign-malignancy classification of pulmonary nodules. These two strategies are (a) the end-to-end training of the deep networks (baseline) and (b) deep feature extractions followed by fine-tuning steps (feature engineering steps). Table 5.3, which summarizes Tables 1 and 2 of the original paper, presents the obtained discrimination power by the two mentioned approaches. Accordingly, the main findings of this study can be reported as follows.

Baseline models vs. augmented deep features. Numerical values show that the augmented deep features outperformed the end-to-end baseline networks by rather large margins regardless of the network architecture or the type of the image data. In fact, the best discrimination power observed by the baseline models is related to the dual pathway DenseNet models, which achieved an AUROC score of 0.824. On the other hand, the worst classification power of the augmented deep features was observed in the VAE-based features extracted from target nodule images (AUROC=0.851). In other words, the unsupervised VAE model, when followed by fine-tuning steps, could achieve higher discriminator power than end-to-end supervised models. In addition, it can be observed that the best classification power achieved by augmented features is almost 0.11 higher than that of the baseline models (0.936 vs. 0.824).

Table 5.3 Discrimination power of baseline deep learning models and augmented deep features after applying 5-fold cross-validation in terms of AUROC. [From Astaraki et al. 2021. Reprinted with permission.]

| Model | Baseline | | | Augmented Deep Features | | |
|--------------|---------------|----------------|----------------------|-------------------------|----------------|----------------------|
| | <i>Target</i> | <i>Context</i> | <i>Dual pathways</i> | <i>Target</i> | <i>Context</i> | <i>Dual pathways</i> |
| VAE | NA | NA | NA | 0.851 | 0.868 | 0.855 |
| VGG | 0.801 | 0.795 | 0.821 | 0.898 | 0.917 | 0.920 |
| ResNet | 0.785 | 0.763 | 0.794 | 0.902 | 0.903 | 0.909 |
| DenseNet | 0.792 | 0.806 | 0.824 | 0.906 | 0.924 | 0.936 |
| EfficientNet | 0.783 | 0.798 | 0.808 | 0.905 | 0.927 | 0.931 |

Dual pathway models vs. conventional single pathways. Experiments show that augmented deep features extracted from context nodule images carry more discrimination power than target nodule images regardless of the network architecture. The same trend can be observed for DenseNet and EfficientNet end-to-end models, in which the classification powers of the context images are slightly higher than target images. It is even more interesting that the combination of target and context information through the dual pathway models leads to improving the classification performance in all the supervised models, whether they were trained with end-to-end fashion or whether fine-tuning steps were applied to the extracted features. The strategy of the dual pathway technique, however, was not beneficial for the unsupervised VAE model. In this domain, another set of experiments was conducted by using only regions around the nodules, that is, by masking out the nodule region from the context images (Table 2 in Supplementary Materials). However, such images exclude the important characteristics of the nodules and therefore carry less discriminative power (0.768 for baseline and 0.879 for augmented features).

Effect of feature augmentation. Table 1 in Supplementary Materials numerically presents the effects of the feature augmentation step on the learned deep features. In fact, the outstanding performance achieved by augmented deep features is due to balancing the class labels in the feature space. By comparing the performance of the end-to-end VGG model against the extracted VGG features trained with an RF model without applying the augmentation step, one can observe that the end-to-end baseline model performs much better than the other method (0.821 vs. 0.773). More importantly, to ensure the RF models were not overfitted by the augmented features, the synthesized augmented features were used only as the training set, while the real extracted deep features were used as the test sets. Comparing the results of the second column in Table 1 in Supplementary Materials to those of Table 5.3 (VGG row after augmentation) can increase our confidence in this hypothesis.

Effect of feature fractioning. To examine the impact of the size of the feature pools on the performance of the RF model, different experiments were conducted by assigning

different fractions of the whole feature set to the training split. In particular, experiments were performed on 25, 50, and 70 percent of the extracted features as the training sets, and the predictive power was measured over the remaining test sets (see Table 3 in the paper). Interestingly, training the RF model with even 25% of the augmented features resulted in an AUROC score of 0.869, which is remarkably higher than the best accuracy achieved by the end-to-end baseline models.

5.2.2 Paper IV

A Comparative Study of Radiomics and Deep-Learning Based Methods for Pulmonary Nodule Malignancy Prediction in Low Dose CT Images

In this paper, the performance of radiomic features for lung nodule malignancy prediction is presented and compared against the predictive power of end-to-end DL models as well as deep features, which were introduced in Paper III. In addition, the hybrid feature sets were formed by combining radiomics and deep features to investigate their potential for malignancy prediction as well.

Handcrafted radiomics. In this approach, 1,334 radiomic descriptors were extracted from target nodule and context nodule images separately. These feature sets were analyzed with and without feature selection methods over eight different learning algorithms. A joint feature set was formed by combining extracted target and context features and was analyzed with the same approach. All the analyses were performed on the original imbalanced features as well as augmented balanced feature sets. Numerical values in terms of AUROC show that synthesizing new samples in the feature space improved the classification power of radiomic features by up to 0.09 in terms of AUROC. Tables 1, 2, 3 in the original paper and Tables 2, 3, and 4 in the Supplementary Materials present detailed quantitative comparisons for target, context, and joint feature sets. For instance, the predictive power of the target nodule features trained with the Adaboost learning algorithm without any feature selection before and after feature augmentation are $AUROC_{unbalanced} = 0.77$ and $AUROC_{balanced} = 0.88$. Comparing the discrimination power of radiomics extracted from target and context nodule images does not lead to drawing a solid conclusion in favor of one of them as their performances are relatively similar. However, integrating the target and context radiomic features together improved the accuracy and resulted in the highest discrimination power of the radiomics family. Comparing the performance of the models after applying the feature selection methods, it can be inferred that the FFS method consistently improved the classification accuracy. Accordingly, the highest classification

accuracy was achieved by integrating the FFS method into the Adab learning algorithm. With this setting, the predictive powers of target, context, and joint feature sets are 0.911, 0.916, and 0.921, respectively. On the other hand, the best performances achieved by deep feature analyses from the target, context, and combined feature sets are 0.906, 0.927, and 0.936, respectively. (See Table 5.3 above and Table 4 in the original paper.) From this comparison, it can be observed that deep features carry more predictive power than radiomics for context and combined nodule images.

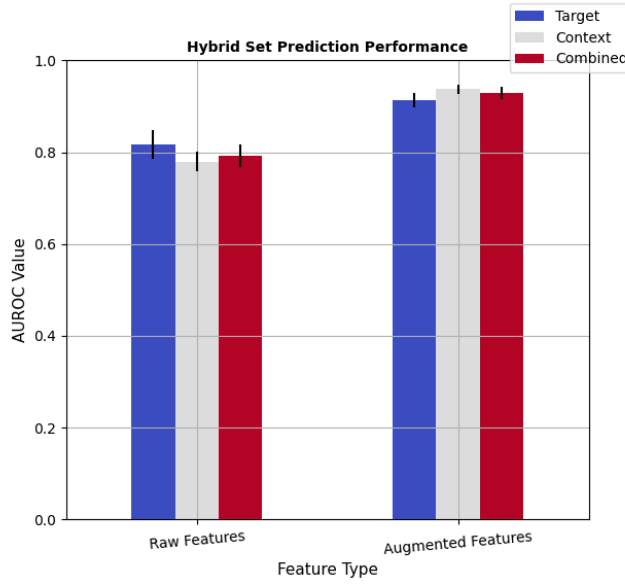


Figure 5.3. The predictive power of hybrid feature sets with different settings. [From Astaraki et al. 2021. Reprinted with permission.]

Hybrid feature analysis. Three separate feature pools were created by combining target, context, and combined radiomics with the deep features extracted from the single- and dual-pathway networks. These merged descriptors were analyzed both as raw features and augmented ones. Figure 5.3 illustrates the quantitative comparisons between the mentioned feature sets. The results show that merging the unbalanced radiomics with unbalanced deep features could improve the accuracy of the unbalanced radiomic set from $AUROC_{radiomic-combined} = 0.774$ to $AUROC_{hybrid-combined} = 0.793$. However, this improvement still cannot beat the accuracy of dual-pathway end-to-end deep classifiers, which achieved an AUROC of 0.824. On the other hand, remarkable improvements were observed in the accuracy of the hybrid feature set after augmenting the radiomics and deep features. In other words, merging the balanced radiomics with balanced deep features not only successfully improved the discrimination power of radiomics itself but also led to a slight improvement in the predictive power of the deep features as well. In addition, the

hybrid context feature set carries more predictive power than the hybrid target set, which is consistent with both the radiomics and deep feature sets alone. Moreover, the best discrimination power (AUROC = 0.938) was achieved by merging the balanced context radiomics with the balanced deep features extracted from context nodule images. Tables 8 and 9 in the Supplementary Materials represent the statistical comparison of the AUROC curves between radiomics and deep features extracted from the context, target nodule images, and their combination. Finally, Table 5 in the original paper provides a summary of the best results achieved by different settings to better compare the performance of each method.

5.2.3 Paper V

Early survival prediction in non-small cell lung cancer from PET/CT images using an intra-tumor partitioning method

The main findings of this paper are based on the following: (1) feature sets stability, (2) predictive power comparisons, and (3) feature set interpretability. In the feature sets stability phase, a reproducibility test was employed to identify reproducible features by calculating the CCC metric. In addition, FFS, as a wrapper algorithm, was applied to both the radiomic and SALoP feature sets to reduce the dimensionality of the feature sets and preserve only the features with higher predictive power. The reproducibility of the proposed SALoP feature set on CT data was substantiated by achieving a high agreement (CCC > 0.92). In fact, the proposed SALoP features were extracted from the longitudinal RIDER test-retest dataset and underwent the reproducibility test from which all the extracted SALoP features were recognized as highly reproducible descriptors. On the other hand, when the conventional radiomics were extracted from the same RIDER dataset, only 47% of them were recognized as reproducible (CCC \geq 0.8). Therefore, in all the further analyses, all SALoP features were used, while 53% of the CT radiomics were removed from the feature sets. In addition, dimensionality reduction with the FFS methods on delta radiomic features led to selecting more than 10 radiomic features for each of the PET, CT, and PET-CT feature sets. However, the same FFS method resulted in two PET features, three CT features, and three PET-CT features. Interestingly, the selected SALoP PET features come from the inner layers (closer to the tumor core), while a combination of inner and outer layers formed the more prognostic SALoP CT features.

The predictive power of the proposed SALoP feature set and conventional radiomics were compared on the following grounds:

Delta SALoP vs. Delta Radiomics. Table 5.4, which is a summary of Tables 1 and 2 in the original paper, represents the calculated prognostic values of the two delta feature sets and their combination as well as with and without applying the FFS method in terms of AUROC. The term delta stands for the differences between the extracted features from pretreatment and second scans. Comparing the prognostic power of SALoP set against the radiomics set, it can be inferred that the proposed SALoP outperformed the radiomic descriptors by a great margin when dimensionality reduction was not applied. On the other hand, for CT and PET images, the radiomics set performs better than the SALoP set after removing the non-informative features from the pools. More interestingly, the highest predictive power was achieved when the SALoP and radiomics sets were combined into a single set and followed by applying the FFS method. The statistical comparison of AUROC between SALoP and radiomics (Table 4, Supplementary Material) indicates significant differences between the SALoP and radiomics features extracted from PET and PET-CT images.

Table 5.4 - Prognostic values of SALoP, radiomics, and their combination with and without feature selection after applying 10-fold cross-validation in terms of AUROC metric. [From Astaraki et al. 2019. Reprinted with permission.]

| Modality | Without FFS | | | With FFS | | |
|----------|--------------|------------------|------------------------|--------------|------------------|------------------------|
| | <i>SALoP</i> | <i>Radiomics</i> | <i>SALoP+Radiomics</i> | <i>SALoP</i> | <i>Radiomics</i> | <i>SALoP+Radiomics</i> |
| CT | 0.64 | 0.58 | 0.66 | 0.68 | 0.85 | 0.86 |
| PET | 0.87 | 0.62 | 0.63 | 0.88 | 0.89 | 0.95 |
| PET-CT | 0.90 | 0.63 | 0.69 | 0.94 | 0.90 | 0.94 |

Imaging modality. Comparing the predictive power delivered by the PET imaging modality to those of the CT modality, one can infer that both radiomics and SALoP followed the same pattern. Specifically, for both cases, the predictive values of PET-based features are higher than CT-based features, and their combination even improves the predictive power. Surprisingly, the highest prediction score was achieved from the combination of the SALoP and radiomics extracted from PET images which was slightly higher than PET-CT images.

Single-timepoint analysis. This set of analyses refers to separately computing the prediction score of the features extracted from pretreatment and second scans. Tables 5 and 6 in the Supplementary Materials report the quantitative values in detail. In brief, the single-timepoint feature sets carry lower prognostic power than the delta features. In other words, delta features can capture the differences in the tumor characteristics from the longitudinal data, which are highly associated with the treatment effects.

5.2.4 Paper VI

Spherical Convolutional Neural Networks for Survival Rate Prediction in Cancer Patients

The proposed SphCNN-based features were compared against 17 baseline methods, which combined different types of feature sets and learning algorithms. We used the Concordance Index (C-Index), which is a commonly used metric for problems with right-censored data such as SRP [261]. This metric measures how well the survival times of a set of patients are ranked and can be seen as a generalization of the AUROC metric. Since two different datasets for NSCLC patients were studied, both intra- and inter-dataset experiments were conducted. While intra-dataset performance was evaluated using a 5-fold cross-validation procedure, the inter-dataset performance was assessed by training the models on the Lung1 dataset and testing it on the Lung3 dataset. Extracted CNN-based deep features after feature selection (DIF32) trained with EGB resulted in the highest C-Index for the Lung1 dataset (0.62 ± 0.04) that outperformed the best result achieved by the proposed SphCNN-based features. In fact, to investigate the effect of segmentation accuracy on the SRP scores, the features were extracted from tumoral regions defined by both the manual masks and the automatically generated masks. However, none of them could achieve as good a performance (0.58 ± 0.04 and 0.59 ± 0.03 for manual and automatic, respectively) as the combination of the EGB-DIF32 setting. On the other hand, having trained the models on Lung1 and testing them on Lung3 and HN datasets, the proposed method yields the best performance (C-Index = 0.64), followed by the combination of EGB and selected radiomics features (RF32) that achieved a C-Index of 0.63. In addition, the efficacy of the EGB-DIF32, which resulted in the highest performance in the intra-dataset analyses, was reduced to 0.61 in terms of the C-Index.

For each of the intra- and inter-dataset analyses, the validation set of the best-performing model was stratified into risk groups based on the median of the predicted risk scores. Therefore, from each validation set, two subgroups were formed to be evaluated by a non-parametric Kaplan-Meier (KM) estimator (see Figure 5.4). The observed trends in the KM curves are in line with the quantitative values of features-based analyses. While the KM curves of intra-data analyses show mixed observations, more promising stratifications can be observed from the inter-dataset analyses.

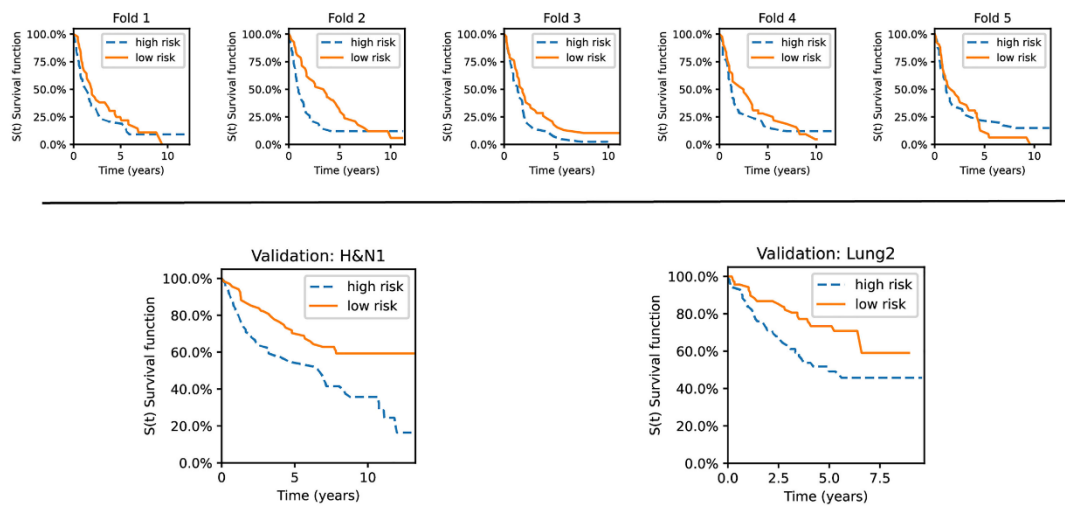


Figure 5.4. Kaplan-Meier curves of the validation sets. The best-performing model was used to stratify the data into low and high-risk groups. The first row shows the curves from the cross-validation over the Lung1 dataset. The second row shows the curves of the model trained on the Lung1 dataset and tested on HN and Lung2 datasets. [From Sinzinger et al. 2022. Reprinted with permission.]

Chapter 6

Discussion

As the average lifespan of humans is increasing across the world, medical experts aim to diagnose and treat diseases at an earlier stage by picking up the signs of serious diseases as they arise. In the domain of a disease or health condition, Disability-Adjusted Life Years (DALYs) are the sum of the years of life lost due to premature mortality and the years lived with a disability due to prevalent cases of the disease or health condition in a population [262]. According to the statistics provided by WHO, cancer imposes the largest burden across the world (244.6 million DALYs), both in men (137.4 million DALYs) and in women (107.1 million DALYs) [263].

Large-scale data have the potential to change how oncologists manage, analyze, and leverage data across different applications. Properly analyzing the health care data has the potential to reduce the costs of cancer diagnosis, prognosis, treatment, and outcome predictions which, in general, will improve the quality of a cancer patient's life. Personalized approaches span the full spectrum of cancer care. In fact, a personalized risk evaluation can identify patients at high risk of developing specific types of cancer, so they can be followed with more careful screening and prevention strategies, which will result in more effective early diagnoses and treatment methods [264]. The applications of Personalized Medicine (PM) in cancer are quite broad, entailing screening, diagnosis, prognosis, prediction of treatment efficacy, patient follow-up after surgery/treatment for assessing the early response, detection of recurrence, and the stratification of cancer stages. The objective of PM is to replace the "one size fits all" model of medicine, which has centered on reaction to disease based on average responses to care [265]. As one of the main pillars of cancer care systems, medical imaging modalities provide valuable quantitative information through a non-invasive or minimally-invasive procedure. Medical imaging data plays a central role in cancer management, including screening [266], prediction [267], biopsy guidance [268], staging [269], surgery/treatment planning [270], therapy response [271], recurrence [272], and palliation [273]. The essential need for medical images, however, has resulted in the acquisition of a huge number of imaging scans. For instance, in the United States (US) alone,

around 600 million medical imaging procedures are performed each year, and this number is estimated to approach 625 million scans in 2022 [274]. This estimation means that, on average, all US citizen undergoes two imaging scans in 2022. Considering the growing role of medical imaging data on one side and the challenges of manually examining such an abundance of data on the other side, the development of computerized tools to automatically or semi-automatically examine the image data is of great interest. The abundance of acquired digital medical images increases the demand for more advanced software to facilitate the process of cancer screening. Parallel to this, the rapid advances in ML methods provide a growing number of smart tools that have been incorporated into many different applications. Hence, a variety of ML tools have been specified for the medical imaging domain, aiming to assist clinicians with their repetitive tasks in their practice workflow.

Tumor phenotypes, as can be seen in medical images, contain more information than can be readily processed by the naked human eye. Indeed, recent studies have shown that intricate morphological, intensity, and textural descriptors extracted from tumor mass can capture, to some extent, the heterogeneous characteristics of the cancerous regions and provide valuable information about the potential outcomes of the disease [250], [275], [276]. QIBs aim to extract quantifiable descriptors from medical images to assess the severity, degree of changes, and status of a disease or chronic condition relative to normal findings. In other words, QIBs can stand for underlying biological and pathogenic processes or responses to therapeutic interventions [277]. In this context, radiomics can translate the oncological imaging data into mineable space by quantifying the image phenotypic characteristics in an automated and objective way [119]. However, prior to quantifying the image volume data, a set of image preprocessing steps needs to be performed, and most importantly, the boundaries of the tumoral regions need to be delineated.

In general, the pipeline of imaging biomarkers development includes the following essential steps: standard image acquisition, image preprocessing, target region segmentation, feature extraction, and predictive model development. While the standard image acquisition step falls outside the scope of the conducted studies, this thesis incorporates six studies that, taken together, cover the mentioned steps. The image preprocessing techniques such as contrast enhancement, intensity normalization, and image registration were used in the six studies to prepare the image data for further processing steps. Studies I and II present two different DL-based approaches for tumor segmentation. The last four studies focus on adopting and developing imaging biomarkers with classical and DL-based methods for cancer diagnosis and prognosis. These diagnosis/prognosis

models were developed with end-to-end DL models (Studies III, IV, and VI) and/or with feature engineering strategies (Studies IV, V, and VI). In specific, Studies I and II focus on developing segmentation models as infrastructure that can be used to segment the tumoral regions for further quantification steps. In Studies III and IV, *prognosis* imaging biomarkers were developed and tested on early-stage lung tumors to classify the benign tumors from the malignant ones. In Study V, *response* imaging biomarkers were developed by introducing the SALoP feature set to quantify the response of the tumors to the applied therapy as early as two weeks after the beginning of treatment sessions. Finally, SphCNN-features were employed as *predictive* biomarkers in Study VI to predict the survival rate of NSCLC and HN cancer patients.

As explained in Section 3.4.2, the definition and development of imaging biomarkers require proper justifications to determine the relationship between the extracted features and the studied disease. In this context, the dual-pathway model used in Studies III and IV was developed based on the fact that challenging lung tumors can be better stratified by considering both intra-tumor heterogeneity-based features and contextual attributes of the tumors. Therefore, the dual-pathway model was developed to consider both sources of information simultaneously. In Study V, the SALoP feature set was proposed to quantify the changes in the tumor characteristics caused by the applied treatment from more homogeneous subregions. In fact, different cancer cells with different genetic characteristics form different heterogeneous subclones within the tumoral mass. In practice, different tumoral subclones respond to the applied treatment differently. Therefore, the SALoP biomarkers proposed to capture the treatment effects from different tumoral layers. Lastly, it has been hypothesized that the survival status of cancer patients would not depend on the orientation of the tumor. Thus, the SphCNN model in Study VI was developed to learn and extract the rotational invariance biomarkers.

The reliability of the developed imaging biomarkers can be validated by conducting phantom-based experimental studies or comparing them against the ground truth clinical measurements (see Section 3.4.5). In this domain, manual contour annotations by expert radiologists were set as the ground truth, and the accuracy of the segmentation models in Studies I and II was examined by comparing the predicted segmentation masks against the ground truth. In Studies III and IV the ground truth of each subject was determined by a clinical diagnosis of pathologies. The ground-truth labels of Studies V and VI were determined by following the subjects after the applied treatments. Hence, in all six studies, the clinical measurements were employed as the ground truth, and the models were

developed and optimized to achieve the ground truth values. Of course, in each study, several baseline models were analyzed as well to evaluate the performance of the proposed model with respect to the conventional methods.

These six studies contribute to the field of oncological precision medicine by introducing novel techniques for tumor identification, diagnosis, prognosis, and treatment response evaluation. In the following sections, the primary outcomes and implications of all studies will be discussed analytically, and the pros and cons of the proposed methods will also be outlined. Finally, the limitations of the studies and potential future works will be discussed as well.

6.1 Deep Learning and Tumor Segmentation

The segmentation of tumoral regions from different medical imaging modalities is an essential step in clinical diagnosis and interventions such as radiation therapy and surgery. Relying on manual delineations requires not only expert radiologists but is also a time-consuming and error-prone task. With the outstanding achievements of DL methods in the computer vision society for natural image segmentation and the success of the U-Net-based models in different medical image segmentation applications, DL methods have become the first choice for automatic tumor segmentation tasks. Many advanced segmentation models have been introduced for segmenting different types of tumors such as lung, head-neck, brain, liver, kidney, breast, and colorectal [278]. However, to achieve more accurate segmentation results, the model architectures and training procedures are becoming more complicated, which in turn often requires more training data. Nevertheless, labeling medical images requires enormous manual effort, which is not an optimal solution. On the other hand, UAD methods aim to lift the dependency on the labeled data, but the current knowledge in this field is not mature enough to achieve reliable results yet. The first two studies of this thesis focused on elaborating the deep segmentation models as attempts to address the two problems mentioned. Study I aims to efficiently employ the currently available tools in order to improve the segmentation accuracy by obtaining prior information regarding the pathologies without needing more labeled data. Study II aims to achieve acceptable segmentation accuracy in a purely unsupervised training procedure.

6.1.1 Prior-aware segmentation model

A segmentation prior encodes a particular type of prior knowledge into a segmentation model that has been shown to be a beneficial strategy to improve the segmentation accuracy.

Such prior information includes but is not limited to shape, appearance, size, location, boundaries, edges, and atlases of the target region and can be applied locally or globally [18], [243]. Although leveraging prior information in rule-based segmentation models such as energy functional minimization [279] and template matching [280] techniques is a well-established approach, in the context of CNN models, integrating prior knowledge into the segmentation models is an ongoing problem. Training convolutional autoencoders [281] and VAEs [282] to learn shape manifold space, integrating conditional random field into CNNs in a joint training procedure [283] or a post-processing step [284] to constraint the segmentation boundaries, and modifying the conventional objective functions into topological-aware loss functions [282], [285]–[287] for coding geometric and shape priors are among the recent efforts for adding prior knowledge into segmentation networks. Employing such techniques has led to substantial improvements in the segmentation accuracy of healthy anatomies such as the liver [285], cardiac muscle [243], [288], and multiple abdominal organs [289], [290]. However, obtaining prior knowledge of pathological regions remains an open challenge. In the context of lung pathologies, the presence of various subtypes of the same pathologies with different characteristics makes it even more challenging to obtain robust prior information from the pathologies. For instance, while non-solid nodules and initial Covid-19 infection appear as Ground Glass Opacity (GGO) with low-intensity contrast against the healthy backgrounds, the borders of homogeneous consolidation masses as well as juxta-pleural and juxta-vascular nodules cannot be easily distinguished from the nearby tissues [291]–[293]. Lastly, while the diameter of pulmonary nodules lies in the range of 3 to 30 mm, advanced stages of lung tumors can become larger than 50 mm [294]. This size can be even worse in the case of “crazy-paving” COVID-19 lesions, which can occupy a major capacity of the lungs [295]. To address the mentioned difficulties in obtaining prior information about the lung pathologies, we propose to first capture the prior knowledge for healthy lung structure instead of lesions and then derive the prior knowledge of pathological parts indirectly. We hypothesized that the proposed NAA model would remove the pathological regions and generate high-quality pathology-free images from which the prior information can be estimated by computing the differences between the pathological input images and the reconstructed pathology-free images. The calculated prior information was then integrated into a U-Net-like network as an additional channel along with the original image to improve the segmentation performance.

The design of the supervised NAA model is based on the idea that the abstract latent attributes of healthy lungs are too complicated to be captured by unsupervised methods. However, training a supervised model essentially requires target label images which, in this

case, are the pathology-free images. Such faked images were synthetically generated by a robust inpainting model. This supervised training procedure reinforced the NAA model to preserve the healthy anatomical structures while removing pathological regions. Hence, the NAA model could, to some extent, learn the appearance of healthy lungs even without applying some preprocessing steps such as registering the images into an atlas in order to reduce the effect of anatomy variations. This method is in sharp contrast to conventional representation learning models that aim to learn the appearance of healthy organs, such as UAD methods. In fact, such representation learning models are trained based on unsupervised learning schemes and often are not able to reconstruct high-quality images that preserve fine-grained details [21].

The prior images obtained from the output of the NAA model consist of highlighted regions, along with other structures, that represent the prior information about the shape and location of the pathologies. Integrating the prior image into the segmentation model as an additional channel seems to be a constructive strategy to improve the segmentation accuracy. In particular, the prior image can be considered to be an attention channel showing the importance of different regions that can guide the network training to more meaningful regions and reduce the segmentation uncertainties. This has been verified by looking at the quantitative results where the segmentation accuracy of the prior-aware models outperformed the baseline models by substantial margins for all types of studied lung pathologies. The overall segmentation performance of the prior-aware model demonstrated the capabilities of the NAA model to obtain the prior image, which can boost the model functionalities in learning the sensitive pathological features.

Although prior knowledge of the lung pathologies was not estimated directly, the proposed pipeline shows great potential in learning the appearance of healthy lungs to deduce the prior knowledge of pathologies indirectly. To the best knowledge of the author, Study I is the first study attempting to capture the prior knowledge of lung pathologies. The numerical metrics show the efficacy of the prior-aware segmentation model, which significantly outperformed the baseline segmentation networks.

6.1.2 Unsupervised segmentation model

Recent breakthroughs in DL models have led to automated medical image segmentation models which even achieve expert-level performance in the detection and segmentation of anatomical organs or pathologies. Despite the outstanding segmentation accuracy, these supervised methods have encountered some disadvantages. First, their supervised training

procedure requires large-scale and diverse annotated image data, which in practice are costly to obtain. Second, the performance of the trained model is limited to detecting certain types of lesions, which are similar to those of the training set. Therefore, if certain lesions are not presented or are under-presented in the training sets, the model would fail to predict meaningful results. Recently, to overcome the necessity of expensive labeled data, unsupervised-based DL methods have emerged as promising tools to detect pathologies from arbitrary types, which are recognized as UAD methods. These methods aim to resemble how radiologists examine imaging scans. In fact, expert radiologists are trained to learn the appearance of healthy anatomical regions. Therefore, they do not need data with pixel-level annotations because they can detect arbitrary abnormalities as outliers with respect to the healthy anatomies [240][296].

One of the major limitations of conventional UAD models is that they hardly learn the appearance of healthy anatomical structures with fine-grained details. Instead, they often tend to learn a general representation of anatomical structures without preserving anatomical constraints. One of the objectives of Study II has been to introduce a model with the capabilities of generating high-resolution medical images while preserving the anatomical details in the process of representation learning. The conventional AE-based models are often trained by optimizing per-pixel loss functions that tend to reconstruct blurry images. Therefore, one potential approach is to modify the objective function in order to improve the quality of the reconstructed images. Therefore, more advanced types of objective functions such as perceptual loss and style loss can potentially increase the conceptual and textural quality of the generated images. However, integrating these objective functions into the conventional representation learning models would degrade their ability to learn the latent characteristics of the healthy anatomies. In other words, such modified models tend to learn a wide range of image-based details and can hardly discriminate normal structures from anomalies. In fact, such fortified objective functions increase the risk of model overfitting with respect to representation learning tasks. On the other hand, limiting the convolutional operators with image subregions can regularize the learning process of representation learning models and avoid the overfitting problem. In particular, while the powerful objective function is prone to overfitting on the details of anatomical structures, localizing the functionality of convolutional operators can potentially counteract this tendency. Accordingly, considering the functionality of the PConv [225] operators, they can be a perfect choice for this problem as they deal with local convolutions instead of ordinary global convolutions. As a result, the representation learning process in Study II was turned from conventional AE- and GAN-based models into an image inpainting

problem. In practice, leveraging the inpainting model with multi-term objective function as an optimization algorithm and Gconv operators [226] as localized convolutional backbones could successfully enforce the model to synthesize the high fidelity realistic-looking medical images while preserving the anatomical constraints, regardless of the imaging modality.

The proposed autoinpainting pipeline for unsupervised tumor segmentation yielded interesting results when the PET imaging modality was used either as single modal images or multimodal images. In such scenarios, the segmentation accuracy of the proposed unsupervised pipeline was not far behind the performance of the supervised nnU-Net model. Of course, the hyperactivities caused by tumoral regions facilitate tumor localization and, therefore, tumor inpainting. Nevertheless, the capabilities of the proposed inpainting model were not limited only to hyperintensity signals of PET images as the pipeline could detect and inpaint the challenging NSCLC tumors in CT images as well. Highly similar visual attributes of NSCLC tumors with respect to the surrounding soft tissues make them challenging for segmentation tasks even with supervised models. Nevertheless, the proposed model could inpaint challenging cases and lead to promising results in the context of unsupervised models. Regarding the limitations of the proposed pipeline in the context of HN tumors in CT images, it is worth mentioning that even the supervised segmentation methods can hardly detect the HN tumors in full resolution CT images. For instance, in [297], a promising Dice score of 0.48 was reported for the HN tumor segmentation in CT images by analyzing a cropped region around the tumors.

Comparing the segmentation accuracy of the proposed pipeline against the conventional UAD methods can highlight the great potential of the autoinpainting model. In fact, the UAD models were not able to thoroughly learn the appearance of normal anatomies even with the PET images. Hence, these models could not successfully reconstruct tumor-free images. Even if they managed to remove the tumors, they were not able to preserve the anatomical constraint and produced images with irrelevant anatomical structures, which in turn substantially increased the false-positive rates.

Although this study has compared the performance of unsupervised methods with the powerful nnU-net supervised model, it should be noted that such comparison is not fair. In fact, the only reason that the supervised nnU-Net model was examined is to estimate the maximum accuracy, which can be achieved on the same datasets. While the unsupervised segmentation methods can potentially overcome the disadvantages of supervised models, the current models have not been robust enough to yield results that are as accurate as supervised models. In Study II, an inpainting-based UAD method was proposed to segment

the NSCLC and HN tumors in multimodal and single-modal images. To the best knowledge of the author, this is the first attempt to segment such challenging tumors with unsupervised DL methods. The quantitative results verified the potential of the proposed pipeline, which significantly outperformed the conventional UAD models.

6.2 Imaging Biomarkers and Benign-Malignancy Classification

Distinguishing malignant pulmonary nodules from benign nodules is considered to be one of the most challenging steps carried out by clinicians in the lung cancer management pipeline. In fact, the identification of malignant nodules is an important step toward the early detection of lung cancer, offering the best chance for a cure. In clinical practice, radiologists visually examine the image slices and estimate the likelihood of malignancy by relying on morphological characteristics. However, the accurate identification of malignancy is made with invasive biopsy-based cytological analysis [298]. Therefore, robust image-based characterization of lung nodules can save the patients from such invasive interventions. Studies III and IV aimed to discern malignant pulmonary nodules from benign nodules in LDCT images by using imaging biomarkers. In practice, the presence of highly similar visual characteristics shared between benign and malignant nodules on the one hand and the concept of intra-nodule heterogeneities on the other make this classification task a challenging problem. In addition, the sensitivity of the classification results to the delineation accuracy and the class imbalance of the dataset are among the technical difficulties of the problem.

The dual-pathway deep classifier framework was proposed to learn the latent characteristics of the nodules both from target nodule images and context nodule images. The underlying reasons for this choice are (a) that target nodule images can represent the intra-nodule heterogeneity attributes and (b) that context nodule images can capture the relative position of the nodules concerning surrounding tissues, which is associated with the malignancy [23]. In addition, the context images cover the nodule regions and minimize the dependency of the classification power on the delineation accuracy. Accordingly, it was expected that the dual-pathway model learns intra-nodule attributes along with context features jointly in one single framework. This strategy further assures that the final classification scores produced by the DL models are derived from the pathological regions.

The quantitative results achieved in Study III show that the designed dual-pathway architecture has a strong relationship with nodule malignancy. In particular, the strategy of joint learning could successfully improve the classification performance of the end-to-end

models compared to scenarios where the target or context nodule images were trained separately. In addition, comparing the discrimination power of context images and target images implies that context images contain more predictive power than target images. This observation can be explained by the fact that context images not only constitute the intra-nodule characteristics but also represent the relative location of the nodules with respect to the surrounding tissues as well. This finding can lift the need for accurately segmenting the nodule regions, which is a very challenging problem. Instead, employing a cropped region around the nodules is often considered a more straightforward problem. As was expected, training supervised deep classifiers resulted in higher predictive power than the unsupervised CNNs. In fact, supervised models are trained to directly learn the abstract features that are more relevant with respect to the class labels. On the other hand, unsupervised CNN-based features tend to capture more semantic features. Nevertheless, unsupervised models can still be good candidates for training image volumes with a limited number of labeled data [299].

While achieving impressive results, the architecture and training protocols of DL models have become more complicated. These complexities help the models to better learn and select the relevant features. However, it is still difficult to understand how a network draws its decisions [300]. Therefore, the interpretability of DL models is an ongoing problem. On the other hand, explainable properties such as intra-nodule heterogeneity and the morphological characteristics of the nodules make radiomics an alternative approach for benign-malignancy classification. Therefore, radiomic descriptors were extracted from target nodule and context nodule images to conduct a fair comparison against the DL models. In the literature, a wide range of feature selection methods were integrated into a variety of learning algorithms in order to build effective prediction models [301]–[305]. However, a consensus among the proposed methods can hardly be reached. Therefore, to objectively determine the best scenario for analyzing the radiomics pipeline, the extracted radiomic descriptors were analyzed with eight different conventional learning algorithms and seven different standard feature selection methods. An evaluation of the discrimination powers obtained by radiomics and CNNs implies that end-to-end training of the deep classifiers without any further fine-tuning steps can perform better than the radiomics pipeline. This, of course, is related to the inherent advantage of the end-to-end learning procedure of CNNs over hand-crafted radiomics. On the other hand, fine-tuned radiomic features carry more predictive power than fine-tuned CNN-based features.

In binary classification tasks such as benign-malignancy problems, class imbalance occurs when one class—the minority class—consists of substantially fewer samples than the other class—the majority class. The minority class is often the class of interest, that is, the malignant subjects. Properly training a learning model over such imbalanced datasets can be challenging because the models tend to over-classify the majority class due to its increased prior probability [306]. Therefore, the samples belonging to the malignant class are misclassified more often than those belonging to the benign class. However, synthesizing new samples in the feature space by using SMOTE algorithm could successfully avoid such a bias toward the benign class. In fact, SMOTE generated new samples from the malignant class to balance the distributions of the two classes. This strategy reduced the misclassification of the malignant group and therefore improved the classification accuracy by a substantial margin. To assure that the synthesized features were not only a replication of the original features, which would increase the risk of overfitting, a new experiment was conducted. The generated features by SMOTE were included only in the training set. This essentially means that if the synthesized features were a replication of the original features, then the classification accuracy of original features when the model was trained on synthesized features should reach the highest possible value as a sign of overfitting. However, the presence of the relatively large values of false negatives can increase our confidence in the efficacy of the employed SMOTE without being concerned about the overfitting issues.

The discriminative power of the model was improved by combining deep features with radiomics. These improvements point to the fact that the hybrid feature set can represent a wide range of lung nodule characteristics, from handcrafted textures to abstract deep features. This observation aligns with other relevant studies on lung nodules [307], [308]. Nevertheless, combining these two distinct feature sets in order to enhance the predictive power should be conducted carefully. In other words, simply concatenating the two feature sets will dramatically increase the length of the features, which in turn can increase the risk of overfitting. More importantly, it can even worsen the class imbalance issue, which would adversely impact the performance of the models. The observed inferior performance of the raw hybrid feature sets can be caused by the described reasons. However, fine-tuning the hybrid feature sets by balancing the classes and reducing the dimensionality of the features by effective FS methods such as FFS are beneficial strategies to efficiently gain from the hybrid feature pools that result in the highest predictive power.

In Studies III and IV, the problem of lung nodule benign-malignancy classification was investigated. A dual-pathway framework was proposed to extract features from target nodule and context nodule images. The performance of end-to-end deep classifiers, radiomics, DL-based radiomics analysis, and hybrid feature sets was objectively compared before and after fine-tuning the prediction models. It can be concluded that the end-to-end deep classifiers resulted in more accurate classification power compared to the classical radiomics when the fine-tunings and optimizations were not applied. The experiments show that context and nodule images can have a complementary role as the combination of their features improved the classification power both for the deep classifiers and radiomics pipeline. In addition, quantitative comparisons show that incorporating the radiomic features into the learned deep features can be beneficial and improve the classification accuracy if proper fine-tuning steps such as feature selection and class label balancing are applied to the feature sets.

6.3 Imaging Biomarkers and Evaluation of Early Response to Therapy

The primary goal of Study V is to develop a novel, physiologically meaningful imaging biomarker set that can capture intratumor heterogeneity based on the characteristics of CT and PET images. The proposed SALoP aims to partition the tumor volumes into separate concentric zones; and from each zone, mean intensity changes of PET and CT values were quantified to describe changes in the tumor characteristics caused by the applied treatment. In other words, dividing a tumor mass into several zones based on the distance from the tumor borders can represent a single heterogeneous mass by a set of constituent homologous subregions. Since remarkable intensity variations within each partitioned region were not observed, it can be inferred that tumor properties could be quantified from more homogeneous local areas.

In addition to the concepts of biological and clinical validations of imaging biomarkers, which were described in Section 3.4.5, the interpretability of imaging biomarkers can reveal their relevance with respect to the physiological characteristics of the tumors. The proposed SALoP feature set was designed to include such physiological characteristics of the tumors; therefore, it can be, rather, easily interpreted. The voxel values of CT images represent the tissue density while the PET intensity values reflect the metabolic activities. Hence, changes in these values within the constituent tumor subregions indicate the impact of the applied therapy on the density of tissues as well as their regional glucose uptakes. Accordingly, the proposed SALoP set can quantify physiological attributes of distinct parts of the tumors,

such as variations in metabolic activity, oxygenation, and tissue inflammation. As a result, the quantitative values captured by SALoP are not just a set of uninterpretable numerical metrics; instead, they can be related to the factual reactions of tumoral cells to the applied treatment. On the other hand, changes in the values of radiomic features cannot be easily explained in terms of clinical routines [309]. This concept of interpretability would even worsen when it comes to the DL models. Therefore, the proposed SALoP set is superior to both radiomics and DL models in terms of interpretability. Despite the limited number of features quantified by the SALoP set, its prognostic power outperformed the high throughput radiomic feature sets. In fact, the SALoP set represents each tumor in PET-CT images with only 20 numerical values, which is only about 4% of the size of the studied radiomic descriptors. This implies that each subregion of the tumor has distinguished properties that are captured well by the proposed method, and their combination together can represent the diversity inside the tumor volumes, which is closely associated with the therapy response. On the other hand, the relatively poor performance of the radiomic features can be explained by the fact that many of the extracted descriptors are either redundant or non-informative. Hence, their combination is not an effective strategy to add value to the predictive power. Nevertheless, reducing the dimensionality of the radiomic feature set with the FFS method led to eliminating the redundant and non-informative features and preserving the radiomic signatures, which resulted in a remarkable improvement in the predictive power. This is in line with the observations in other studies as well [310]–[312]. The identified radiomics signature after applying the FFS method was reported by other relevant studies as well [313], [314]. More importantly, a majority of the radiomic signatures belong to the second-order feature categories, which point to the importance of the textural features aiming to capture intra-tumor heterogeneities [315], [316]. Nevertheless, comparing the performance of the selected SALoP features to the original SALoP set, we can infer that no significant improvement was achieved by removing the non-informative features. This implies that the proposed SALoP set has great potential to capture the tumor phenotypes regardless of whether the dimensionality reduction applies or not.

Combining the SALoP set with conventional radiomic features improved the predictive power. This interesting observation suggests that these features might complement each other. In fact, SALoP features quantify, to some extent, both geometric and first-order properties of the tumor volumes.

One of the advantages of FDG-PET image data is that they can quantify the changes in the molecular levels of the tissues even when such tiny changes cannot be detected anatomically. Accordingly, PET-based imaging features are expected to carry more prognostic power than CT-based features [317], [318]. This expectation was realized in Study V when the PET-based descriptors from both the SALoP set and radiomics delivered more predictive power than those of the CT features. Nevertheless, given the types of the characteristics acquired by CT and PET images, merging the PET-features with CT-descriptors allows for capturing a wider range of tumor characteristics, which in practice strengthens the predictive power.

Several studies have highlighted the prognostic potential of radiomics in patients diagnosed with NSCLC [319]–[322]. The underlying hypothesis is that genomic and proteomic characteristics of cancers can be expressed in imaging biomarkers, which can be quantified with optimal image analysis methods [1]. This could extend to the concept of radiogenomics, which implies that genomic tumor heterogeneities are associated with intratumoral heterogeneity in the imaging levels [1]. In this context, FDG-PET-CT-based imaging biomarkers have been shown to carry invaluable information for staging, evaluating treatment response, detecting recurrence, and predicting the prognosis of NSCLC [323]. In assessing the treatment response, previous studies have shown strong associations between the textural features and responses to radiotherapy [324] and chemoradiotherapy [316].

The SALoP feature set was proposed as a new way to capture the intra-tumor heterogeneities. This feature set showed promising results when employed to predict the survival status of the NSCLC patients. In fact, the prognostic power of the proposed SALoP set even outperformed the conventional radiomic features in assessing the early tumor response to treatment. The highly reproducible SALoP feature set can be interpreted physiologically as well. In addition, predicting the survival status after two years by analyzing the features extracted from the scans of the first two weeks of treatment with an accuracy of 0.96 is an outstanding performance achieved by the combination of SALoP and radiomics sets. Accordingly, this feature set has the potential to be employed as prognostic imaging biomarkers for NSCLC studies.

6.4 Imaging Biomarkers and Survival Rate Prediction

The prediction of the duration of time until an event occurs, such as cancer recurrence or death, underlies important clinical decisions in oncology. Survival analysis, in particular, holds great value for patients, oncologists, clinicians, and researchers. The naive cancer

prognosis prediction models rely on population-level measurements for a specific site of cancer and cancer stage [325]. Nevertheless, such methods fail to take into account the individualized characteristics of patients, even such basic ones as age at diagnosis. To address such limitations, patient-specific methods have been introduced in clinical practice that consider the combinations of clinical information and laboratory examinations with well-established biomarkers. Nonetheless, the outcome of survival prediction models often depends on the clinician’s subjective interpretation and intuition, which in turn limits the accuracy and reproducibility of the models [326]. In addition, laboratory examinations do not necessarily yield accurate results. For instance, the genetic heterogeneity of the cancerous cells degrades the value of biopsy-based genomic analysis. In this context, however, medical imaging modalities can potentially illustrate the entire tumor in a non- or minimally invasive and repeatable procedure. To discover the correlation between medical images and underlying genetic characteristics of the tumors, which can be used for survival analysis applications, various imaging biomarkers have been proposed, such as radiomics and DL-based methods. On the other hand, predicting the survival rate by relying only on the shape, size, and texture of the tumor is a challenging task because it is not clear yet whether the imaging data would be representative enough or not. In addition, the prediction might be affected by different factors, including the presence of right-censored data, image acquisition parameters, inaccurate segmentation masks, the selected imaging biomarkers, and the prediction model itself. The primary goal of Study VI was to predict the survival rate from CT images with a fully automatic pipeline based on SphCNNs and compare it to the relevant conventional methods.

The quantitative results from the experiments show that the selected features from the pre-trained deep classifier resulted in the highest predictive power for intra-data analysis. On the other hand, the predictive power of the entire radiomics pool was slightly higher than the subset of selected radiomics. The observation that the same feature selection method leads to improving the performance of DL-based features but not the radiomics pool can be explained by the fact that DL-based features were extracted from a single 2D tumoral slice while the radiomics were extracted from the complete tumor volumes. Hence, the less diverse features of the tumors in 2D slices, which were captured by the DL model, are more prone to show a stronger association with the target labels compared to the 3D radiomics, which were extracted from the irregular tumor volumes with a wide range of texture, intensity, and morphological characteristics. The proposed SphCNN method automatically quantifies morphological features in the spherical domain. In other words, the spherical mapping methods can be seen as a compact representation of tumor surface texture, size,

and shape. Employing these spherical signals combined with a rotation-invariant SphCNN led to obtaining C-indices comparable to radiomics and deep features in the intra-dataset experiments. However, the proposed method slightly outperformed the conventional ones when referring to the inter-dataset analysis. These results suggest the robustness of the proposed SphCNN-based survival rate prediction for the test phase on completely unseen datasets. The quantitative metrics also indicate a rather similar performance on both the NSCLC and HN datasets. This finding suggests that the morphological features that the SphCNN internalized during training might have prognostic relevance for tumors in general. However, since the differences between the proposed method and the best-performing baselines were not substantial, it can be understood that the proposed method results in competitive performance overall.

6.5 Advances and Limitations of Machine Learning Methods in Cancer Screening

ML and its subfields, such as DL, offer the prospect of improving diagnosis, prognosis, and treatment analysis in the health care systems. Among all the healthcare domains in which ML methods may be applied, the medical imaging domain plays a central role because of the heavy burdens related to the manual analysis of high-dimensional image data by expert radiologists. Successful integration of such methods in clinical practice depends on different factors such as robustness, accuracy, stability, computational time, costs, and ethical conduct. Although many of the developed methods have already been approved by regulatory authorities such as the U.S. Food and Drug Administration [327], [328], a vast body of these automatic methods is still in their initial phase, undergoing the critical appraisal and independent assessments [329].

Rapid advances in ML algorithms involve several key technologies in the medical imaging domain, including but not limited to image reconstruction [330], image enhancement [331], image registration [331], image classification [332], and image segmentation [333]. This thesis consists of six studies focusing on cancer image analysis, from which four studies contributed to diagnosis and prognosis applications through image classification, and two studies contributed to the tumor segmentation problem. Despite the novelties introduced in these studies and the encouraging results achieved for different tasks, there still exists much room to develop more accurate and/or efficient models for classification and segmentation tasks. However, considering the unprecedented achievements of recent years in oncological image analysis applications, one can expect that the rapidly growing ML techniques will lead to even more promising breakthroughs in the

future. As an example, the best segmentation accuracy reported in the Decathlon challenge [334], [335] for the lung tumor task in 2018 was 0.69 in terms of Dice metrics. However, the best performance reported on the same dataset with more advanced methods improved up to 0.77 [336]. Although these results indicate that current methods are not accurate enough to be accepted for clinical applications, such a great improvement achieved from limited training data is evidence of the rapid development of robust DL models. As another example, one can point to the Brain Tumor Segmentation (BraTS) challenge [337]. While the winner of the BraTS 2017 challenge achieved Dice metrics of 0.729, 0.785, and 0.886 for enhancing part, core part, and whole tumors, respectively [338], the top performance in BraTS 2021 was improved to 0.908, 0.941, and 0.946 for the same subregions [339], [340]. It should be noted that such substantial improvements have been achieved in the light of more powerful computational resources, more labeled training data, and of course, more elaborated pipelines.

The six studies presented in this thesis aim to provide relevant contributions to the oncological image analysis field. Nevertheless, each of these studies faced some limitations that should be considered for investigation in the future. These limitations are outlined in the following subsections.

6.5.1 Imaging datasets

Image data is the basis of many ML-based methods. The small size of the datasets increases the risk of overfitting; however, we must be cautious regarding the potential biases caused by a single large-scale dataset. A practical way to reduce such uncertainties is to utilize several large-scale public and/or private datasets to test the performance of the new methods objectively. In Study V, for example, the performance of the model was tested over a dataset containing only 30 subjects. Although the employed method was not a data-greedy DL model and several strategies were applied to reduce the risk of overfitting, the potential of the proposed SALoP biomarkers may be further investigated on large-scale datasets of longitudinal PET-CT scans. In Studies III and IV, a large-scale LDCT of pulmonary nodules were studied for benign-malignancy classification; however, it is still necessary to validate the performance of the dual pathway model on another comprehensive dataset such as LIDC-IDRI [244]. The same limitation exists within Study II in which the performance of the proposed autoinpainting pipeline was examined only on one single NSCLC and one single HN dataset.

In addition to the sample size, the generalization power of the methods could be examined in other types of cancers as well. For instance, the abilities of the SALoP biomarkers in capturing the tumor characteristics could be investigated on highly heterogeneous brain tumors as well, given the availability of longitudinal datasets. In addition, the spherical CNN model of Study VI could be examined for brain tumor survival prediction, which is already known as a challenging task [341]. In the same context, the ability of the unsupervised autoinpainting model can be examined on the brain tumor segmentation task from MR images, which is the center of attention for many UAD methods.

6.5.2 Volumetric analysis

While in Studies III to VI, volumetric medical images were analyzed, the segmentation models introduced in Studies I and II deal with 2D slices of volumetric images. In other words, the volumetric context of the images was not being used to segment the tumors. Of course, this limitation arises from the challenges of volumetric inpainting models. Extending the inpainting model from 2D to 3D did not lead to acceptable results in the limited conducted experiments. Therefore, the pipelines in their entirety were simplified into 2D models with the cost of losing the volumetric context. In fact, the objective of studies I and II was not to develop a 3D inpainting model; however, considering the potential role of a robust volumetric inpainting model, it is worth focusing on this problem in future studies.

6.5.3 Incorporating domain knowledge

As previously mentioned, most of the ML models in the domain of medical image analysis are adopted from the models developed explicitly for general computer vision applications. Nevertheless, medical images are more complicated due to the presence of large variations in human anatomies, high inter-class similarities, the limited size of labeled data, and noisy annotations. Although integration of weak domain knowledge, such as anatomical attributes of medical images, into ML models is a relatively straightforward task, it is much more demanding to capture and integrate the strong domain-specific knowledge that expert radiologists rely on. The idea of partitioning a tumor into subregions, and dual pathway models used in Studies III, IV, and V, to some extent, were inspired by such domain knowledge. However, the integration of more relevant domain knowledge regarding the tumor heterogeneities may help increase the performance of the survival analysis and classification models. In addition, learning the distribution of healthy anatomies and integrating them into the UAD models has been studied for a long time. Regardless of how

they have been done, all these attempts—including Studies I and II in this thesis—leave room for improvements. Therefore, a potential direction for future studies can focus on the encoding of domain-specific knowledge in the models.

6.5.4 Methodology

The continuous success of DL models originates not only from the greater complexity of the network architecture and learning paradigms but also the more careful data preparation and preprocessing. Although the methodology of the conducted studies was carefully examined, other options exist, and these may be beneficial to overcome the difficulties of the proposed methods. Some of the potential methods that can be further investigated are discussed below.

Techniques to address class imbalance. In Studies III and IV, different approaches, including class weights and equal sampling across classes, as well as SMOTE, were examined; of these, SMOTE was found to be a beneficial technique to tackle this imbalance issue and improve the classification accuracy. However, there are other techniques [306] to modify the training procedure and/or objective function, which can be examined as well. This imbalance issue is not limited only to the classification problems, but also the performance of segmentation models may be negatively impacted by it. For instance, in the cases of multi-focal pathologies in which some of the lesions appear smaller than the others, the conventional Dice objective function cannot recognize the instance imbalance and the larger lesions dominate minor instances. The functionality of novel objective functions such as blob loss [342] would be helpful in addressing such an issue.

Multimodal fusion models. In Study IV, the extracted radiomic features were merged with learned deep features in a non-learnable approach. In other words, a deep classifier was trained, and the learned features were extracted and concatenated to the radiomics pool. However, this fusion can be done with other approaches [343] such as training an end-to-end classifier model [344].

Contrastive learning. In Study I, the differences between the generated tumor-free images and original tumoral slices were considered to be prior images with respect to the tumor characteristics. The residual images were then integrated into a segmentation network as the second input channel. This set of paired images between tumoral and tumor-free images can be a good candidate for training a segmentation model with a contrastive learning procedure [345]. In addition, the entire pipeline of Study I consists of three separate

modules. Regardless of how, squeezing the entire pipeline into one single model might be an effective strategy to reduce random and systematic errors.

Chapter 7

Conclusions

The six studies included in this thesis have all been conducted with the same endeavor: to refine the current state of oncological image analysis and improve our ability to quantitatively evaluate the tumor characteristics from minimally invasive or non-invasively acquired medical images. The overall results have shown that carefully designed imaging biomarkers can translate the raw voxel values into quantitative features from which discriminant patterns can be derived. In addition, integrating domain-specific knowledge into the segmentation models can increase the segmentation accuracy. Novel methods for different applications of cancer diagnosis and prognosis have been introduced with competing performance against the state-of-the-art methods. Specific conclusions from each of the six included studies are outlined below.

- Study I** A NAA model was proposed to capture the prior knowledge concerning the lung pathologies in CT images. The proposed NAA model could learn the appearance of healthy lung anatomies and indirectly produce prior information regarding the presence of pathologies. The segmentation accuracy of the prior-aware segmentation model outperformed the baseline models with significant margins in different types of lung pathologies, including NSCLC, nodules, and COVID-19 lesions.
- Study II** A purely unsupervised framework was proposed to segment the challenging lung and head-neck tumors based on the ability of an inpainting model to remove the tumors. The inpainting model could reconstruct high-fidelity medical images in full resolution without corrupting the anatomical constraints. The proposed autoinpainting pipeline resulted in promising segmentation accuracy and significantly outperformed a family of autoencoder-based UAD methods.
- Study III** The proposed dual-pathway DL framework is a beneficial strategy for learning disentangled image-based features to classify the benign and malignant

pulmonary nodules in LDCT images. In fact, integrating both target nodule and context nodule images into a single deep classifier could successfully capture and discern the intricate characteristics of nodules.

Study IV A comparative study was conducted to distinguish malignant pulmonary nodules from benign ones. The discrimination power of end-to-end deep classifiers, radiomic feature sets, learned deep feature sets, and hybrid feature sets were examined under the same fine-tuning conditions. The quantitative results suggest that the effective integration of radiomics and deep features improves the classification power.

Study V The introduced SALoP feature set can be implemented easily and interpreted physiologically. It is highly reproducible and is capable of describing intra-tumor heterogeneity. The proposed SALoP set outperformed conventional radiomics for predicting the overall survival status of NSCLC patients. Such specifications of the SALoP feature set can make it a candidate for a robust imaging biomarker in the early assessment of lung tumor treatment response.

Study VI The performance of a fully automatic pipeline for image-based lung cancer survival rate prediction was analyzed. The proposed rotationally invariance SphCNN model resulted in competitive predictive power with respect to the state-of-the-art methods when tested on intra-dataset examination and performed slightly better on inter-dataset assessments.

Bibliography

- [1] P. Lambin *et al.*, “Radiomics: Extracting more information from medical images using advanced feature analysis,” *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012, doi: <https://doi.org/10.1016/j.ejca.2011.11.036>.
- [2] K. D. Toennies, *Guide to Medical Image Analysis*. London: Springer London, 2017. doi: <https://doi.org/10.1007/978-1-4471-7320-5>.
- [3] H. Hooda, O. P. Verma, and T. Singhal, “Brain tumor segmentation: A performance analysis using K-Means, Fuzzy C-Means and Region growing algorithm,” *Proceedings of 2014 IEEE International Conference on Advanced Communication, Control and Computing Technologies, ICACCCT 2014*, pp. 1621–1626, Jan. 2015, doi: <https://doi.org/10.1109/ICACCCT.2014.7019383>.
- [4] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen, “Random Walk and Graph Cut for Co-Segmentation of Lung Tumor on PET-CT Images,” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5854–5867, Dec. 2015, doi: <https://doi.org/10.1109/TIP.2015.2488902>.
- [5] C. Ma, G. Luo, and K. Wang, “Concatenated and Connected Random Forests with Multiscale Patch Driven Active Contour Model for Automated Brain Tumor Segmentation of MR Images,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 1943–1954, Aug. 2018, doi: <https://doi.org/10.1109/TMI.2018.2805821>.
- [6] T. X. Pham, P. Siarry, and H. Oulhadj, “Segmentation of MR Brain Images through Hidden Markov Random Field and Hybrid Metaheuristic Algorithm,” *IEEE Transactions on Image Processing*, vol. 29, pp. 6507–6522, 2020, doi: <https://doi.org/10.1109/TIP.2020.2990346>.
- [7] J. Pedrosa *et al.*, “LNDb challenge on automatic lung cancer patient management,” *Medical Image Analysis*, vol. 70, p. 102027, May 2021, doi: <https://doi.org/10.1016/j.media.2021.102027>.
- [8] P. Wang and A. C. S. Chung, “Relax and focus on brain tumor segmentation,” *Medical Image Analysis*, vol. 75, p. 102259, Jan. 2022, doi: <https://doi.org/10.1016/J.MEDIA.2021.102259>.
- [9] J. Zhao *et al.*, “United adversarial learning for liver tumor segmentation and detection of multi-modality non-contrast MRI,” *Medical Image Analysis*, vol. 73, p. 102154, Oct. 2021, doi: <https://doi.org/10.1016/J.MEDIA.2021.102154>.
- [10] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, “Machine Learning methods for Quantitative Radiomic Biomarkers,” *Scientific Reports*, vol. 5, no. 1, p. 13087, Oct. 2015, doi: <https://doi.org/10.1038/srep13087>.
- [11] I. Dregely, D. Prezzi, C. Kelly-Morland, E. Roccia, R. Neji, and V. Goh, “Imaging biomarkers in oncology: Basics and application to MRI,” *Journal of Magnetic Resonance Imaging*, vol. 48, no. 1, pp. 13–26, Jul. 2018, doi: <https://doi.org/10.1002/JMRI.26058>.
- [12] F. Piludu *et al.*, “MRI-Based Radiomics to Differentiate between Benign and Malignant Parotid Tumors With External Validation,” *Frontiers in Oncology*, vol. 11, p. 1, Apr. 2021, doi: <https://doi.org/10.3389/fonc.2021.656918>.
- [13] J. Shi *et al.*, “MRI-based peritumoral radiomics analysis for preoperative prediction of lymph node metastasis in early-stage cervical cancer: A multi-center study,” *Magnetic Resonance Imaging*, vol. 88, pp. 1–8, May 2022, doi: <https://doi.org/10.1016/J.MRI.2021.12.008>.

- [14] M. Amini *et al.*, “Overall Survival Prognostic Modelling of Non-small Cell Lung Cancer Patients Using Positron Emission Tomography/Computed Tomography Harmonised Radiomics Features: The Quest for the Optimal Machine Learning Algorithm,” *Clinical Oncology*, vol. 34, no. 2, pp. 114–127, Feb. 2022, doi: <https://doi.org/10.1016/J.CLON.2021.11.014>.
- [15] A. Conti, A. Duggento, I. Indovina, M. Guerrisi, and N. Toschi, “Radiomics in breast cancer classification and prediction,” *Seminars in Cancer Biology*, vol. 72, pp. 238–250, Jul. 2021, doi: <https://doi.org/10.1016/J.SEMCANCER.2020.04.002>.
- [16] P. Aonpong, Y. Iwamoto, X. H. Han, L. Lin, and Y. W. Chen, “Genotype-Guided Radiomics Signatures for Recurrence Prediction of Non-Small Cell Lung Cancer,” *IEEE Access*, vol. 9, pp. 90244–90254, 2021, doi: <https://doi.org/10.1109/ACCESS.2021.3088234>.
- [17] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241. doi: https://doi.org/10.1007/978-3-319-24574-4_28.
- [18] L. Joskowicz, D. Cohen, N. Caplan, and J. Sosna, “Automatic segmentation variability estimation with segmentation priors,” *Medical Image Analysis*, vol. 50, pp. 54–64, Dec. 2018, doi: <https://doi.org/10.1016/J.MEDIA.2018.08.006>.
- [19] S. Hansen, S. Gautam, R. Jenssen, and M. Kampffmeyer, “Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels,” *Medical Image Analysis*, vol. 78, p. 102385, May 2022, doi: <https://doi.org/10.1016/J.MEDIA.2022.102385>.
- [20] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, “f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks,” *Medical Image Analysis*, vol. 54, pp. 30–44, May 2019, doi: <https://doi.org/10.1016/J.MEDIA.2019.01.010>.
- [21] F. Meissen, G. Kaissis, and D. Rueckert, “Challenging Current Semi-Supervised Anomaly Segmentation Methods for Brain MRI,” Sep. 2021, doi: <https://doi.org/10.48550/arxiv.2109.06023>.
- [22] G. X. Wu and D. J. Raz, *Lung Cancer Screening*, vol. 170. Springer, Cham, 2016. doi: https://doi.org/10.1007/978-3-319-40389-2_1.
- [23] X. Liu, F. Hou, H. Qin, and A. Hao, “Multi-view multi-scale CNNs for lung nodule type classification from CT images,” *Pattern Recognition*, vol. 77, pp. 262–275, May 2018, doi: <https://doi.org/10.1016/J.PATCOG.2017.12.022>.
- [24] S. S. Alahmari, D. Cherezov, D. B. Goldgof, L. O. Hall, R. J. Gillies, and M. B. Schabath, “Delta Radiomics Improves Pulmonary Nodule Malignancy Prediction in Lung Cancer Screening,” *IEEE Access*, vol. 6, pp. 77796–77806, 2018, doi: <https://doi.org/10.1109/ACCESS.2018.2884126>.
- [25] E. L. van Persijn Van Meerten, H. Gelderblom, and J. L. Bloem, “RECIST revised: Implications for the radiologist. A review article on the modified RECIST guideline,” *European Radiology*, vol. 20, no. 6, pp. 1456–1467, Dec. 2010, doi: <https://doi.org/10.1007/S00330-009-1685-Y/FIGURES/7>.
- [26] W. Lu and W. Chen, “Positron emission tomography/computerized tomography for tumor response assessment—a review of clinical practices and radiomics studies,” *Translational Cancer Research*, vol. 5, no. 4, pp. 364–370, 2016, doi: <https://doi.org/10.21037/tcr.2016.07.12>.
- [27] R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, “From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors,” *Journal of Nuclear Medicine*, vol. 50, no. Suppl_1, pp. 122S–150S, 2009, doi: <https://doi.org/10.2967/jnumed.108.057307>.

- [28] G. Lee *et al.*, "Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art," *European Journal of Radiology*, vol. 86, pp. 297–307, 2017, doi: <https://doi.org/10.1016/j.ejrad.2016.09.005>.
- [29] D. R. Cox, "Regression Models and Life-Tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, Jan. 1972, doi: <https://doi.org/10.1111/J.2517-6161.1972.TB00899.X>.
- [30] S. Leger *et al.*, "A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling," *Scientific Reports* 2017 7:1, vol. 7, no. 1, pp. 1–11, Oct. 2017, doi: <https://doi.org/10.1038/s41598-017-13448-3>.
- [31] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021, doi: <https://doi.org/10.3322/CAAC.21660>.
- [32] L. Rahib, M. Wehner, L. M. Matrisian, and K. T. Nead, "Projection of cancer incidence and death to 2040 in the US: Impact of cancer screening and a changing demographic.," https://doi.org/10.1200/JCO.2020.38.15_suppl.1566, vol. 38, no. 15_suppl, pp. 1566–1566, May 2020, doi: https://doi.org/10.1200/JCO.2020.38.15_SUPPL.1566.
- [33] "European Cancer Information System," 2020. <https://ecis.jrc.ec.europa.eu/>
- [34] N. Howlader, A. M. Noone, and Krapcho. M., "SEER cancer statistics review, 1975–2012," Bethesda, MD. [Online]. Available: http://seer.cancer.gov/csr/1975_2012/
- [35] D. Hanahan and R. A. Weinberg, "The Hallmarks of Cancer," *Cell*, vol. 100, no. 1, pp. 57–70, Jan. 2000, doi: [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
- [36] D. Hanahan and R. A. Weinberg, "Hallmarks of Cancer: The Next Generation," *Cell*, vol. 144, no. 5, pp. 646–674, Mar. 2011, doi: <https://doi.org/10.1016/J.CELL.2011.02.013>.
- [37] Y. A. Fouad and C. Aanei, "Revisiting the hallmarks of cancer.," *Am J Cancer Res*, vol. 7, no. 5, pp. 1016–1036, 2017.
- [38] D. Hanahan, "Hallmarks of Cancer: New Dimensions," *Cancer Discovery*, vol. 12, no. 1, pp. 31–46, Jan. 2022, doi: <https://doi.org/10.1158/2159-8290.CD-21-1059>.
- [39] Y. Assenov, D. Brocks, and C. Gerhäuser, "Intratumor heterogeneity in epigenetic patterns," *Seminars in Cancer Biology*, vol. 51, pp. 12–21, Aug. 2018, doi: <https://doi.org/10.1016/J.SEMCANCER.2018.01.010>.
- [40] A. Matak *et al.*, "Stochastic phenotype switching leads to intratumor heterogeneity in human liver cancer," *Hepatology*, vol. 68, no. 3, pp. 933–948, Sep. 2018, doi: <https://doi.org/10.1002/hep.29679>.
- [41] H. N. Wagner and P. S. Conti, "Advances in Medical Imaging for Cancer Diagnosis and Treatment," *Cancer*, vol. 67, no. S4, pp. 1121–1128, 1991, doi: <https://doi.org/10.1002/1097-0142>.
- [42] F. Diekmann and U. Bick, "Tomosynthesis and contrast-enhanced digital mammography: Recent advances in digital mammography," *European Radiology*, vol. 17, no. 12, pp. 3086–3092, Dec. 2007, doi: <https://doi.org/10.1007/S00330-007-0715-X/FIGURES/6>.
- [43] E. Lin and A. Alessio, "What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT?," *J Cardiovasc Comput Tomogr*, vol. 3, no. 6, p. 403, Nov. 2009, doi: <https://doi.org/10.1016/J.JCCT.2009.07.003>.

- [44] S. J. Swensen, "Functional CT: Lung nodule evaluation," *Radiographics*, vol. 20, no. 4, pp. 1178–1181, Jul. 2000, doi: <https://doi.org/10.1148/radiographics.20.4.g00jl221178>.
- [45] L. Fass, "Imaging and cancer: A review," 2008, doi: <https://doi.org/10.1016/j.molonc.2008.04.001>.
- [46] M. O. Leach, "Application of magnetic resonance imaging to angiogenesis in breast cancer," *Breast Cancer Research*, vol. 3, no. 1, pp. 22–27, Dec. 2001, doi: <https://doi.org/10.1186/BCR266/FIGURES/1>.
- [47] T. L. Chenevert *et al.*, "Diffusion Magnetic Resonance Imaging: an Early Surrogate Marker of Therapeutic Efficacy in Brain Tumors," *JNCI: Journal of the National Cancer Institute*, vol. 92, no. 24, pp. 2029–2036, Dec. 2000, doi: <https://doi.org/10.1093/JNCI/92.24.2029>.
- [48] M. C. Mabray, R. F. Barajas, and S. Cha, "Modern Brain Tumor Imaging," *Brain Tumor Research and Treatment*, vol. 3, no. 1, pp. 8–23, Apr. 2015, doi: <https://doi.org/10.14791/BTRT.2015.3.1.8>.
- [49] J. A. Thie, "Understanding the Standardized Uptake Value, Its Methods, and Implications for Usage," *Journal of Nuclear Medicine*, vol. 45, no. 9, 2004.
- [50] M. A. Lodge, M. A. Chaudhry, and R. L. Wahl, "Noise Considerations for PET Quantification Using Maximum and Peak Standardized Uptake Value," *Journal of Nuclear Medicine*, vol. 53, no. 7, pp. 1041–1047, Jul. 2012, doi: <https://doi.org/10.2967/JNUMED.111.101733>.
- [51] M. D. Farwell, D. A. Pryma, and D. A. Mankoff, "PET/CT imaging in cancer: Current applications and future directions," *Cancer*, vol. 120, no. 22, pp. 3433–3445, Nov. 2014, doi: <https://doi.org/10.1002/CNCR.28860>.
- [52] P. S. T. Shanmugam, *Understanding Cancer Therapies*. CRC press, 2018.
- [53] MayoClinic, "Cancer treatment methods," 2019. <https://www.mayoclinic.org/tests-procedures/cancer-treatment/about/pac-20393344>
- [54] A. Sudhakar, "History of Cancer, Ancient and Modern Treatment Methods," *J Cancer Sci Ther*, vol. 1, no. 2, pp. i–iv, 2009, doi: <https://doi.org/10.4172/1948-5956.100000E2>.
- [55] P. Therasse *et al.*, "New Guidelines to Evaluate the Response to Treatment in Solid Tumors," *JNCI: Journal of the National Cancer Institute*, vol. 92, no. 3, pp. 205–216, Feb. 2000, doi: <https://doi.org/10.1093/jnci/92.3.205>.
- [56] E. A. Eisenhauer *et al.*, "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)," *European Journal of Cancer*, vol. 45, no. 2, pp. 228–247, Jan. 2009, doi: <https://doi.org/10.1016/j.ejca.2008.10.026>.
- [57] L. H. Schwartz *et al.*, "RECIST 1.1 – Standardisation and disease-specific adaptations: Perspectives from the RECIST Working Group," *European Journal of Cancer*, vol. 62, pp. 138–145, Jul. 2016.
- [58] S. Litière, S. Collette, E. G. E. de Vries, L. Seymour, and J. Bogaerts, "RECIST — learning from the past to build the future," *Nature Reviews Clinical Oncology*, vol. 14, no. 3, pp. 187–192, Mar. 2017, doi: <https://doi.org/10.1038/nrclinonc.2016.195>.
- [59] L. Seymour *et al.*, "iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics," *The Lancet Oncology*, vol. 18, no. 3, pp. e143–e152, Mar. 2017, doi: [https://doi.org/10.1016/S1470-2045\(17\)30074-8](https://doi.org/10.1016/S1470-2045(17)30074-8).
- [60] A. Gouliamos, J. Andreou, and P. Kosmidis, *Imaging in Clinical Oncology*, Second. Springer, 2014.

- [61] R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, "From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors," *Journal of Nuclear Medicine*, vol. 50, no. Suppl_1, pp. 122S-150S, 2009, doi: <https://doi.org/10.2967/jnumed.108.057307>.
- [62] M. J. Joyner and N. Paneth, "Promises, promises, and precision medicine," *The Journal of Clinical Investigation*, vol. 129, no. 3, pp. 946–948, Mar. 2019, doi: <https://doi.org/10.1172/JCI126119>.
- [63] A. Giardino *et al.*, "Role of Imaging in the Era of Precision Medicine," *Academic Radiology*, vol. 24, no. 5, pp. 639–649, May 2017, doi: <https://doi.org/10.1016/J.ACRA.2016.11.021>.
- [64] T. Mokry *et al.*, "Accuracy of Estimation of Graft Size for Living-Related Liver Transplantation: First Results of a Semi-Automated Interactive Software for CT-Volumetry," *PLoS ONE*, vol. 9, no. 10, p. e110201, Oct. 2014, doi: <https://doi.org/10.1371/journal.pone.0110201>.
- [65] D. Prezzi, A. Khan, and V. Goh, "Perfusion CT imaging of treatment response in oncology," *European Journal of Radiology*, vol. 84, no. 12, pp. 2380–2385, Dec. 2015, doi: <https://doi.org/10.1016/j.ejrad.2015.03.022>.
- [66] W. M. Thaiss, A. W. Sauter, M. Bongers, M. Horger, and K. Nikolaou, "Clinical applications for dual energy CT versus dynamic contrast enhanced CT in oncology," *European Journal of Radiology*, vol. 84, no. 12, pp. 2368–2379, Dec. 2015, doi: <https://doi.org/10.1016/j.ejrad.2015.06.001>.
- [67] G. Buizza *et al.*, "Early Tumor Response Prediction for Lung Cancer Patients Using Novel Longitudinal Pattern Features from Sequential PET/CT Image Scans," *Physica Medica*, vol. 54, pp. 21–29, Oct. 2018, doi: <https://doi.org/10.1016/J.EJMP.2018.09.003>.
- [68] L. (Louis) Marti-Bonmati, A. Alberich-Bayarri, L. Martí-Bonmatí, and A. Alberich-Bayarri, *Imaging Biomarkers Development and Clinical Integration*, First. Springer, 2017.
- [69] P. D. Chang, D. S. Chow, P. H. Yang, C. G. Filippi, and A. Lignelli, "Predicting Glioblastoma Recurrence by Early Changes in the Apparent Diffusion Coefficient Value and Signal Intensity on FLAIR Images," *American Journal of Roentgenology*, vol. 208, no. 1, pp. 57–65, Jan. 2017, doi: <https://doi.org/10.2214/AJR.16.16234>.
- [70] H. Akbari *et al.*, "Imaging Surrogates of Infiltration Obtained Via Multiparametric Imaging Pattern Analysis Predict Subsequent Location of Recurrence of Glioblastoma," *Neurosurgery*, vol. 78, no. 4, pp. 572–580, Apr. 2016, doi: <https://doi.org/10.1227/NEU.0000000000001202>.
- [71] J. C. Peeken *et al.*, "Deep learning derived tumor infiltration maps for personalized target definition in Glioblastoma radiotherapy," *Radiotherapy and Oncology*, vol. 138, pp. 166–172, Sep. 2019, doi: <https://doi.org/10.1016/J.RADONC.2019.06.031>.
- [72] J. Larry Jameson and D. L. Longo, "Precision Medicine—Personalized, Problematic, and Promising," *Obstetrical & Gynecological Survey*, vol. 70, no. 10, pp. 612–614, Oct. 2015, doi: <https://doi.org/10.1097/01.ogx.0000472121.21647.38>.
- [73] C. N. A. M. Oldenhuis, S. F. Oosting, J. A. Gietema, and E. G. E. de Vries, "Prognostic versus predictive value of biomarkers in oncology," *European Journal of Cancer*, vol. 44, no. 7, pp. 946–953, May 2008, doi: <https://doi.org/10.1016/J.EJCA.2008.03.006>.
- [74] M. Zhou *et al.*, "Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches," *American Journal of Neuroradiology*, vol. 39, no. 2, pp. 208–216, Feb. 2018, doi: <https://doi.org/10.3174/AJNR.A5391>.
- [75] J. R. Ferreira Junior, M. Koenigkam-Santos, F. E. G. Cipriano, A. T. Fabro, and P. M. de Azevedo-Marques, "Radiomics-based features for pattern recognition of lung cancer histopathology and

- metastases,” *Computer Methods and Programs in Biomedicine*, vol. 159, pp. 23–30, Jun. 2018, doi: <https://doi.org/10.1016/J.CMPB.2018.02.015>.
- [76] Y. Huang *et al.*, “Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non—Small Cell Lung Cancer,” *Radiology*, vol. 281, no. 3, pp. 947–957, Dec. 2016, doi: <https://doi.org/10.1148/radiol.2016152234>.
 - [77] R. J. Gillies, P. E. Kinahan, and H. Hricak, “Radiomics: Images Are More than Pictures, They Are Data,” *Radiology*, vol. 278, no. 2, pp. 563–577, 2016, doi: <https://doi.org/10.1148/radiol.2015151169>.
 - [78] A. Saha, X. Yu, D. Sahoo, and M. A. Mazurowski, “Effects of MRI scanner parameters on breast cancer radiomics,” *Expert Systems with Applications*, vol. 87, pp. 384–391, 2017, doi: <https://doi.org/10.1016/j.eswa.2017.06.029>.
 - [79] C. Shen *et al.*, “2D and 3D CT Radiomics Features Prognostic Performance Comparison in Non-Small Cell Lung Cancer,” *Translational Oncology*, vol. 10, no. 6, pp. 886–894, Dec. 2017, doi: <https://doi.org/10.1016/J.TRANON.2017.08.007>.
 - [80] M. F. McNitt-Gray, E. M. Hart, N. Wyckoff, J. W. Sayre, J. G. Goldin, and D. R. Aberle, “A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results,” *Medical Physics*, vol. 26, no. 6, pp. 880–888, Jun. 1999, doi: <https://doi.org/10.1118/1.598603>.
 - [81] F. Giesel *et al.*, “CT radiomic analysis using lymph-node-density profile in correlation to SUV-value for PET/CT based N-Staging,” *Journal of Nuclear Medicine*, p. jnumed.116.179648, Sep. 2016, doi: <https://doi.org/10.2967/JNUMED.116.179648>.
 - [82] C. A. Karlo *et al.*, “Radiogenomics of Clear Cell Renal Cell Carcinoma: Associations between CT Imaging Features and Mutations,” *Radiology*, vol. 270, no. 2, pp. 464–471, Feb. 2014, doi: <https://doi.org/10.1148/radiol.13130663>.
 - [83] V. K. Dhingra, A. Mahajan, and S. Basu, “Emerging clinical applications of PET based molecular imaging in oncology: the promising future potential for evolving personalized cancer care.,” *Indian J Radiol Imaging*, vol. 25, no. 4, pp. 332–41, 2015, doi: <https://doi.org/10.4103/0971-3026.169467>.
 - [84] M. Lazzeroni *et al.*, “Evaluation of third treatment week as temporal window for assessing responsiveness on repeated FDG-PET-CT scans in Non-Small Cell Lung Cancer patients,” *Physica Medica*, vol. 46, pp. 45–51, Feb. 2018, doi: <https://doi.org/10.1016/J.EJMP.2018.01.012>.
 - [85] P. W. Grigsby, B. A. Siegel, F. Dehdashti, J. Rader, and I. Zoberi, “Posttherapy [¹⁸F] Fluorodeoxyglucose Positron Emission Tomography in Carcinoma of the Cervix: Response and Outcome,” *Journal of Clinical Oncology*, vol. 22, no. 11, pp. 2167–2171, Jun. 2004, doi: <https://doi.org/10.1200/JCO.2004.09.035>.
 - [86] A. Cameron, F. Khalvati, M. A. Haider, and A. Wong, “MAPS: A Quantitative Radiomics Approach for Prostate Cancer Detection,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 6, pp. 1145–1156, Jun. 2016, doi: <https://doi.org/10.1109/TBME.2015.2485779>.
 - [87] A. Kotrotsou, P. O. Zinn, and R. R. Colen, “Radiomics in Brain Tumors: An Emerging Technique for Characterization of Tumor Environment,” *Magnetic Resonance Imaging Clinics of North America*, vol. 24, no. 4, pp. 719–729, Nov. 2016, doi: <https://doi.org/10.1016/j.mric.2016.06.006>.
 - [88] J. Wang *et al.*, “Identifying Triple-Negative Breast Cancer Using Background Parenchymal Enhancement Heterogeneity on Dynamic Contrast-Enhanced MRI: A Pilot Radiomics Study,” *PLOS ONE*, vol. 10, no. 11, p. e0143308, Nov. 2015, doi: <https://doi.org/10.1371/journal.pone.0143308>.

- [89] M. L. Giger *et al.*, “Radiomics for ultrafast dynamic contrast-enhanced breast MRI in the diagnosis of breast cancer: a pilot study,” in *Medical Imaging 2018: Computer-Aided Diagnosis*, Feb. 2018, vol. 10575, p. 138. doi: <https://doi.org/10.1117/12.2293644>.
- [90] S. D. McGarry *et al.*, “Magnetic Resonance Imaging-Based Radiomic Profiles Predict Patient Prognosis in Newly Diagnosed Glioblastoma Before Therapy,” *Tomography*, vol. 2, no. 3, pp. 223–228, Sep. 2016, doi: <https://doi.org/10.18383/j.tom.2016.00250>.
- [91] X. Feng, N. Tustison, and C. Meyer, “Brain Tumor Segmentation Using an Ensemble of 3D U-Nets and Overall Survival Prediction Using Radiomic Features,” Springer, Cham, 2019, pp. 279–288. doi: https://doi.org/10.1007/978-3-030-11726-9_25.
- [92] A. Valsecchi, E. Bermejo, S. Damas, and O. Cordón, “Metaheuristics for Medical Image Registration,” in *Handbook of Heuristics*, Cham: Springer International Publishing, 2018, pp. 1079–1101. doi: https://doi.org/10.1007/978-3-319-07124-4_56.
- [93] A. Oikonomou *et al.*, “Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy,” *Scientific Reports*, vol. 8, no. 1, p. 4003, Dec. 2018, doi: <https://doi.org/10.1038/s41598-018-22357-y>.
- [94] M. Kirienko *et al.*, “Ability of FDG PET and CT radiomics features to differentiate between primary and metastatic lung lesions,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 45, no. 10, pp. 1649–1660, Sep. 2018, doi: <https://doi.org/10.1007/s00259-018-3987-2>.
- [95] S. Bakas *et al.*, “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features,” *Scientific Data*, vol. 4, no. 1, p. 170117, Dec. 2017, doi: <https://doi.org/10.1038/sdata.2017.117>.
- [96] T. P. Coroller *et al.*, “Radiomic phenotype features predict pathological response in non-small cell lung cancer,” *Radiotherapy and Oncology*, vol. 119, no. 3, pp. 480–486, Jun. 2016, doi: <https://doi.org/10.1016/j.radonc.2016.04.004>.
- [97] E. Huynh *et al.*, “CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer,” *Radiotherapy and Oncology*, vol. 120, no. 2, pp. 258–266, Aug. 2016, doi: <https://doi.org/10.1016/j.radonc.2016.05.024>.
- [98] M. Elter and A. Horsch, “CADx of mammographic masses and clustered microcalcifications: A review,” *Medical Physics*, vol. 36, no. 6Part1, pp. 2052–2068, May 2009, doi: <https://doi.org/10.1118/1.3121511>.
- [99] Y. Tan, L. H. Schwartz, and B. Zhao, “Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field,” *Medical Physics*, vol. 40, no. 4, p. 043502, Mar. 2013, doi: <https://doi.org/10.1118/1.4793409>.
- [100] H. Zaidi and I. El Naqa, “PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 11, pp. 2165–2187, Nov. 2010, doi: <https://doi.org/10.1007/s00259-010-1423-3>.
- [101] H. Li *et al.*, “MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays,” *Radiology*, vol. 281, no. 2, pp. 382–391, Nov. 2016, doi: <https://doi.org/10.1148/radiol.2016152110>.
- [102] B. H. Menze *et al.*, “Fully Automatic Brain Tumor Segmentation from Multiple MR Sequences using Hidden Markov Fields and Variational.” 2013.

- [103] M. Winkels and T. S. Cohen, "Pulmonary nodule detection in CT scans with equivariant CNNs," *Medical Image Analysis*, vol. 55, pp. 15–26, Jul. 2019, doi: <https://doi.org/10.1016/J.MEDIA.2019.03.010>.
- [104] V. Oreiller *et al.*, "Head and neck tumor segmentation in PET/CT: The HECKTOR challenge," *Medical Image Analysis*, vol. 77, p. 102336, Apr. 2022, doi: <https://doi.org/10.1016/J.MEDIA.2021.102336>.
- [105] A. Duran, G. Dussert, O. Rouvière, T. Jaouen, P. M. Jodoin, and C. Lartizien, "ProstAttention-Net: A deep attention model for prostate cancer segmentation by aggressiveness in MRI scans," *Medical Image Analysis*, vol. 77, p. 102347, Apr. 2022, doi: <https://doi.org/10.1016/J.MEDIA.2021.102347>.
- [106] Y. Balagurunathan *et al.*, "Test–Retest Reproducibility Analysis of Lung CT Image Features," *Journal of Digital Imaging*, vol. 27, no. 6, pp. 805–823, 2014, doi: <https://doi.org/10.1007/s10278-014-9716-x>.
- [107] B. Zhao *et al.*, "Reproducibility of radiomics for deciphering tumor phenotype with imaging," *Scientific Reports*, vol. 6, pp. 1–7, 2016, doi: <https://doi.org/10.1038/srep23428>.
- [108] R. T. H. M. Larue *et al.*, "4DCT imaging to assess radiomics feature stability: An investigation for thoracic cancers," *Radiotherapy and Oncology*, vol. 125, no. 1, pp. 147–153, 2017, doi: <https://doi.org/10.1016/j.radonc.2017.07.023>.
- [109] C. J. Herold *et al.*, "Imaging in the Age of Precision Medicine: Summary of the Proceedings of the 10th Biannual Symposium of the International Society for Strategic Studies in Radiology," *Radiology*, vol. 279, no. 1, pp. 226–238, Apr. 2016, doi: <https://doi.org/10.1148/radiol.2015150709>.
- [110] R. J. Gillies, A. R. Anderson, R. A. Gatenby, and D. L. Morse, "The biology underlying molecular imaging in oncology: from genome to anatome and back again," *Clinical Radiology*, vol. 65, no. 7, pp. 517–521, Jul. 2010, doi: <https://doi.org/10.1016/J.CRAD.2010.04.005>.
- [111] H. J. W. L. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, 2014, doi: <https://doi.org/10.1038/ncomms5006>.
- [112] I. El Naqa, *A guide to outcome modeling in radiotherapy and oncology : listening to the data*, Series in. CRC press, 2018.
- [113] A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, "Image biomarker standardisation initiative," *arXiv1612.07003*, vol. 11, Dec. 2019.
- [114] M. Avanzo, J. Stancanella, and I. El Naqa, "Beyond imaging: The promise of radiomics," *Physica Medica*, vol. 38, pp. 122–139, 2017, doi: <https://doi.org/10.1016/j.ejmp.2017.05.071>.
- [115] B. Ganeshan, S. Abaleke, R. C. D. Young, C. R. Chatwin, and K. A. Miles, "Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage," *Cancer Imaging*, vol. 10, no. 1, p. 137, 2010, doi: <https://doi.org/10.1102/1470-7330.2010.0021>.
- [116] W. Chen, M. L. Giger, L. Lan, and U. Bick, "Computerized interpretation of breast MRI: Investigation of enhancement-variance dynamics," *Medical Physics*, vol. 31, no. 5, pp. 1076–1082, Apr. 2004, doi: <https://doi.org/10.1118/1.1695652>.
- [117] P.-P. Ypsilantis *et al.*, "Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks," *PLOS ONE*, vol. 10, no. 9, p. e0137036, Sep. 2015, doi: <https://doi.org/10.1371/journal.pone.0137036>.

- [118] U. Bashir, M. M. Siddique, E. Mclean, V. Goh, and G. J. Cook, "Imaging Heterogeneity in Lung Cancer: Techniques, Applications, and Challenges," *American Journal of Roentgenology*, vol. 207, no. 3, pp. 534–543, Sep. 2016, doi: <https://doi.org/10.2214/AJR.15.15864>.
- [119] A. Zwanenburg *et al.*, "The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, May 2020, doi: <https://doi.org/10.1148/radiol.2020191145>
- [120] H. J. W. L. Aerts, "The Potential of Radiomic-Based Phenotyping in Precision Medicine," *JAMA Oncology*, vol. 2, no. 12, p. 1636, Dec. 2016, doi: <https://doi.org/10.1001/jamaoncol.2016.2631>.
- [121] V. P. Parmar, A. K. Pandya, and C. K. Kumbharana, "Determining the character replacement rules and implementing them for phonetic identification of given words to identify similar pronunciation words," in *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, Feb. 2015, pp. 272–277. doi: <https://doi.org/10.1109/ABLAZE.2015.7155010>.
- [122] Y. Liu *et al.*, "Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas," *Clinical Lung Cancer*, vol. 17, no. 5, pp. 441–448.e6, Sep. 2016, doi: <https://doi.org/10.1016/j.clcc.2016.02.001>.
- [123] P. Kickingeder *et al.*, "Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models," *Radiology*, vol. 280, no. 3, pp. 880–889, Sep. 2016, doi: <https://doi.org/10.1148/radiol.2016160845>.
- [124] V. Parekh and M. A. Jacobs, "Radiomics: a new application from established techniques," *Expert Review of Precision Medicine and Drug Development*, vol. 1, no. 2, pp. 207–226, Mar. 2016, doi: <https://doi.org/10.1080/23808993.2016.1164013>.
- [125] R. R. Colen *et al.*, "Radiomics to predict immunotherapy-induced pneumonitis: proof of concept," *Investigational New Drugs*, vol. 36, no. 4, pp. 601–607, Aug. 2018, doi: <https://doi.org/10.1007/s10637-017-0524-2>.
- [126] L. E. L. Lu, L. Li, H. Yang, L. H. Schwartz, and B. Zhao, "Radiomics for Classification of Lung Cancer Histological Subtypes Based on Nonenhanced Computed Tomography," *Academic Radiology*, vol. 26, no. 9, pp. 1245–1252, Sep. 2019, doi: <https://doi.org/10.1016/J.ACRA.2018.10.013>.
- [127] M. Bogowicz, S. Tanadini-Lang, M. Guckenberger, and O. Riesterer, "Combined CT radiomics of primary tumor and metastatic lymph nodes improves prediction of loco-regional control in head and neck cancer," *Scientific Reports*, vol. 9, no. 1, p. 15198, Dec. 2019, doi: <https://doi.org/10.1038/s41598-019-51599-7>.
- [128] H. Cho, S. Lee, J. Kim, and H. Park, "Classification of the glioma grading using radiomics analysis," *PeerJ*, vol. 6, p. e5982, Nov. 2018, doi: <https://doi.org/10.7717/peerj.5982>.
- [129] W. Guo *et al.*, "Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data," *Journal of Medical Imaging*, vol. 2, no. 4, p. 041007, Sep. 2015, doi: <https://doi.org/10.1117/1.JMI.2.4.041007>.
- [130] B. Zhang *et al.*, "Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma," *Clinical Cancer Research*, vol. 23, no. 15, pp. 4259–4269, Aug. 2017, doi: <https://doi.org/10.1158/1078-0432.CCR-16-2910>.
- [131] C. Lian, S. Ruan, T. Denœux, F. Jardin, and P. Vera, "Selecting radiomic features from FDG-PET images for cancer treatment outcome prediction," *Medical Image Analysis*, vol. 32, pp. 257–268, 2016, doi: <https://doi.org/10.1016/j.media.2016.05.007>.

- [132] P. Yin *et al.*, "Comparison of radiomics machine-learning classifiers and feature selection for differentiation of sacral chordoma and sacral giant cell tumour based on 3D computed tomography features," *European Radiology*, vol. 29, no. 4, pp. 1841–1847, Apr. 2019, doi: <https://doi.org/10.1007/s00330-018-5730-6>.
- [133] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: <https://doi.org/10.1016/j.COMPELECENG.2013.11.024>.
- [134] B. Zhang *et al.*, "Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma," *Cancer Letters*, vol. 403, pp. 21–27, 2017, doi: <https://doi.org/10.1016/j.canlet.2017.06.004>.
- [135] H. Cho and H. Park, "Classification of low-grade and high-grade glioma using multi-modal image radiomics features," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2017, pp. 3081–3084. doi: <https://doi.org/10.1109/EMBC.2017.8037508>.
- [136] C. Parmar *et al.*, "Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer," *Scientific Reports*, vol. 5, no. 1, p. 11044, Sep. 2015, doi: <https://doi.org/10.1038/srep11044>.
- [137] X. Zhu *et al.*, "Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer," *European Radiology*, vol. 28, no. 7, pp. 2772–2778, Jul. 2018, doi: <https://doi.org/10.1007/s00330-017-5221-1>.
- [138] X. Zhang *et al.*, "Radiomics assessment of bladder cancer grade using texture features from diffusion-weighted imaging," *Journal of Magnetic Resonance Imaging*, vol. 46, no. 5, pp. 1281–1288, Nov. 2017, doi: <https://doi.org/10.1002/jmri.25669>.
- [139] C. C. Ko *et al.*, "Pre-operative MRI Radiomics for the Prediction of Progression and Recurrence in Meningiomas," *Frontiers in Neurology*, vol. 12, p. 599, May 2021, doi: <https://doi.org/10.3389/fneur.2021.636235>.
- [140] C. Liang *et al.*, "The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer," *Oncotarget*, vol. 7, no. 21, May 2016, doi: <https://doi.org/10.18632/oncotarget.8919>.
- [141] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge," Springer, Cham, 2018, pp. 287–297. doi: https://doi.org/10.1007/978-3-319-75238-9_25.
- [142] M. Ingrisch *et al.*, "Radiomic Analysis Reveals Prognostic Information in T1-Weighted Baseline Magnetic Resonance Imaging in Patients With Glioblastoma," *Investigative Radiology*, vol. 52, no. 6, pp. 360–366, Jun. 2017, doi: <https://doi.org/10.1097/RLI.0000000000000349>.
- [143] S. A. Mattonen *et al.*, "Detection of Local Cancer Recurrence After Stereotactic Ablative Radiation Therapy for Lung Cancer: Physician Performance Versus Radiomic Assessment," *International Journal of Radiation Oncology*Biophysics*Physics*, vol. 94, no. 5, pp. 1121–1128, Apr. 2016, doi: <https://doi.org/10.1016/j.IJROBP.2015.12.369>.
- [144] N. Emaminejad *et al.*, "Fusion of Quantitative Image and Genomic Biomarkers to Improve Prognosis Assessment of Early Stage Lung Cancer Patients," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 5, pp. 1034–1043, 2016, doi: <https://doi.org/10.1109/TBME.2015.2477688>.

- [145] N. M. Braman *et al.*, "Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI," *Breast Cancer Research*, vol. 19, no. 1, p. 57, Dec. 2017, doi: <https://doi.org/10.1186/s13058-017-0846-1>.
- [146] R. Nakanishi *et al.*, "Radiomics Texture Analysis for the Identification of Colorectal Liver Metastases Sensitive to First-Line Oxaliplatin-Based Chemotherapy," *Annals of Surgical Oncology*, vol. 28, no. 6, pp. 2975–2985, Jun. 2021, doi: <https://doi.org/10.1245/S10434-020-09581-5/FIGURES/4>.
- [147] J. Wu *et al.*, "Early-Stage Non–Small Cell Lung Cancer: Quantitative Imaging Characteristics of ¹⁸F Fluorodeoxyglucose PET/CT Allow Prediction of Distant Metastasis," *Radiology*, vol. 281, no. 1, pp. 270–278, Oct. 2016, doi: <https://doi.org/10.1148/radiol.2016151829>.
- [148] X. Bai *et al.*, "MRI radiomics-based nomogram for individualised prediction of synchronous distant metastasis in patients with clear cell renal cell carcinoma," *European Radiology*, vol. 31, no. 2, pp. 1029–1042, Feb. 2021, doi: <https://doi.org/10.1007/s00330-020-07184-y>.
- [149] S. Wu *et al.*, "A Radiomics Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer," *Clinical Cancer Research*, vol. 23, no. 22, pp. 6904–6911, Nov. 2017, doi: <https://doi.org/10.1158/1078-0432.CCR-17-1510>.
- [150] A. Laghi and C. Voena, "CT-based Radiomics for Biliary Tract Cancer: A Possible Solution for Predicting Lymph Node Metastases," *Radiology*, vol. 290, no. 1, pp. 99–100, Jan. 2019, doi: <https://doi.org/10.1148/radiol.2018182158>.
- [151] T. P. Coroller *et al.*, "Radiomic-Based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC," *Journal of Thoracic Oncology*, vol. 12, no. 3, pp. 467–476, 2017, doi: <https://doi.org/10.1016/j.jtho.2016.11.2226>.
- [152] L. Wei *et al.*, "Variational Autoencoder Graph-based Radiomics Outcome Modeling of Intrahepatic Progression Risk and Overall Survival for HCC post-SBRT Patients," *International Journal of Radiation Oncology*Biophysics*, vol. 105, no. 1, pp. S83–S84, Sep. 2019, doi: <https://doi.org/10.1016/j.ijrobp.2019.06.550>.
- [153] O. Morin *et al.*, "Integrated models incorporating radiologic and radiomic features predict meningioma grade, local failure, and overall survival," *Neuro-Oncology Advances*, vol. 1, no. 1, May 2019, doi: <https://doi.org/10.1093/oaajnl/vdz011>.
- [154] W. Wu *et al.*, "Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology," *Frontiers in Oncology*, vol. 6, p. 71, Mar. 2016, doi: <https://doi.org/10.3389/fonc.2016.00071>.
- [155] Y. Zhu *et al.*, "Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma," *Scientific Reports*, vol. 5, no. 1, p. 17787, Dec. 2015, doi: <https://doi.org/10.1038/srep17787>.
- [156] M. Sollini, L. Cozzi, L. Antunovic, A. Chiti, and M. Kirienko, "PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology," *Scientific Reports*, vol. 7, no. 1, p. 358, Dec. 2017, doi: <https://doi.org/10.1038/s41598-017-00426-y>.
- [157] F. Binczyk, W. Prazuch, P. Bozek, and J. Polanska, "Radiomics and artificial intelligence in lung cancer screening," *Translational Lung Cancer Research*, vol. 10, no. 2, pp. 1186–1199, Feb. 2021, doi: <https://doi.org/10.21037/TLCR-20-708>.
- [158] S. Gore, T. Chougule, J. Jagtap, J. Saini, and M. Ingalthalikar, "A Review of Radiomics and Deep Predictive Modeling in Glioma Characterization," *Academic Radiology*, vol. 28, no. 11, pp. 1599–1621, Nov. 2021, doi: <https://doi.org/10.1016/J.ACRA.2020.06.016>.

- [159] M. Schmier *et al.*, "Repeatability of Multiparametric Prostate MRI Radiomics Features," *Scientific Reports*, vol. 9, no. 1, p. 9441, Dec. 2019, doi: <https://doi.org/10.1038/s41598-019-45766-z>.
- [160] S. K. B. Spohn *et al.*, "Radiomics in prostate cancer imaging for a personalized treatment approach - current aspects of methodology and a systematic review on validated studies," *Theranostics*, vol. 11, no. 16, pp. 8027–8042, 2021, doi: <https://doi.org/10.7150/THNO.61207>.
- [161] S. Rathore *et al.*, "Technical note: a radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma," in *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, Mar. 2018, vol. 10576, p. 110. doi: <https://doi.org/10.1117/12.2323331>.
- [162] P. Giraud *et al.*, "Radiomics and Machine Learning for Radiotherapy in Head and Neck Cancers," *Frontiers in Oncology*, vol. 9, p. 174, Mar. 2019, doi: <https://doi.org/10.3389/fonc.2019.00174>.
- [163] G. Bruixola *et al.*, "Radiomics and radiogenomics in head and neck squamous cell carcinoma: Potential contribution to patient management and challenges," *Cancer Treatment Reviews*, vol. 99, p. 102263, Sep. 2021, doi: <https://doi.org/10.1016/j.ctrv.2021.102263>.
- [164] M. A. Hofman, "Evolution of the human brain: When bigger is better," *Frontiers in Neuroanatomy*, vol. 8, no. MAR, p. 15, Mar. 2014, doi: <https://doi.org/10.3389/fnana.2014.00015>.
- [165] H. Greenside, "Theoretical Nonequilibrium Physics." <https://webhome.phy.duke.edu/~hsg/414/images/brain-vs-computer.html>
- [166] "Brain Neurons & Synapses | Action Potentials & Neurotransmission." <https://human-memory.net/brain-neurons-synapses/> (accessed Dec. 09, 2019).
- [167] S. K. Zhou, H. Greenspan, and D. Shen, *Deep learning for medical image analysis*, Elsevier. Academic Press, 2017.
- [168] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, Dec. 1989, doi: <https://doi.org/10.1007/BF02551274>.
- [169] F. Chollet, *Deep learning with Python*. Manning Publications, 2017.
- [170] S. K. Zhou, H. Greenspan, and D. Shen, *Deep learning for medical image analysis*, Elsevier. Academic Press, 2017. [Online]. Available: <https://www.sciencedirect.com/book/9780128104088/deep-learning-for-medical-image-analysis>
- [171] F. Chollet, *Deep learning with Python*. Manning Publications, 2017. [Online]. Available: <https://www.manning.com/books/deep-learning-with-python>
- [172] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2016. [Online]. Available: <https://mitpress.mit.edu/books/deep-learning>
- [173] D. Liu, "A Practical Guide to ReLU - Danqing Liu - Medium." <https://medium.com/@danqing/a-practical-guide-to-relu-b83ca804f1f7> (accessed Dec. 09, 2019).
- [174] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," *arXiv:1511.07289*, Nov. 2016.
- [175] W. Shang, K. Sohn, D. Almeida, and H. Lee, "Understanding and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units," *arXiv:1603.05201*, Mar. 2016.
- [176] A. Krizhevsky and A. Krizhevsky, "Convolutional deep belief networks on cifar-10," 2010.

- [177] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning (Adaptive Computation and Machine Learning series)*. London: MIT press, 2016.
- [178] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: <https://doi.org/10.1162/neco.2006.18.7.1527>.
- [179] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556*, Sep. 2015.
- [180] C. Szegedy *et al.*, "Going Deeper with Convolutions," Sep. 2015.
- [181] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2016.
- [182] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning." pp. 1139–1147, Feb. 13, 2013.
- [183] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, "Regularization of Neural Networks using DropConnect." pp. 1058–1066, Feb. 13, 2013.
- [184] S. I. Wang and C. D. Manning, "Fast dropout training," 2013.
- [185] X. Wang *et al.*, "SSA-Net: Spatial Self-Attention Network for COVID-19 Pneumonia Infection Segmentation with Semi-supervised Few-shot Learning," *Medical Image Analysis*, p. 102459, Apr. 2022, doi: <https://doi.org/10.1016/j.media.2022.102459>.
- [186] J. Li, L. Qi, Q. Chen, Y.-D. Zhang, and X. Qian, "A dual meta-learning framework based on idle data for enhancing segmentation of pancreatic cancer," *Medical Image Analysis*, vol. 78, p. 102342, May 2022, doi: <https://doi.org/10.1016/j.media.2021.102342>.
- [187] Q. Lu *et al.*, "A transfer learning approach to few-shot segmentation of novel white matter tracts," *Medical Image Analysis*, vol. 79, p. 102454, Jul. 2022, doi: <https://doi.org/10.1016/j.media.2022.102454>.
- [188] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: <https://doi.org/10.1109/5.726791>.
- [189] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: <https://doi.org/10.1145/3065386>.
- [190] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556*, Sep. 2015, [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [191] C. Szegedy *et al.*, "Going Deeper with Convolutions," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, pp. 1–9, Sep. 2014, doi: <https://doi.org/10.48550/arxiv.1409.4842>.
- [192] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-December, pp. 770–778. doi: <https://doi.org/10.48550/arXiv.1512.03385>.
- [193] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: <https://doi.org/10.1016/j.media.2017.07.005>.

- [194] K. Kamnitsas *et al.*, “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” *Medical Image Analysis*, vol. 36, pp. 61–78, Feb. 2017, doi: <https://doi.org/10.1016/j.media.2016.10.004>.
- [195] Y. Yin, C. He, B. Xu, and Z. Li, “Coronary Plaque Characterization From Optical Coherence Tomography Imaging With a Two-Pathway Cascade Convolutional Neural Network Architecture,” *Front Cardiovasc Med*, vol. 8, Jun. 2021, doi: <https://doi.org/10.3389/fcvm.2021.670502>.
- [196] J. Zhou *et al.*, “Automatic Detection and Classification of Focal Liver Lesions Based on Deep Convolutional Neural Networks: A Preliminary Study,” *Frontiers in Oncology*, vol. 10, p. 3261, Jan. 2021, doi: <https://doi.org/10.3389/fonc.2020.581210>.
- [197] A. la Greca Saint-Estevan *et al.*, “A 2.5D convolutional neural network for HPV prediction in advanced oropharyngeal cancer,” *Computers in Biology and Medicine*, vol. 142, p. 105215, Mar. 2022, doi: <https://doi.org/10.1016/j.compbiomed.2022.105215>.
- [198] M. Havaei *et al.*, “Brain tumor segmentation with Deep Neural Networks,” *Medical Image Analysis*, vol. 35, pp. 18–31, Jan. 2017, doi: <https://doi.org/10.1016/J.MEDIA.2016.05.004>.
- [199] S. Valverde *et al.*, “Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach,” *Neuroimage*, vol. 155, pp. 159–168, Jul. 2017, doi: <https://doi.org/10.1016/J.NEUROIMAGE.2017.04.034>.
- [200] S. Trebeschi *et al.*, “Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR,” *Scientific Reports*, vol. 7, no. 1, p. 5301, Dec. 2017, doi: <https://doi.org/10.1038/s41598-017-05728-9>.
- [201] M. A. Mazurowski, M. Buda, A. Saha, and M. R. Bashir, “Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI,” *Journal of Magnetic Resonance Imaging*, vol. 49, no. 4, pp. 939–954, Apr. 2019, doi: <https://doi.org/10.1002/jmri.26534>.
- [202] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” Jun. 2016.
- [203] J. Schlemper *et al.*, “Attention gated networks: Learning to leverage salient regions in medical images,” *Medical Image Analysis*, vol. 53, pp. 197–207, Apr. 2019, doi: <https://doi.org/10.1016/J.MEDIA.2019.01.012>.
- [204] Z. Zhang, C. Wu, S. Coleman, and D. Kerr, “DENSE-INception U-net for medical image segmentation,” *Computer Methods and Programs in Biomedicine*, vol. 192, p. 105395, Aug. 2020, doi: <https://doi.org/10.1016/J.CMPB.2020.105395>.
- [205] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, “H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018, doi: <https://doi.org/10.1109/TMI.2018.2845918>.
- [206] B. Ji *et al.*, “A multi-scale recurrent fully convolution neural network for laryngeal leukoplakia segmentation,” *Biomedical Signal Processing and Control*, vol. 59, p. 101913, May 2020, doi: <https://doi.org/10.1016/J.BSPC.2020.101913>.
- [207] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11045 LNCS, pp. 3–11, 2018, doi: https://doi.org/10.1007/978-3-030-00889-5_1.

- [208] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation," *Neuroinformatics* 2018 16:3, vol. 16, no. 3, pp. 383–392, May 2018, doi: <https://doi.org/10.1007/S12021-018-9377-X>.
- [209] J. Hong, B. yong Park, M. J. Lee, C. S. Chung, J. Cha, and H. Park, "Two-step deep neural network for segmentation of deep white matter hyperintensities in migraineurs," *Computer Methods and Programs in Biomedicine*, vol. 183, p. 105065, Jan. 2020, doi: <https://doi.org/10.1016/J.CMPB.2019.105065>.
- [210] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods* 2020 18:2, vol. 18, no. 2, pp. 203–211, Dec. 2020, doi: <https://doi.org/10.1038/s41592-020-01008-z>.
- [211] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995, doi: <https://doi.org/10.1006/CVIU.1995.1004>.
- [212] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, Jun. 2001, doi: <https://doi.org/10.1109/34.927467>.
- [213] M. Sonka, B. P. F. Lelieveldt, S. C. Mitchell, J. G. Bosch, R. J. Van der Geest, and J. H. C. Reiber, "Active appearance motion model segmentation," in *Proceedings Second International Workshop on Digital and Computational Video*, pp. 64–68. doi: <https://doi.org/10.1109/DCV.2001.929943>.
- [214] T. Dietenbeck, M. Alessandrini, D. Bardosa, J. D'hooge, D. Friboulet, and O. Bernanrd, "Detection of the whole myocardium in 2D-echocardiography for multiple orientations using a geometrically constrained level-set," *Medical Image Analysis*, vol. 16, no. 2, pp. 386–401, Feb. 2012, doi: <https://doi.org/10.1016/J.MEDIA.2011.10.003>.
- [215] T. Chan and W. Zhu, "Level Set Based Shape Prior Segmentation," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 1164–1170. doi: <https://doi.org/10.1109/CVPR.2005.212>.
- [216] M. Rousson and N. Paragios, "Prior Knowledge, Level Set Representations & Visual Grouping," *International Journal of Computer Vision*, vol. 76, no. 3, pp. 231–243, Mar. 2008, doi: [10.1007/s11263-007-0054-z](https://doi.org/10.1007/s11263-007-0054-z).
- [217] S. Safar and M.-H. Yang, "Learning shape priors for object segmentation via neural networks," in *2015 IEEE International Conference on Image Processing (ICIP)*, Sep. 2015, pp. 1835–1839. doi: <https://doi.org/10.1109/ICIP.2015.7351118>.
- [218] H. Ravishankar, R. Venkataramani, S. Thiruvankadam, P. Sudhakar, and V. Vaidya, "Learning and Incorporating Shape Models for Semantic Segmentation," Springer, Cham, 2017, pp. 203–211. doi: https://doi.org/10.1007/978-3-319-66182-7_24.
- [219] O. Oktay *et al.*, "Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384–395, Feb. 2018, doi: <https://doi.org/10.1109/TMI.2017.2743464>.
- [220] A. V. Dalca, J. Guttag, and M. R. Sabuncu, "Anatomical Priors in Convolutional Networks for Unsupervised Biomedical Segmentation." pp. 9290–9299, 2018.
- [221] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, and Etal, "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge," Nov. 2018.

- [222] N. Wang, Y. Zhang, and L. Zhang, "Dynamic Selection Network for Image Inpainting," *IEEE Transactions on Image Processing*, vol. 30, pp. 1784–1798, 2021, doi: <https://doi.org/10.1109/TIP.2020.3048629>.
- [223] J. Li, N. Wang, L. Zhang, B. Du, and D. Tao, "Recurrent feature reasoning for image inpainting," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 7757–7765, 2020, doi: <https://doi.org/10.1109/CVPR42600.2020.00778>.
- [224] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative Image Inpainting with Contextual Attention," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 5505–5514, Dec. 2018, doi: <https://doi.org/10.1109/CVPR.2018.00577>.
- [225] G. Liu, F. A. Reda, K. J. Shih, T. C. Wang, A. Tao, and B. Catanzaro, "Image Inpainting for Irregular Holes Using Partial Convolutions," in *Lecture Notes in Computer Science*, Sep. 2018, vol. 11215 LNCS, pp. 89–105. doi: https://doi.org/10.1007/978-3-030-01252-6_6.
- [226] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-October, pp. 4470–4479, Oct. 2019, doi: <https://doi.org/10.1109/ICCV.2019.00457>.
- [227] J. v. Manjón *et al.*, "Blind mri brain lesion inpainting using deep learning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12417 LNCS, pp. 41–49, 2020, doi: https://doi.org/10.1007/978-3-030-59520-3_5.
- [228] R. Zhang *et al.*, "A Progressive Generative Adversarial Method for Structurally Inadequate Medical Image Data Augmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 7–16, Jan. 2022, doi: <https://doi.org/10.1109/JBHI.2021.3101551>.
- [229] H. Liao *et al.*, "Generative Mask Pyramid Network for CT/CBCT Metal Artifact Reduction with Joint Projection-Sinogram Correction," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11769 LNCS, pp. 77–85, 2019, doi: https://doi.org/10.1007/978-3-030-32226-7_9.
- [230] C. Bukas *et al.*, "Patient-Specific Virtual Spine Straightening and Vertebra Inpainting: An Automatic Framework for Osteoplasty Planning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12904 LNCS, pp. 529–539, 2021, doi: https://doi.org/10.1007/978-3-030-87202-1_51.
- [231] A. Chatsias *et al.*, "Disentangled representation learning in cardiac image analysis," *Medical Image Analysis*, vol. 58, p. 101535, Dec. 2019, doi: <https://doi.org/10.1016/j.media.2019.101535>.
- [232] S. Liao, Y. Gao, A. Oto, and D. Shen, "Representation Learning: A Unified Deep Learning Framework for Automatic Prostate MR Segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8150 LNCS, no. PART 2, pp. 254–261, 2013, doi: https://doi.org/10.1007/978-3-642-40763-5_32.
- [233] Z. Ning, C. Tu, X. Di, Q. Feng, and Y. Zhang, "Deep cross-view co-regularized representation learning for glioma subtype identification," *Medical Image Analysis*, vol. 73, p. 102160, Oct. 2021, doi: <https://doi.org/10.1016/j.media.2021.102160>.
- [234] Y. Liu *et al.*, "Incomplete multi-modal representation learning for Alzheimer's disease diagnosis," *Medical Image Analysis*, vol. 69, p. 101953, Apr. 2021, doi: <https://doi.org/10.1016/j.media.2020.101953>.

- [235] D. Zimmerer, F. Isensee, J. Petersen, S. Kohl, and K. Maier-Hein, "Unsupervised anomaly localization using variational auto-encoders," in *Lecture Notes in Computer Science*, Oct. 2019, vol. 11767 LNCS, pp. 289–297. doi: https://doi.org/10.1007/978-3-030-32251-9_32.
- [236] D. Zimmerer, S. A. A. Kohl, J. Petersen, F. Isensee, and K. H. Maier-Hein, "Context-encoding Variational Autoencoder for Unsupervised Anomaly Detection," Dec. 2018, doi: <https://doi.org/10.48550/arXiv.1812.05941>.
- [237] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab, "Scale-Space Autoencoders for Unsupervised Anomaly Segmentation in Brain MRI," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12264 LNCS, pp. 552–561, Oct. 2020, doi: https://doi.org/10.1007/978-3-030-59719-1_54.
- [238] C. Baur, R. Graf, B. Wiestler, S. Albarqouni, and N. Navab, "SteGANomaly: Inhibiting CycleGAN Steganography for Unsupervised Anomaly Detection in Brain MRI," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12262 LNCS, pp. 718–727, Oct. 2020, doi: https://doi.org/10.1007/978-3-030-59713-9_69.
- [239] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab, "Deep autoencoding models for unsupervised anomaly segmentation in brain MR images," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Sep. 2019, vol. 11383 LNCS, pp. 161–169. doi: https://doi.org/10.1007/978-3-030-11723-8_16.
- [240] C. Baur, S. Denner, B. Wiestler, N. Navab, and S. Albarqouni, "Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study," *Medical Image Analysis*, vol. 69. Elsevier B.V., p. 101952, Apr. 01, 2021. doi: <https://doi.org/10.1016/j.media.2020.101952>.
- [241] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv:1312.6114*, Dec. 2014.
- [242] C. Doersch, "Tutorial on Variational Autoencoders," *arXiv:1606.05908*, Jun. 2016.
- [243] O. Oktay *et al.*, "Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384–395, Feb. 2018, doi: <https://doi.org/10.1109/TMI.2017.2743464>.
- [244] S. G. Armato *et al.*, *Data From LIDC-IDRI, The Cancer Imaging Archive*. 2015. doi: <https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX>.
- [245] "MeVisLab." MeVis, Medical Solution AG, , 2016. [Online]. Available: <https://www.mevislab.de/>
- [246] P. E. Kinahan and J. W. Fletcher, "PET/CT Standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy," *Semin Ultrasound CT MR*, vol. 31, no. 6, p. 496, Dec. 2010, doi: <https://doi.org/10.1053/J.SULT.2010.10.001>.
- [247] "Mialab, Medical Image Analysis." [Online]. Available: <http://mialab.org/>
- [248] H. J. W. L. Aerts *et al.*, "Data From NSCLC-Radiomics [Data set]," *The Cancer Imaging Archive*, 2019. <https://doi.org/10.7937/K9/TCIA.2015.PF0M9REI>
- [249] H. J. W. L. Aerts *et al.*, "Data From NSCLC-Radiomics-Genomics [Data set]," Jun. 03, 2015. <https://doi.org/10.7937/K9/TCIA.2015.L4FRET6Z>
- [250] J. J. M. van Griethuysen *et al.*, "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Research*, vol. 77, no. 21, pp. e104–e107, Nov. 2017, doi: <https://doi.org/10.1158/0008-5472.CAN-17-0339>.

- [251] Martin Vallières, "Radiomics, MATLAB programming tool." [Online]. Available: <https://github.com/mvallieres/radiomics>
- [252] I. Toma-Dasu *et al.*, "Evaluating Tumor Response of Non-Small Cell Lung Cancer Patients With 18F-Fludeoxyglucose Positron Emission Tomography: Potential for Treatment Individualization," *International Journal of Radiation Oncology*Biophysics*, vol. 91, no. 2, pp. 376–384, Feb. 2015, doi: <https://doi.org/10.1016/j.IJROBP.2014.10.012>.
- [253] K. Loverdos, A. Fotiadis, C. Kontogianni, M. Iliopoulou, and M. Gaga, "Lung nodules: A comprehensive review on current approach and management.," *Ann Thorac Med*, vol. 14, no. 4, pp. 226–238, 2019, doi: https://doi.org/10.4103/atm.ATM_110_19.
- [254] American college of Radiology, "Lung CT screening reporting and data system (Lung-RADS)," 2014. <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads>
- [255] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2261–2269, Aug. 2016, Accessed: Feb. 03, 2021. [Online]. Available: <http://arxiv.org/abs/1608.06993>
- [256] M. Tan and Q. v. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *36th International Conference on Machine Learning, ICML 2019*, vol. 2019-June, pp. 10691–10700, May 2019, doi: <https://doi.org/10.48550/arXiv.1905.11946>.
- [257] F. Sinzinger, M. Astaraki, Ö. Smedby, and R. Moreno, "Spherical Convolutional Neural Networks for Survival Rate Prediction in Cancer Patients," *Frontiers in Oncology*, vol. 0, p. 1524, Apr. 2022, doi: <https://doi.org/10.3389/FONC.2022.870457>.
- [258] N. v Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: <https://doi.org/10.1613/jair.953>.
- [259] B. Zhang *et al.*, "Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma," *Clinical Cancer Research*, vol. 23, no. 15, pp. 4259–4269, Aug. 2017, doi: <https://doi.org/10.1158/1078-0432.CCR-16-2910>.
- [260] M. Avanzo, J. Stancanella, and I. el Naqa, "Beyond imaging: The promise of radiomics," *Physica Medica*, vol. 38, pp. 122–139, 2017, doi: <https://doi.org/10.1016/j.ejmp.2017.05.071>.
- [261] H. Steck, B. Krishnapuram, C. Dehing-oberije, P. Lambin, and V. Raykar, "On Ranking in Survival Analysis: Bounds on the Concordance Index," 2007.
- [262] WHO, "WHO methods and data sources for global burden of disease estimates 2000-2019," Geneva, Dec. 2020.
- [263] C. Mattiuzzi and G. Lippi, "Current Cancer Epidemiology," *Journal of Epidemiology and Global Health*, vol. 9, no. 4, p. 217, Dec. 2019, doi: <https://doi.org/10.2991/JEGH.K.191008.001>.
- [264] F. Meric-Bernstam, C. Farhangfar, J. Mendelsohn, and G. B. Mills, "Building a Personalized Medicine Infrastructure at a Major Cancer Center," *Journal of Clinical Oncology*, vol. 31, no. 15, p. 1849, May 2013, doi: <https://doi.org/10.1200/JCO.2012.45.3043>.
- [265] M. Diamandis, N. M. A. White, and G. M. Yousef, "Personalized medicine: Marking a new epoch in cancer patient management," *Molecular Cancer Research*, vol. 8, no. 9, pp. 1175–1187, Sep. 2010, doi: <https://doi.org/10.1158/1541-7786.MCR-10-0264>.

- [266] B. S. Kramer, C. D. Berg, D. R. Aberle, and P. C. Prorok, "Lung cancer screening with low-dose helical CT: Results from the national lung screening trial (NLST)," *Journal of Medical Screening*, vol. 18, no. 3, pp. 109–111, Sep. 2011, doi: <https://doi.org/10.1258/jms.2011.011055>.
- [267] N. Horeweg *et al.*, "Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening," *The Lancet Oncology*, vol. 15, no. 12, pp. 1332–1341, Nov. 2014, doi: [https://doi.org/10.1016/S1470-2045\(14\)70389-4](https://doi.org/10.1016/S1470-2045(14)70389-4).
- [268] W. Jarmoowski, X. Ren, P. Wielgosz, and T. M. Peters, "Image-guidance for surgical procedures," *Physics in Medicine & Biology*, vol. 51, no. 14, p. R505, Jun. 2006, doi: <https://doi.org/10.1088/0031-9155/51/14/R01>.
- [269] S. J. Uyebico, C. C. Wu, R. D. Suh, N. H. Le, K. Brown, and M. S. Krishnam, "Lung cancer staging essentials: The new TNM staging system and potential imaging pitfalls," *Radiographics*, vol. 30, no. 5, pp. 1163–1181, Sep. 2010, doi: <https://doi.org/10.1148/rg.305095166>.
- [270] G. C. Pereira, M. Traughber, and R. F. Muzic, "The role of imaging in radiation therapy planning: Past, present, and future," *BioMed Research International*, vol. 2014, 2014, doi: <https://doi.org/10.1155/2014/231090>.
- [271] C. Shi *et al.*, "Imaging Beyond Seeing: Early Prognosis of Cancer Treatment," *Small Methods*, vol. 5, no. 3, p. 2001025, Mar. 2021, doi: <https://doi.org/10.1002/SMTD.202001025>.
- [272] K. Kubota *et al.*, "FDG-PET delayed imaging for the detection of head and neck cancer recurrence after radio-chemotherapy: Comparison with MRI/CT," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 31, no. 4, pp. 590–595, Jan. 2004, doi: <https://doi.org/10.1007/s00259-003-1408-6>.
- [273] A. Ford, S. Bydder, and M. A. Ebert, "The use of On-Board Imaging to plan and deliver palliative radiotherapy in a single cohesive patient appointment," *Journal of Medical Imaging and Radiation Oncology*, vol. 55, no. 6, pp. 633–638, Dec. 2011, doi: <https://doi.org/10.1111/j.1754-9485.2011.02321.x>.
- [274] MedSuite, "Medical Imaging Market Analysis, Size, Trends ," 2016. <https://idataresearch.com/product/medical-imaging-market-united-states/>
- [275] M. Mottola *et al.*, "Reproducibility of CT-based radiomic features against image resampling and perturbations for tumour and healthy kidney in renal cancer patients," *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–11, Jun. 2021, doi: <https://doi.org/10.1038/s41598-021-90985-y>.
- [276] A. Chaddad *et al.*, "Magnetic Resonance Imaging Based Radiomic Models of Prostate Cancer: A Narrative Review," *Cancers 2021, Vol. 13, Page 552*, vol. 13, no. 3, p. 552, Feb. 2021, doi: <https://doi.org/10.3390/CANCERS13030552>.
- [277] D. C. Sullivan *et al.*, "Metrology standards for quantitative imaging biomarkers1," *Radiology*, vol. 277, no. 3, pp. 813–825, Dec. 2015, doi: <https://doi.org/10.1148/radiol.2015142202>.
- [278] H. Jiang, Z. Diao, and Y. D. Yao, "Deep learning techniques for tumor segmentation: a review," *Journal of Supercomputing*, vol. 78, no. 2, pp. 1807–1851, Feb. 2022, doi: <https://doi.org/10.1007/s11227-021-03901-6>.
- [279] T. Dietenbeck, M. Alessandrini, D. Bardosa, J. D'hooge, D. Friboulet, and O. Bernanrd, "Detection of the whole myocardium in 2D-echocardiography for multiple orientations using a geometrically constrained level-set," *Medical Image Analysis*, vol. 16, no. 2, pp. 386–401, Feb. 2012, doi: <https://doi.org/10.1016/j.media.2011.10.003>.

- [280] C. Dong *et al.*, “Segmentation of liver and spleen based on computational anatomy models,” *Computers in Biology and Medicine*, vol. 67, pp. 146–160, Dec. 2015, doi: <https://doi.org/10.1016/j.compbiomed.2015.10.007>.
- [281] H. Ravishankar, R. Venkataramani, S. Thiruvankadam, P. Sudhakar, and V. Vaidya, “Learning and Incorporating Shape Models for Semantic Segmentation,” Springer, Cham, 2017, pp. 203–211. doi: https://doi.org/10.1007/978-3-319-66182-7_24.
- [282] R. J. Araújo, J. S. Cardoso, and H. P. Oliveira, “A deep learning design for improving topology coherence in blood vessel segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Oct. 2019, vol. 11764 LNCS, pp. 93–101. doi: https://doi.org/10.1007/978-3-030-32239-7_11.
- [283] S. Zheng *et al.*, “Conditional random fields as recurrent neural networks,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1529–1537. doi: <https://doi.org/10.1109/ICCV.2015.179>.
- [284] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: <https://doi.org/10.1109/TPAMI.2017.2699184>.
- [285] H. Zheng *et al.*, “Semi-supervised Segmentation of Liver Using Adversarial Learning with Deep Atlas Prior,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Oct. 2019, vol. 11769 LNCS, pp. 148–156. doi: https://doi.org/10.1007/978-3-030-32226-7_17.
- [286] A. BenTaieb and G. Hamarneh, “Topology aware fully convolutional networks for histology gland segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Oct. 2016, vol. 9901 LNCS, pp. 460–468. doi: https://doi.org/10.1007/978-3-319-46723-8_53.
- [287] A. Mosinska, P. Marquez-Neila, M. Kozinski, and P. Fua, “Beyond the Pixel-Wise Loss for Topology-Aware Delineation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2018, pp. 3136–3145. doi: <https://doi.org/10.1109/CVPR.2018.00331>.
- [288] C. Zotti, Z. Luo, A. Lalande, and P. M. Jodoin, “Convolutional Neural Network with Shape Prior Applied to Cardiac MRI Segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1119–1128, May 2019, doi: <https://doi.org/10.1109/JBHI.2018.2865450>.
- [289] R. Wang, S. Cao, K. Ma, Y. Zheng, and D. Meng, “Pairwise learning for medical image segmentation,” *Medical Image Analysis*, vol. 67, p. 101876, Jan. 2021, doi: <https://doi.org/10.1016/j.media.2020.101876>.
- [290] Y. Zhou *et al.*, “Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation,” *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-October, pp. 10671–10680, Apr. 2019, doi: <https://doi.org/10.48550/arXiv.1904.06346>.
- [291] D. Gu, G. Liu, and Z. Xue, “On the performance of lung nodule detection, segmentation and classification,” *Computerized Medical Imaging and Graphics*, vol. 89, p. 101886, Apr. 2021, doi: <https://doi.org/10.1016/j.compmedimag.2021.101886>.
- [292] S. Wang *et al.*, “Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation,” *Medical Image Analysis*, vol. 40, pp. 172–183, Aug. 2017, doi: <https://doi.org/10.1016/j.media.2017.06.014>.

- [293] D. Wang, T. Zhang, M. Li, R. Bueno, and J. Jayender, "3D deep learning based classification of pulmonary ground glass opacity nodules with automatic segmentation," *Computerized Medical Imaging and Graphics*, vol. 88, p. 101814, Mar. 2021, doi: <https://doi.org/10.1016/j.compmedimag.2020.101814>.
- [294] D. M. Hansell, A. A. Bankier, H. MacMahon, T. C. McCloud, N. L. Müller, and J. Remy, "Fleischner Society: Glossary of Terms for Thoracic Imaging," *Radiology*, vol. 246, no. 3, pp. 697–722, Mar. 2008, doi: <https://doi.org/10.1148/radiol.2462070712>.
- [295] D. Di *et al.*, "Hypergraph learning for identification of COVID-19 with CT imaging," *Medical Image Analysis*, vol. 68, p. 101910, Feb. 2021, doi: <https://doi.org/10.1016/j.media.2020.101910>.
- [296] W. H. L. Pinaya *et al.*, "Unsupervised Brain Imaging 3D Anomaly Detection and Segmentation with Transformers," *Medical Image Analysis*, p. 102475, May 2022, doi: <https://doi.org/10.1016/j.media.2022.102475>.
- [297] V. Andrearczyk *et al.*, "Automatic Segmentation of Head and Neck Tumors and Nodal Metastases in PET-CT scans," in *Medical Imaging with Deep Learning*, 2020, pp. 33–43. [Online]. Available: <https://proceedings.mlr.press/v121/andrearczyk20a.html>
- [298] X. Xu *et al.*, "MSCS-DeepLN: Evaluating lung nodule malignancy using multi-scale cost-sensitive neural networks," *Medical Image Analysis*, vol. 65, p. 101772, Oct. 2020, doi: <https://doi.org/10.1016/j.media.2020.101772>.
- [299] Y. Xie, J. Zhang, and Y. Xia, "Semi-supervised adversarial model for benign–malignant lung nodule classification on chest CT," *Medical Image Analysis*, vol. 57, pp. 237–248, Oct. 2019, doi: <https://doi.org/10.1016/j.media.2019.07.004>.
- [300] C. Reimers, J. Runge, and J. Denzler, "Determining the Relevance of Features for Deep Neural Networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12371 LNCS, pp. 330–346, 2020, doi: https://doi.org/10.1007/978-3-030-58574-7_20.
- [301] P. Sun, D. Wang, V. C. Mok, and L. Shi, "Comparison of Feature Selection Methods and Machine Learning Classifiers for Radiomics Analysis in Glioma Grading," *IEEE Access*, vol. 7, pp. 102010–102020, 2019, doi: <https://doi.org/10.1109/ACCESS.2019.2928975>.
- [302] J. E. van Timmeren, R. T. H. Leijenaar, W. van Elmpt, B. Reymen, and P. Lambin, "Feature selection methodology for longitudinal cone-beam CT radiomics," *Acta Oncologica*, vol. 56, no. 11, pp. 1537–1543, Nov. 2017, doi: <https://doi.org/10.1080/0284186X.2017.1350285>.
- [303] P. Yin *et al.*, "Comparison of radiomics machine-learning classifiers and feature selection for differentiation of sacral chordoma and sacral giant cell tumour based on 3D computed tomography features," *European Radiology*, vol. 29, no. 4, pp. 1841–1847, Apr. 2019, doi: <https://doi.org/10.1007/s00330-018-5730-6>.
- [304] A. Demircioğlu, "Evaluation of the dependence of radiomic features on the machine learning model," *Insights into Imaging*, vol. 13, no. 1, Dec. 2022, doi: <https://doi.org/10.1186/S13244-022-01170-2>.
- [305] Z. R. Bardosi *et al.*, "Benchmarking Eliminative Radiomic Feature Selection for Head and Neck Lymph Node Classification," *Cancers 2022, Vol. 14, Page 477*, vol. 14, no. 3, p. 477, Jan. 2022, doi: <https://doi.org/10.3390/CANCERS14030477>.
- [306] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, pp. 1–54, Dec. 2019, doi: <https://doi.org/10.1186/s40537-019-0192-5>.

- [307] Y. Xie, J. Zhang, Y. Xia, M. Fulham, and Y. Zhang, "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT," *Information Fusion*, vol. 42, pp. 102–110, Jul. 2018, doi: <https://doi.org/10.1016/J.INFFUS.2017.10.005>.
- [308] J. L. Causey *et al.*, "Highly accurate model for prediction of lung nodule malignancy with CT scans," *Scientific Reports*, vol. 8, no. 1, p. 9286, Dec. 2018, doi: <https://doi.org/10.1038/s41598-018-27569-w>.
- [309] P. Papadimitroulas *et al.*, "Artificial intelligence: Deep learning in oncological radiomics and challenges of interpretability and data harmonization," *Physica Medica*, vol. 83, pp. 108–121, Mar. 2021, doi: <https://doi.org/10.1016/J.EJMP.2021.03.009>.
- [310] H. J. W. L. Aerts *et al.*, "Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC," *Scientific Reports*, vol. 6, no. 1, p. 33860, Sep. 2016, doi: <https://doi.org/10.1038/srep33860>.
- [311] T. P. Coroller *et al.*, "Radiomic-Based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC," *Journal of Thoracic Oncology*, vol. 12, no. 3, pp. 467–476, 2017, doi: <https://doi.org/10.1016/j.jtho.2016.11.2226>.
- [312] S. R. Digumarthy, A. M. Padole, R. lo Gullo, L. v. Sequist, and M. K. Kalra, "Can CT radiomic analysis in NSCLC predict histology and EGFR mutation status?," *Medicine (United States)*, vol. 98, no. 1, p. E13963, Jan. 2019, doi: <https://doi.org/10.1097/MD.00000000000013963>.
- [313] C. L. Yanqi Huang, Zaiyi Liu, Lan He, Xin Chen, Dan Pan, Zelan Ma, Cuishan Liang, Jie Tian, "Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non—Small Cell Lung Cancer," *Radiology*, vol. 281, no. 3, 2016, doi: <https://doi.org/10.1148/radiol.2016152234>.
- [314] E. E. C. de Jong *et al.*, "Applicability of a prognostic CT-based radiomic signature model trained on stage I-III non-small cell lung cancer in stage IV non-small cell lung cancer.," *Lung Cancer*, vol. 124, pp. 6–11, Oct. 2018, doi: <https://doi.org/10.1016/j.lungcan.2018.07.023>.
- [315] D. v Fried *et al.*, "Prognostic Value and Reproducibility of Pretreatment CT Texture Features in Stage III Non-Small Cell Lung Cancer," *International Journal of Radiation Oncology Biology Physicsy Biology Physics*, vol. 90, no. 4, pp. 834–842, 2014, doi: <https://doi.org/10.1016/j.ijrobp.2014.07.020>.
- [316] X. Z. Xinzhe Dong, Xiaorong Sun, Lu Sun, Peter G. Maxim, Lei Xing, Yong Huang, Wenwu Li, Honglin Wan, "Early Change in Metabolic Tumor Heterogeneity during Chemoradiotherapy and Its Prognostic Value for Patients with Locally Advanced Non-Small Cell Lung Cancer," *PLOS ONE*, vol. 11, no. 6, 2016, doi: <https://doi.org/10.1371/journal.pone.0157836>.
- [317] P. Lovinfosse *et al.*, "FDG PET/CT radiomics for predicting the outcome of locally advanced rectal cancer," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 45, no. 3, pp. 365–375, Mar. 2018, doi: <https://doi.org/10.1007/s00259-017-3855-5>.
- [318] J. W. Lee and S. M. Lee, "Radiomics in Oncological PET/CT: Clinical Applications," *Nuclear Medicine and Molecular Imaging 2017 52:3*, vol. 52, no. 3, pp. 170–189, Oct. 2017, doi: <https://doi.org/10.1007/S13139-017-0500-Y>.
- [319] T. P. Coroller *et al.*, "Radiomic phenotype features predict pathological response in non-small cell lung cancer," *Radiotherapy and Oncology*, vol. 119, no. 3, pp. 480–486, Jun. 2016, doi: <https://doi.org/10.1016/J.RADONC.2016.04.004>.

- [320] J. E. van Timmeren *et al.*, “Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images,” *Radiotherapy and Oncology*, vol. 123, no. 3, pp. 363–369, Jun. 2017, doi: <https://doi.org/10.1016/j.radonc.2017.04.016>.
- [321] S. Park *et al.*, “Performance of radiomics models for survival prediction in non-small-cell lung cancer: influence of CT slice thickness,” *European Radiology*, vol. 31, no. 5, pp. 2856–2865, May 2021, doi: <https://doi.org/10.1007/S00330-020-07423-2/TABLES/4>.
- [322] N. Chen *et al.*, “Progression-Free Survival Prediction in Small Cell Lung Cancer Based on Radiomics Analysis of Contrast-Enhanced CT,” *Frontiers in Medicine*, vol. 9, p. 292, Feb. 2022, doi: <https://doi.org/10.3389/fmed.2022.833283>.
- [323] E. D’Arnese, G. W. di Donato, E. del Sozzo, M. Sollini, D. Sciuto, and M. D. Santambrogio, “On the Automation of Radiomics-based Identification and Characterization of NSCLC,” *IEEE Journal of Biomedical and Health Informatics*, 2022, doi: <https://doi.org/10.1109/JBHI.2022.3156984>.
- [324] T. Pyka *et al.*, “Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy,” *Radiat Oncol*, vol. 10, no. 1, Apr. 2015, doi: <https://doi.org/10.1186/S13014-015-0407-7>.
- [325] L. A. Vale-Silva and K. Rohr, “Long-term cancer survival prediction using multimodal deep learning,” *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–12, Jun. 2021, doi: <https://doi.org/10.1038/s41598-021-92799-4>.
- [326] S. Cheon *et al.*, “The accuracy of clinicians’ predictions of survival in advanced cancer: A review,” *Annals of Palliative Medicine*, vol. 5, no. 1, pp. 22–29, 2016, doi: <https://doi.org/10.3978/J.ISSN.2224-5820.2015.08.04>.
- [327] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nature Medicine 2019 25:1*, vol. 25, no. 1, pp. 44–56, Jan. 2019, doi: <https://doi.org/10.1038/s41591-018-0300-7>.
- [328] S. Benjamens, P. Dhunoo, and B. Meskó, “The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database,” *npj Digital Medicine 2020 3:1*, vol. 3, no. 1, pp. 1–8, Sep. 2020, doi: <https://doi.org/10.1038/s41746-020-00324-0>.
- [329] A. L. Beam and I. S. Kohane, “Big Data and Machine Learning in Health Care,” *JAMA*, vol. 319, no. 13, pp. 1317–1318, Apr. 2018, doi: <https://doi.org/10.1001/JAMA.2017.18391>.
- [330] G. Wang, Y. Zhang, X. Ye, and X. Mou, “Machine Learning for Tomographic Imaging,” *Machine Learning for Tomographic Imaging*, Dec. 2019, doi: <https://doi.org/10.1088/978-0-7503-2216-4>.
- [331] M. Gaillochet, K. C. Tezcan, and E. Konukoglu, “Joint Reconstruction and Bias Field Correction for Undersampled MR Imaging,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12262 LNCS, pp. 44–52, 2020, doi: https://doi.org/10.1007/978-3-030-59713-9_5.
- [332] J. Cheng *et al.*, “ResGANet: Residual group attention network for medical image classification and segmentation,” *Medical Image Analysis*, vol. 76, p. 102313, Feb. 2022, doi: <https://doi.org/10.1016/J.MEDIA.2021.102313>.
- [333] R. Wang, S. Chen, C. Ji, J. Fan, and Y. Li, “Boundary-aware context neural network for medical image segmentation,” *Medical Image Analysis*, vol. 78, p. 102395, May 2022, doi: <https://doi.org/10.1016/J.MEDIA.2022.102395>.

- [334] “The Medical Segmentation Decathlon Leaderboard,” 2018. <http://medicaldecathlon.com/results/>
- [335] M. Antonelli *et al.*, “The Medical Segmentation Decathlon,” Jun. 2021, doi: <https://doi.org/10.48550/arXiv.2106.05735>.
- [336] “The Medical Segmentation Decathlon Post-challenge Leaderboard,” 2021. <https://decathlon-10.grand-challenge.org/evaluation/challenge/leaderboard/>
- [337] B. H. Menze *et al.*, “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015, doi: <https://doi.org/10.1109/TMI.2014.2377694>.
- [338] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, and Etal, “Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge,” Nov. 2018, doi: <https://doi.org/10.48550/arXiv.1811.02629>.
- [339] M. Futrega, A. Milesi, M. Marcinkiewicz, and P. Ribalta, “Optimized U-Net for Brain Tumor Segmentation,” Oct. 2021, doi: <https://doi.org/10.48550/arxiv.2110.03352>.
- [340] “BraTS 2021 Leaderboard,” 2021. <https://www.synapse.org/#!Synapse:syn25829067/wiki/612712>
- [341] L. Weninger, O. Rippel, S. Koppers, and D. Merhof, “Segmentation of brain tumors and patient survival prediction: Methods for the braTS 2018 challenge,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11384 LNCS, pp. 3–12, 2019, doi: https://doi.org/10.1007/978-3-030-11726-9_1.
- [342] F. Kofler *et al.*, “blob loss: instance imbalance aware loss functions for semantic segmentation,” May 2022, doi: <https://doi.org/10.48550/arXiv.2205.08209>.
- [343] J. Gao *et al.*, “A Survey on Deep Learning for Multimodal Data Fusion,” *Neural Computation*, vol. 32, no. 5, pp. 829–864, May 2020, doi: https://doi.org/10.1162/neco_a_01273.
- [344] Y. S. Choi *et al.*, “Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics,” *Neuro-Oncology*, vol. 23, no. 2, pp. 304–313, Feb. 2021, doi: <https://doi.org/10.1093/NEUONC/NOAA177>.
- [345] D. Zeng *et al.*, “Positional Contrastive Learning for Volumetric Medical Image Segmentation,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12902 LNCS, pp. 221–230, 2021, doi: https://doi.org/10.1007/978-3-030-87196-3_21.