

VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER SCIENCE AND ENGINEERING



REPORT
CAPSTONE PROJECT

**SURVIVAL PREDICTION OF LUNG
CANCER PATIENTS USING
MEDICAL IMAGES WITH DEEP
LEARNING**

Major: Computer Science

COMMITTEE: Council 4 CS-CN

SUPERVISOR: Vo Thanh Hung, MSc.

SECRETARY: Mr. Le Binh Dang

REVIEWER: Mr. Duong Duc Tin

—o0o—

STUDENT: Nguyen Minh Quang - 2153724

HO CHI MINH CITY, MAY 2025

Supervisor's Signature

This is to certify that the topic prepared by Nguyễn Minh Quang, entitled, **SURVIVAL PREDICTION OF LUNG CANCER PATIENTS USING MEDICAL IMAGES WITH DEEP LEARNING** submitted in fulfillment of the requirement for the degree of Computer Science Engineer complies with the regulation of the University and meets the accepted standards with respect to originality.

Signed by advisor

Name

Signature

Date

(Advisor)

(Co-Advisor)

Declaration

I hereby declare that this research is entirely my own work, conducted under the supervision and guidance of MSc. Võ Thanh Hùng. The results presented in this report are authentic and have not been published in any form prior to this submission. All materials used in the research process were collected independently from various sources and are appropriately cited in the References section.

Furthermore, I acknowledge that this research includes the use of results from other authors and organizations. These contributions have been clearly cited and referenced.

In the event of any plagiarism or copyright infringement, I take full responsibility for my actions. Ho Chi Minh City University of Technology is not liable for any violations of copyright that may arise from this research.

Acknowledgements

This capstone project was completed under the scientific guidance of MSc. Võ Thanh Hùng, Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology - Vietnam National University Ho Chi Minh City. I would like to express my sincere gratitude to MSc. Võ Thanh Hùng for his invaluable expertise, guidance, discussions, feedback, and the favorable conditions he provided, enabling me to complete this project successfully.

I also extend my special thanks to the reviewers, who took the time to read my work and provided constructive feedback to help me improve my specialized project. My deepest appreciation goes to the lecturers of the Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology - Vietnam National University Ho Chi Minh City, who have imparted their profound knowledge and created an excellent environment for me to learn and grow throughout our academic journey.

I wish all the faculty members good health, success, and continued contributions to educating future generations of students.

Thank you sincerely.

Abstract

Lung cancer remains a leading cause of mortality worldwide, emphasizing the urgent need for advancements in early diagnosis and survival prediction. This project proposes a deep learning-based pipeline to address this challenge, integrating medical imaging and clinical data for accurate survival analysis of lung cancer patients. The pipeline consists of two stages: tumor segmentation and survival prediction.

In the first stage, the nnU-Net architecture is employed to segment lung tumors in CT images. This automated and adaptable model ensures robust segmentation by optimizing preprocessing, architecture, and postprocessing steps for the dataset. Extracted tumor features such as size, shape, and location are then integrated with clinical variables, including demographics and medical history, to form a comprehensive feature set.

In the second stage, the combined features are used as input for various survival analysis models, including traditional Cox Proportional Hazards and Accelerated Failure Time models, as well as the neural network-based DeepSurv. Additionally, an enhanced architecture is explored, in which CT slices and clinical data are processed through a dual-branch neural network that learns representations directly from imaging and tabular data. Although this enhanced method did not outperform the primary feature-based approach, it demonstrated promising performance and potential for further improvement.

The proposed framework shows that incorporating CT-derived tumor characteristics significantly improves survival prediction performance. This research highlights the effectiveness of integrating imaging and clinical data, and sets the foundation for future extensions to other cancer types and multimodal applications in personalized oncology.

Contents

1	Introduction	1
1.1	Introduction	1
1.2	Objectives	2
1.3	Scope	3
1.4	Report Outline	4
2	Problem Requirements Analysis	5
3	Theoretical Background	7
3.1	Overview of Machine Learning and Deep Learning	7
3.1.1	Machine Learning	7
3.1.2	Deep Learning	8
3.2	Computer Vision	9
3.3	Survival Analysis Fundamentals	12
3.3.1	Censoring Data	13
3.3.2	Survival and Hazard	14
3.3.3	Hazard Ratio	14
3.3.4	Univariate Analysis	15
3.3.5	Covariate Analysis	17
4	Related Works	19
4.1	Image Segmentation	19
4.1.1	Fully Convolutional Networks	19
4.1.2	U-Net-based Segmentation Models	20
4.1.3	Mask R-CNN	22
4.1.4	Transformer	23
4.2	Survival Analysis	24
5	Methodology	26

CONTENTS

5.1	Overview of the Proposed Framework	26
5.2	Lung Tumor Segmentation	27
5.2.1	Network's Architecture	28
5.2.2	Loss Function	30
5.3	Radiomics-Based Feature Extraction	31
5.3.1	Feature Extraction using SimpleITK	31
5.3.2	Selected Features and Their Significance	32
5.4	Deep Learning-Based Feature Fusion	33
5.4.1	Image Branch: ResNet-Based Feature Extractor	34
5.4.2	Clinical Branch: Multi-Layer Perceptron	36
5.4.3	Feature Fusion via Concatenation	37
5.5	Survival Analysis Model	38
6	Experiments and Results	41
6.1	Dataset Analysis and Preprocessing	41
6.1.1	Dataset	41
6.1.2	Preprocessing	44
6.2	Experiment Setup	46
6.2.1	Tumor Segmentation Experiments	46
6.2.2	Survival Analysis Experiments	46
6.2.3	Evaluation Metrics	47
6.3	Results	49
6.3.1	Tumor Segmentation Performance	49
6.3.2	Survival Prediction: CT vs Clinical vs Combined Models	51
6.3.3	Primary Method vs Proposed Enhancement	53
7	Conclusion and Discussion	55
7.1	Achieved Results	55
7.2	Limitations	56
7.3	Future Works	56
References		57

List of Tables

5.3.1	Selected radiomic features extracted from segmented tumor regions and their descriptions	33
5.4.1	Slice selection strategy based on the number of tumor-containing slices	36
6.1.1	Summary of NSCLC Radiogenomics's data types	42
6.1.2	Summary of demographic and clinical cohort characteristics	43
6.3.1	Quantitative comparisons of segmentation performance of proposed model with other deep learning techniques	50
6.3.2	C-index comparison of survival models using CT, clinical, and combined features	51
6.3.3	C-index comparison between primary method and proposed enhancement .	54

List of Figures

3.2.1	Image classification	10
3.2.2	Object recognition in an urban environment with YOLOv7	10
3.2.3	Semantic segmentation	11
3.2.4	Instance segmentation	11
3.2.5	Image generation	12
4.1.1	Fully convolutional network ^[26]	20
4.1.2	U-Net architecture ^[29]	21
5.2.1	nnU-Net automated method configuration ^[19]	29
5.2.2	nnU-Net cascade architecture ^[19]	30
5.3.1	Radiomics-based feature extraction workflow	32
5.4.1	Deep learning-based feature fusion	34
5.4.2	Distribution of patients by tumor slice count	35
5.4.3	Architecture of the MLP used for extracting clinical feature vectors	37
5.5.1	DeepSurv architecture ^[24]	40
6.3.1	Comparison of tumor segmentation results across models	50
6.3.2	Comparison of distributional fits to the survival data	52

List of Abbreviations

AFT Accelerated Failure Time.

AGs Attention Gates.

AI Artificial Intelligence.

C-Index Concordance Index.

CNN Convolutional Neural Network.

CNNs Convolutional Neural Networks.

Cox PH Cox Proportional Hazards.

CT Computed Tomography.

CV Computer Vision.

DICOM Digital Imaging and Communications in Medicine.

DL Deep Learning.

DSI Dice Similarity Index.

FCN Fully Convolutional Network.

FCNs Fully Convolutional Networks.

GANs Generative Adversarial Networks.

GDPR General Data Protection Regulation.

GPUs Graphics Processing Units.

HIPAA Health Insurance Portability and Accountability Act.

LIST OF ABBREVIATIONS

HR Hazard Ratio.

HU Hounsfield Units.

IoU Intersection over Union.

ML Machine Learning.

MLP Multilayer Perceptron.

MRI Magnetic Resonance Imaging.

NIfTI Neuroimaging Informatics Technology Initiative.

NLP Natural Language Processing.

NSCLC Non-Small Cell Lung Cancer.

PCA Principal Component Analysis.

PET Positron Emission Tomography.

R-CNN Region-based Convolutional Neural Network.

RNA Ribonucleic Acid.

RNNs Recurrent Neural Networks.

RPN Region Proposal Network.

SA Survival Analysis.

SGD Stochastic Gradient Descent.

SVM Support Vector Machines.

ViT Vision Transformer.

WHO World Health Organization.

Chapter 1

Introduction

This section provides an overview of the research topic, current advances, and challenges within the field, along with a clear definition of the problem addressed in this project. It outlines the research purpose and significance, explaining the necessity for improved diagnostic methods in medical imaging, particularly for cancer detection and analysis. This foundation enables me to set specific goals to achieve within the scope of this specialized project.

1.1 Introduction

Cancer remains one of the leading causes of death globally, with the number of new cases and fatalities rising annually. According to 2022 statistics from GLOBOCAN^[3], supported by the World Health Organization (WHO), cancer was responsible for almost one in six deaths (16.8%) worldwide. In that year alone, an estimated 20 million new cases were diagnosed, resulting in 9.7 million deaths due to cancer. Specifically, lung cancer emerges as a major concern, being the most commonly diagnosed cancer worldwide, comprising 12.4% of all cancer cases. In general, the rate of patients diagnosed with and dying from cancer is increasing rapidly, posing a significant challenge to the healthcare system and increasing the need for effective diagnostic and treatment methods.

In response to this pressing need, early diagnosis and timely intervention are among the most critical factors in improving patient outcomes, particularly for rapidly progressing cancer types. In this context, medical imaging plays a crucial role in the detection and assessment of cancer. Diagnostic imaging methods such as Positron Emission Tomography (PET), Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) have become indispensable tools for locating, sizing, and staging tumors. However, accurately analyzing and processing medical images by eye requires a high level of expertise and is

1.2. Objectives

often time-consuming.

Survival Analysis (SA) is a set of statistical and analytical methods that estimate the time until a particular event occurs for an individual or group in a study. In cancer treatment, SA is often used to predict a patient's survival time based on their current health status. This approach is highly valuable, as it assists doctors in making accurate diagnoses and determining appropriate treatment options for patients. While survival analysis methods based on tabular data have been widely studied for a long time, extending survival analysis to more complex data types, such as medical imaging (CT, PET, MRI), remains relatively new. Notably, recent studies reveal significant improvements in survival analysis applications on image data, driven by advancements in deep learning. This progress opens up new possibilities for combining survival analysis with medical imaging data, enabling more accurate and effective predictions to support diagnosis and treatment planning for patients.

Therefore, this study aims to develop a SA model for cancer patients by integrating data from medical images. Leveraging information from medical imaging allows the model to access and process detailed insights on tumor location, size, and characteristics, providing a more comprehensive and accurate assessment. It is hoped that this project will serve as a foundation for developing a fully functional tool to support doctors in making treatment decisions and planning patient care more effectively.

1.2 Objectives

The goal of this project is to enhance the timeliness and effectiveness of both the diagnosis and treatment processes for cancer. By leveraging the advancements and effectiveness of deep learning, the project aims to segment tumor images from medical imaging data, extracting key features of the tumor, which will serve as important input for the survival analysis model. To achieve this goal, the project will follow the following objectives:

- Study the foundational knowledge of SA in various fields.
- Review the knowledge of machine learning and deep learning techniques.
- Review related works and research on the application of SA for cancer patients using medical imaging data.
- Investigate, collect, and preprocess data for the problem.
- Implement and evaluate key works relevant to the approach.

1.3. Scope

- Propose, test, and perform evaluations of potential solutions.

1.3 Scope

This project focuses on researching and developing deep learning-based methods for medical image analysis, specifically for tumor segmentation and survival prediction in lung cancer patients. The specific scope of the project includes:

Type of cancer studied:

- The study is limited to the analysis and processing of medical images related to lung cancer, as it is a highly prevalent type of cancer with significant importance in diagnosis and treatment. The results of this research can potentially be expanded to other types of cancer in the future.

Model Usage and Training:

- The project utilizes a specific dataset that includes medical images with annotated tumor masks and accompanying clinical data.
- A deep learning model will be trained on this dataset to generate tumor segmentation results.
- Once the model is trained, it may be applied to new datasets, which may not have available tumor masks but contain sufficient input data for survival analysis.

Boundaries of the Project:

- Data: Access to medical datasets is challenging due to privacy concerns and the sensitive nature of patient information. This limitation affects the ability to acquire large and diverse datasets, which is crucial for training and validating deep learning models.
- Cancer Type: The project is solely focused on lung cancer, and thus the research findings may not directly apply to other types of cancer.
- Model Scope: The development is limited to two main stages: tumor segmentation and survival analysis. Other aspects, such as cancer detection or the integration of non-imaging data (e.g., genomic data), are beyond the scope of this study.

1.4. Report Outline

1.4 Report Outline

The outline of our specialized project is organized as follows:

- **Chapter 1: Introduction**

Provide a comprehensive introduction to the global cancer situation, the need for the project, the objective, and the contribution to the project.

- **Chapter 2: Problem Requirements Analysis**

Presents the challenges and difficulties of the problem in the current era.

- **Chapter 3: Related Works**

Discuss the methods and results of works related to the task of medical image segmentation and survival analysis.

- **Chapter 4: Theoretical Background**

Provide an overview of relevant knowledge, particularly focusing on the theory behind machine learning and deep learning models applied in computer vision, as well as the fundamentals of survival analysis in the medical field.

- **Chapter 5: Methodology**

Propose some detailed methods and an overall model to solve the problem.

- **Chapter 6: Experiments and Results**

Give a summary of the datasets, describe the setup for the experimental methodology, and present the evaluation outcomes from various experiments.

- **Chapter 7: Conclusion and Discussion**

Summarize the results achieved, the limitations, and potential areas for further improvement.

Chapter 2

Problem Requirements Analysis

This chapter presents the challenges and difficulties of the problem in the current era. Additionally, the specific requirements will also be discussed.

Lung cancer remains one of the leading causes of cancer-related deaths worldwide, making accurate survival prediction a crucial task in modern healthcare. However, this problem presents significant challenges in today's context. One major difficulty lies in the availability and quality of data. While medical imaging datasets are available, they often lack standardization, exhibit significant variability across imaging devices and institutions, or contain noise and artifacts. These issues complicate the preprocessing and normalization steps required for model training. Furthermore, integrating imaging data, such as tumor segmentation outputs, with clinical data is a complex task due to the differences in data types and scales. The lack of complete clinical information makes this challenge even harder.

Deep learning models must not only deliver high accuracy but also be interpretable and reliable enough for clinical applications. Inaccuracies in the segmentation phase can significantly impact survival prediction while overfitting remains a persistent concern due to the limited size of medical datasets. Computational demands also pose a challenge, as training 3D models for medical imaging requires substantial hardware resources, which may not always be accessible. Moreover, ethical and regulatory constraints, such as ensuring patient privacy and complying with standards like General Data Protection Regulation (GDPR) or Health Insurance Portability and Accountability Act (HIPAA), add another layer of complexity to the problem.

To address these challenges, the problem requires high-quality annotated datasets that include medical images (e.g., CT scans) and accompanying clinical data, such as patient demographics and medical history. Robust deep-learning architectures must be developed to perform accurate lung tumor segmentation and integrate extracted features with

clinical data for survival analysis. Evaluation metrics like the Dice similarity coefficient for segmentation and Concordance index for survival prediction are necessary to validate model performance.

Finally, adherence to privacy and ethical guidelines is essential, along with ensuring scalability and reproducibility of the proposed solutions. By overcoming these challenges and meeting these requirements, it is possible to develop innovative and clinically relevant approaches to improve lung cancer survival prediction.

Chapter 3

Theoretical Background

This chapter provides an overview of the concepts and foundations of machine learning and deep learning, along with the essential theories of survival analysis. Since this project focuses on medical images, including PET, CT, and MRI scans.

3.1 Overview of Machine Learning and Deep Learning

3.1.1 Machine Learning

Machine Learning (ML) is a significant branch of Artificial Intelligence (AI) that focuses on developing algorithms and models enabling computers to learn and make decisions based on data without requiring explicit programming. Instead of relying on fixed rules coded manually, ML leverages data to identify patterns and relationships, allowing it to make predictions or perform specific tasks. This capability makes ML a powerful tool for solving complex problems, particularly when dealing with large and diverse datasets, such as those found in the medical field.

Supervised Learning

Supervised learning involves training a model on labeled data, where each input is paired with its corresponding output. The model learns to predict outputs for new inputs by minimizing the error between its predictions and the actual labels during training. Common algorithms for supervised learning include Linear Regression, Logistic Regression, Support Vector Machines (SVM), Decision Trees, Random Forests, and Neural Networks. For instance, in medical imaging, supervised learning is frequently used for

3.1. Overview of Machine Learning and Deep Learning

cancer classification, where models are trained on labeled datasets of images to distinguish between "cancerous" and "non-cancerous" cases.

Unsupervised Learning

Unsupervised Learning deals with unlabeled data and focuses on identifying hidden patterns or structures within the dataset. This type of learning is particularly useful for clustering, association, or dimensionality reduction tasks. Algorithms like K-Means Clustering, Hierarchical Clustering, Principal Component Analysis (PCA), and Autoencoders are commonly used in this domain. In medical imaging, unsupervised learning can be applied to cluster CT scans based on anatomical similarities or tumor features, helping to uncover patterns that might not be immediately apparent.

Reinforcement Learning

Reinforcement learning involves training an agent to make decisions by interacting with an environment and receiving feedback through rewards or penalties. While it is less commonly applied in medical imaging, it can be used in areas such as optimizing treatment plans or controlling robotic-assisted surgery.

3.1.2 Deep Learning

Deep Learning (DL), a subset of Machine Learning, has gained tremendous attention in recent years due to its ability to model and solve complex problems, particularly those involving large-scale and unstructured data such as images, audio, and text. Unlike traditional ML, which often requires manual feature extraction, DL automates this process by using artificial neural networks with many layers—hence the term "deep." These deep networks can learn hierarchical representations of data, where lower layers capture basic features (such as edges in images) and deeper layers combine these features to recognize more abstract patterns (such as objects or facial features). This ability to automatically learn complex patterns has made DL the backbone of many breakthroughs in various fields, including Computer Vision (CV), Natural Language Processing (NLP), and speech recognition.

In medical imaging, for example, deep learning models like Convolutional Neural Networks (CNNs) have revolutionized the way diseases are diagnosed, enabling machines to outperform human experts in tasks like tumor detection and organ segmentation. The advancement of DL is largely driven by the availability of large datasets, powerful

computational resources (such as Graphics Processing Units (GPUs)), and improvements in neural network architectures and training algorithms.

As a result, DL has moved from a research-focused field to practical applications with significant real-world impact, particularly in areas such as healthcare, autonomous driving, and entertainment.

3.2 Computer Vision

Computer Vision is a field of AI that leverages machine learning and neural networks to enable computers and systems to extract meaningful information from digital images, videos, or other types of visual inputs. If AI gives computers the ability to learn and think, then computer vision provides them with the ability to see, observe, and understand.

CV operates similarly to human vision, but humans have the advantage of years of experience to recognize objects, judge distances, detect movement, and identify anomalies in images. In contrast, computer vision equips machines to perform these tasks using cameras, data, and algorithms instead of biological components like retinas, optic nerves, and the visual cortex. However, with the ability to analyze thousands of products or processes per minute and detect even the smallest defects or issues, computer vision systems can quickly outperform human capabilities.

There is a lot of research being conducted in the field of CV, but its impact does not stop there. Real-world applications highlight the vital role CV plays in industries such as business, entertainment, transportation, healthcare, and daily life. A key factor driving the expansion of these applications is the flood of visual data generated by smartphones, security systems, traffic cameras, and other visually-equipped devices.

Therefore, here are a few examples of computer vision tasks:

- **Image Classification:** This task involves analyzing an image and predicting its overall category or class from a predefined set. The classification considers the entire image as a whole without focusing on specific regions or objects within it. It is one of the fundamental tasks in computer vision, often used as a building block for more complex applications. For instance, figure 3.2.1 will be classified as a tiger.
- **Object Detection and Localization:** This task goes beyond image classification by not only identifying the types of objects in an image but also pinpointing their positions. It uses bounding boxes to outline each detected object, allowing the model to recognize multiple objects of various classes in a single image. Object detection

3.2. Computer Vision



Figure 3.2.1: Image classification

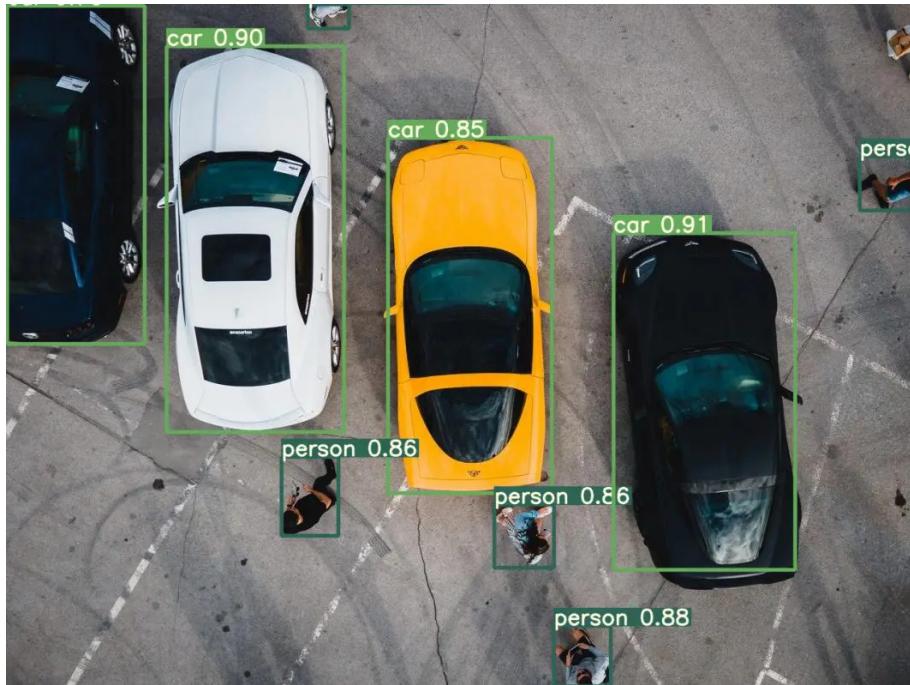


Figure 3.2.2: Object recognition in an urban environment with YOLOv7

is essential in applications requiring detailed spatial awareness, such as autonomous driving or surveillance.

- **Semantic Segmentation:** This task involves assigning a class label to every pixel in an image, effectively dividing it into regions that correspond to different categories. Unlike object detection, semantic segmentation does not differentiate between instances of the same class but focuses on capturing the overall structure

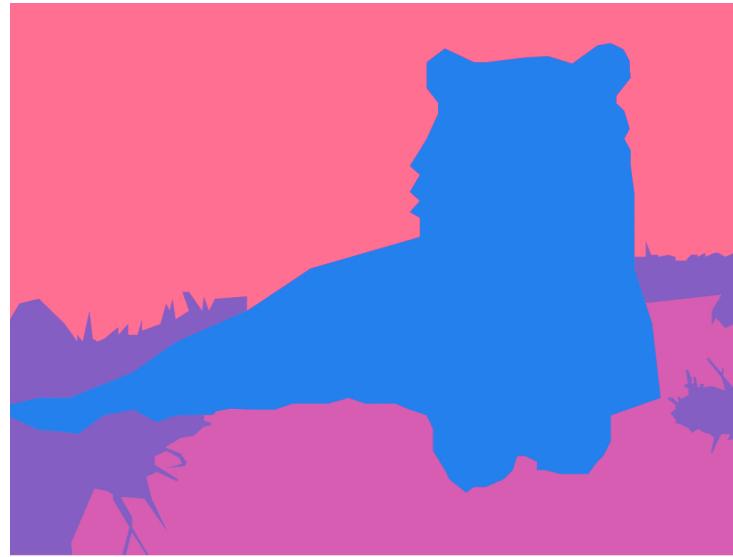


Figure 3.2.3: Semantic segmentation



Figure 3.2.4: Instance segmentation

and context of the image. It is commonly used in medical imaging, self-driving cars, and scene understanding.

- **Instance Segmentation:** This task combines the benefits of semantic segmentation and object detection, providing pixel-level accuracy while distinguishing between individual instances of the same class. Each instance is treated as a separate entity, making it ideal for applications that require precise object differentiation, such as counting objects or analyzing interactions in an image.

3.3. Survival Analysis Fundamentals

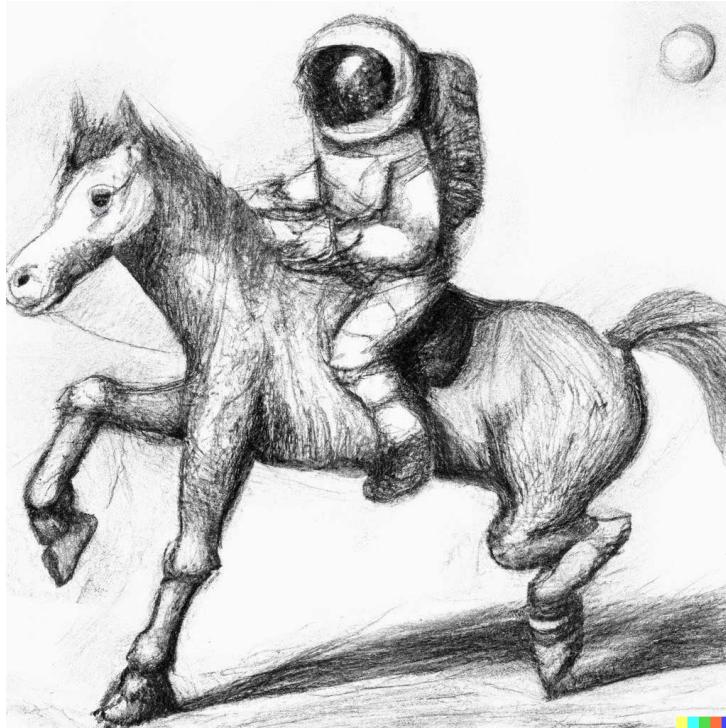


Figure 3.2.5: Image generation

- **Image Generation:** This task uses deep learning techniques to synthesize new images that resemble real ones, based on the patterns and features learned from training data. It can generate images in various contexts, such as recreating missing parts of an image, designing artwork, or generating entirely new content from textual descriptions. Advances in generative models like Generative Adversarial Networks (GANs) and diffusion models have significantly improved the quality and realism of generated images.

This project primarily focuses on the task of instance segmentation, specifically lung tumor segmentation using medical images, to extract meaningful information for survival analysis.

3.3 Survival Analysis Fundamentals

Survival Analysis is a group of statistical methods used to analyze time-to-event data until a significant event occurs. In the field of medicine, the significant event is often the time of "death". Additionally, SA can be applied to various fields, such as the time to bankruptcy in finance, failure, or breakdown of machinery and systems.

In many medical studies, time to death is an event of interest. However, in this project

3.3. Survival Analysis Fundamentals

and any kind of similarity of cancer studies, another key measure is the duration between treatment response and the occurrence of recurrence or relapse-free survival time (also known as disease-free survival time). It is important to define the event of interest and specify the start and end points of the observation period. For instance, the focus might be on relapse occurring within the period between a confirmed treatment response and the first relapse of cancer.

In the article introducing the basic concepts of SA by TG Clark^[7], several fundamental terms and concepts in the field of SA are presented. Among them, one of the most crucial concepts in SA is survival time. Survival time refers to the period from the start of the study or observation until a significant event occurs (e.g., from the start of treatment to the time of disease recurrence).

3.3.1 Censoring Data

A common challenge in SA is the fact that during the analysis process, some groups of individuals may have experienced the significant event, while others have not, leading to unknown survival times for those groups. This phenomenon is known as censoring. There are several reasons leading to this problem:

- Individuals do not experience the significant event by the end of the study.
- Individuals are lost to follow-up during the study.
- Individuals experience another event that prevents the continuation of the observation.

Additionally, there are three types of censoring, which are **right censoring**, **left censoring**, and **interval censoring**.

- **Right censoring:** when a significant event has not occurred for an individual by the time the study or observation ends, or when the individual is lost to follow-up. The significant event may occur at a later time, but the exact timing is unknown.
- **Left censoring:** when a significant event occurs before a certain time, but the exact timing of the event is unknown. For example, a patient may have had the disease before participating in the study, but the exact time of diagnosis is unknown.
- **Interval censoring:** when an event occurs within a certain period, but the exact time of occurrence is unknown. For example, if a patient is examined periodically every 6 months and has no symptoms during the first examination but experiences a

3.3. Survival Analysis Fundamentals

relapse during the second, it is known that the recurrence occurred around 6 months after the first examination, although the exact timing remains unclear.

In general, most cases are right-censored observations, but methods for interval and left-censored data are also available, as discussed by Hosmer and Lemeshow^[18]. In this project, our primary focus is on right-censored data only.

3.3.2 Survival and Hazard

Survival data are typically described and modeled using two related probabilities: survival and hazard. Survival probability (which is also called survival function) $S(t)$ calculates the probability that an individual survives from the time origin (e.g. diagnosis of cancer) to the specific future time t . This function provides an overview of survival rates within the study cohort.

The hazard (denoted by $h(t)$ or $\lambda(t)$) is the probability of an event occurring at time t , conditional on the individual having survived up to that point. This function indicates the probability of an event occurring at a specific time, unlike the survival function, which focuses on the immediate risk instead of the overall probability.

3.3.3 Hazard Ratio

The Hazard Ratio (HR) is a key concept in survival analysis, particularly in models like the Cox Proportional Hazards (Cox PH) model. It is a measure used to compare the risk (or hazard) of an event occurring at any point in time between two groups, often used in clinical studies or time-to-event analysis.

The HR is the ratio of the hazard rates between two groups. It compares the risk of the event occurring in one group relative to another while accounting for the time of the event. The hazard ratio is calculated as below.

$$HR = \frac{\text{Hazard In Group 1}}{\text{Hazard In Group 2}}$$

After computing the hazard ratio, the interpretation is described as follows:

- $HR = 1$: There is no difference in the hazard between the two groups. Both groups have the same risk of the event occurring.
- $HR > 1$: The hazard (risk) is higher in the first group compared to the second. For example, if $HR = 2$, the first group has twice the risk of the event occurring compared to the second group.

3.3. Survival Analysis Fundamentals

- $\text{HR} < 1$: The hazard (risk) is lower in the first group compared to the second. For example, if $\text{HR} = 0.5$, the first group has half the risk of the event occurring compared to the second group.

The hazard ratio provides a clear understanding of the relationship between the covariate(s) and the risk of the event. Clinical studies, for example, allow researchers to quantify the effect of a treatment or intervention on survival outcomes. Clinical Example: In a study comparing two cancer treatments, if the hazard ratio for treatment A versus treatment B is 0.7, this suggests that patients receiving treatment A have a 30% lower risk of the event (e.g., death) compared to those receiving treatment B.

3.3.4 Univariate Analysis

Univariate analysis in survival analysis focuses on evaluating the relationship between a single independent variable and survival time. It aims to summarize survival data, estimate survival probabilities, and assess differences between groups. Common techniques include the Kaplan-Meier estimator and the log-rank test, which are widely used to analyze time-to-event data.

Kaplan-Meier method

The Kaplan-Meier estimator^[23], which is named after Edward L. Kaplan and Paul Meier, is a non-parametric method for estimating the survival function from time-to-event data. It provides a stepwise survival curve that represents the probability of survival over time.

Suppose in the period of the study, there are k patients who have events at distinct time $t_1 < t_2 < t_3 < t_4 < \dots < t_{k-1} < t_k$, as t_k is the time that events occur for patient k . These events are assumed to be independent of one another, so the probabilities of surviving from one interval to the next may be multiplied to calculate the cumulative survival probability. The estimator of the survival function at the time t_j (the likelihood of being alive up to time t_j) is given below.

$$S(t_j) = S(t_{j-1}) \left(1 - \frac{d_j}{n_j}\right)$$

where,

- $S(t_{j-1})$ is the probability of being alive up to time $j - 1$.
- d_j is the number of events that occur at the time j .

3.3. Survival Analysis Fundamentals

- n_j is the number of patients that survive at the time j .
- $t_0 = 0$ and $S(0) = 1$.

The Kaplan-Meier curve graphically represents the survival rate against time. A steeper curve suggests a higher event rate (death rate), indicating a worse survival outcome, while a flatter curve signifies a lower event rate and a better survival outlook. Plateaus or flat sections on the curve indicate periods of relatively stable survival. When multiple curves correspond to different groups, their shapes and patterns can be compared to analyze differences in survival trends. In this case, a log-rank test can be used to compare different groups.

Log-rank test

The log-rank test^[28] is a statistical method used to compare the survival distributions of two groups. This nonparametric test is suitable for data that are censored, provided the censoring is non-informative. It is commonly applied in clinical trials to evaluate the effectiveness of a new treatment compared to a control, with the outcome measured as the time to an event. The test is designed to:

- Compare survival curves derived from groups (e.g., patients receiving different treatments).
- Test the null hypothesis that the survival distributions of the groups are identical.

At each event time, the method determines the expected number of events for each group, assuming no differences exist between them. These expected values are then summed across all event times to calculate the total expected number of events for each group, denoted as E_i for group i . The log-rank test evaluates the difference by comparing the observed number of events O_i in treatment group i to the expected number, using this comparison to compute the test statistic.

$$X^2 = \sum_{i=1}^g \frac{(O_i - E_i)^2}{E_i}$$

This value is compared to a χ^2 distribution with $(g - 1)$ degrees of freedom, where g is the number of groups. P-value may be computed to calculate the statistical significance of the differences between the survival curves of different groups.

- **Null hypothesis:** Both groups have identical distribution curves.

- **Alternative hypothesis:** Both groups have different distribution curves.

If the calculated p-value exceeds 0.05 (in most cases, the significance level is set at this value), the null hypothesis cannot be rejected, indicating that the two groups are assumed to have the same distribution curve based on the available data.

In contrast, if the p-value is less than 0.05, the null hypothesis is rejected, suggesting that the two groups are significantly different.

3.3.5 Covariate Analysis

Covariate analysis in survival analysis aims to explore the influence of one or more covariates (independent variables) on survival time. Unlike univariate analysis, which examines the effect of a single variable, covariate analysis allows for the inclusion of multiple factors that may influence survival.

Two common approaches for covariate analysis are the Cox PH model and the Accelerated Failure Time (AFT) model.

Cox proportional hazard model

The Cox PH model^[9], introduced by D. R. Cox in 1972, is one of the most widely used methods for modeling survival data with covariates. It is a semi-parametric model that assesses the effect of covariates on the hazard function (the risk of an event occurring at a given time), without making assumptions about the underlying baseline hazard function.

The model assumes that the effect of a covariate is multiplicative on the hazard function, i.e., the hazard ratio for an individual's risk remains constant over time. Mathematically, the Cox model is written as

$$h(t) = h_0(t) \times \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_p X_p)$$

where,

- $h(t)$ is the hazard at time t for a set of covariates X .
- $h_0(t)$ is the baseline hazard function (unknown and non-parametric).
- $\beta_1, \beta_2, \dots, \beta_p$ are the regression coefficients, indicating the influence of covariates X_1, X_2, \dots, X_p .

One attractive aspect of the Cox model is that it estimates the baseline hazard function non-parametrically. This means that, unlike many other statistical models, it does not assume that survival times follow a specific statistical distribution.

3.3. Survival Analysis Fundamentals

The hazard ratio is the exponential of the regression coefficient β for a covariate, and it quantifies the effect of the covariate on the hazard function. If the covariate is continuous (e.g., age, dose of a drug), the hazard ratio represents the change in hazard for a one-unit increase in that covariate. If the covariate is categorical (e.g., treatment group), the hazard ratio compares the risk between different categories. The hazard ratio for a covariate X_j is given as below.

$$HR = \exp(\beta_j)$$

For example, if $\beta_1 = 0.3$, the hazard ratio $HR = \exp(0.3) \approx 1.35$, meaning that for each one-unit increase in X_1 , the hazard increases by 35%.

Accelerated Failure Time model

The Accelerated Failure Time model is another approach to survival analysis that directly models the survival time, rather than the hazard function. The AFT model assumes that the effect of covariates accelerates or decelerates the failure time, hence "accelerating" or "slowing down" the event occurrence, meaning that certain factors speed up or slow down the time to event. The model assumes the survival time T follows a distribution (e.g., Weibull, Exponential, Log-normal), and that covariates multiply the survival time by a constant factor.

$$\log(T) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

where,

- T is survival time.
- X_1, X_2, \dots, X_p are the covariates.
- α is constant term.
- ϵ is error term.

Chapter 4

Related Works

The purpose of this chapter is to review and evaluate the state-of-the-art methods and architectures in the field of image segmentation, with a particular focus on tumor segmentation. Additionally, it examines existing works related to survival analysis, especially those incorporating medical imaging data.

4.1 Image Segmentation

Image segmentation, a fundamental task in computer vision, involves partitioning an image into meaningful regions to facilitate the analysis of specific structures or objects. Numerous studies have been conducted in the field of image segmentation, particularly in medical imaging, achieving significant advancements and providing a solid foundation for future segmentation models.

4.1.1 Fully Convolutional Networks

Fully Convolutional Networks (FCNs)^[26], introduced by Long et al. in 2015, represent a seminal advancement in the field of image segmentation. Unlike traditional CNNs, which rely on fully connected layers for classification, FCNs replace these layers with convolutional operations, enabling the network to output dense pixel-wise predictions instead of scalar values. This innovation made FCNs the first deep learning architecture specifically tailored for semantic segmentation tasks.

FCNs are designed to process images of arbitrary size and produce segmentation maps of corresponding dimensions, enabling end-to-end learning. They also use deconvolution (or transpose convolution) layers, which allow the network to upsample feature maps to the resolution of the input image, ensuring accurate localization.

4.1. Image Segmentation

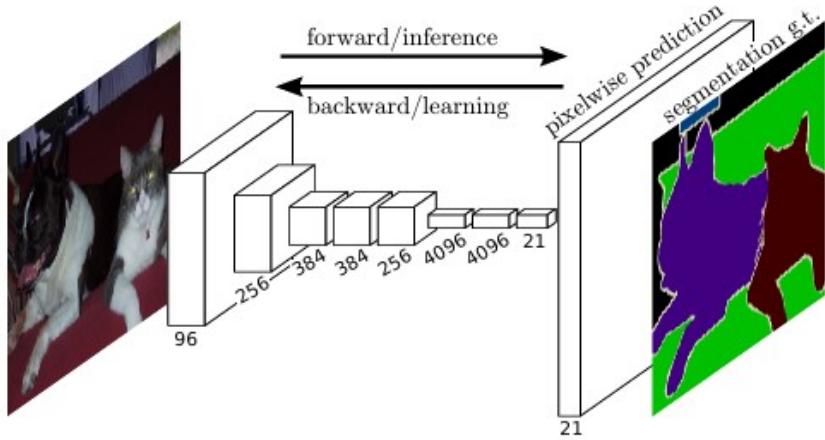


Figure 4.1.1: Fully convolutional network^[26]

FCNs have been widely adopted for segmenting anatomical structures in medical images, including organs, lesions, and tumors. They form the foundation for many subsequent models, such as U-Net, which enhances the Fully Convolutional Network (FCN) architecture by adding skip connections to preserve spatial details better.

Although FCNs mark a significant milestone, their limitations, such as the loss of fine details due to excessive downsampling and limited contextual understanding, have been addressed by later models. Nonetheless, they remain an essential baseline for evaluating segmentation performance.

4.1.2 U-Net-based Segmentation Models

U-Net^[29] has become one of the most widely used and successful architectures in medical image segmentation. Introduced by Ronneberger et al. in 2015, U-Net is a FCN designed for semantic segmentation. It stands out for being relatively simple and efficient while still achieving excellent performance on various medical image segmentation tasks. However, U-Net may struggle with more complex or varied input images, and in such cases, alternative methods could be more suitable.

The U-Net architecture is distinguished by its U-shaped structure, which consists of two main parts:

- **Contracting path (Encoder):** This part reduces the spatial dimensions of the image, while also capturing relevant information about the image.
- **Expanding path (Decoder):** This part upsamples the feature map to produce a relevant segmentation map, leveraging the patterns learned in the contracting path.

4.1. Image Segmentation

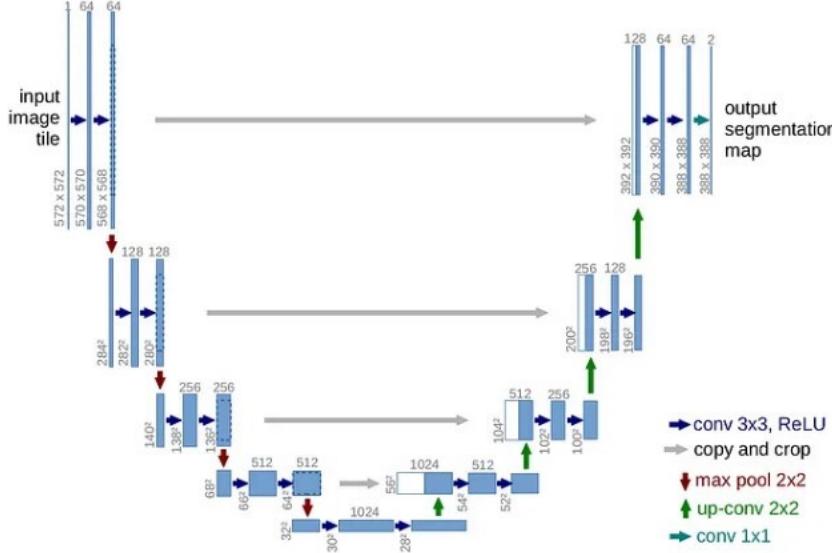


Figure 4.1.2: U-Net architecture^[29]

A key feature of U-Net is its use of skip connections, which link corresponding layers in the encoder and decoder. These connections enable the model to transfer fine-grained details from the encoder to the decoder, ensuring that spatial information is not lost during the downsampling process.

3D U-Net^[6] extends the original U-Net architecture to three-dimensional data, making it particularly well-suited for medical imaging tasks involving volumetric data, such as CT, MRI, and PET scans. Unlike 2D U-Net, which processes individual slices, 3D U-Net works directly on volumetric inputs, capturing spatial context along all three dimensions. This improvement enables better segmentation of structures that span across slices. The architecture mirrors the U-shaped structure of the original U-Net but incorporates 3D convolutional, pooling, and upsampling layers. While 3D U-Net significantly enhances segmentation performance on volumetric datasets, it comes with increased computational cost and memory requirements, which can be challenging for large-scale or high-resolution inputs.

Attention U-Net^[27] introduces attention mechanisms into the U-Net architecture to focus on the most relevant regions in an image for segmentation. This model integrates Attention Gates (AGs), which learn to suppress irrelevant background features and highlight salient regions of interest, improving segmentation accuracy. The AGs work by combining contextual information from lower-resolution layers with finer details from higher-resolution layers, allowing the model to adaptively refine its focus. Attention U-Net has been particularly effective in cases where lesions or abnormalities are small and

4.1. Image Segmentation

difficult to detect, such as in liver or brain tumor segmentation. By emphasizing critical areas in the feature map, it enhances both precision and robustness in challenging medical imaging tasks.

U-Net++^[33] refines the original U-Net by introducing a redesigned skip connection mechanism to better bridge the encoder and decoder. Instead of direct connections, U-Net++ employs nested and dense skip pathways, enabling better information flow and improving gradient propagation during training. These nested connections help the decoder gradually recover spatial information, resulting in more precise segmentation. U-Net++ has demonstrated superior performance in tasks like tumor segmentation, outperforming the original U-Net and Attention U-Net on several benchmark datasets. However, the increased architectural complexity comes with a higher computational cost, which may require careful resource allocation.

DUCK-Net^[11] is an advanced encoder-decoder model designed for robust polyp segmentation in colonoscopy images. It addresses the challenges posed by varying polyp characteristics with a unique DUCK block, which uses six parallel convolutional variations, allowing the network to optimize the best configuration during training. A key feature is its secondary U-Net-like downscaling layer that preserves low-level image details by avoiding direct image processing, ensuring fine detail retention. The combination of custom convolutional blocks and residual downsampling makes DUCK-Net effective in achieving accurate and reliable segmentation results.

4.1.3 Mask R-CNN

Mask Region-based Convolutional Neural Network (R-CNN)^[15], introduced by He et al. in 2017, is an extension of the Faster R-CNN framework, designed to perform pixel-level instance segmentation. This model adds a branch to predict segmentation masks for each detected object, making it capable of simultaneous object detection and semantic segmentation. The architecture consists of a backbone network, typically ResNet or ResNeXt, for feature extraction, a Region Proposal Network (RPN) to generate candidate object regions, and separate branches for classification, bounding box regression, and mask prediction.

One of the key innovations in Mask R-CNN is the use of a RoIAlign operation instead of the traditional RoIPool. This modification improves the spatial precision of object regions, enabling more accurate mask predictions. Mask R-CNN has demonstrated strong performance across various segmentation tasks, including medical imaging.

In medical applications, Mask R-CNN has been successfully utilized for tasks such as tumor localization and organ segmentation. Its ability to separate overlapping structures

4.1. Image Segmentation

and provide instance-specific segmentation makes it particularly valuable in complex scenarios, such as detecting multiple lesions or segmenting anatomically adjacent organs. However, its reliance on region proposals and complex multi-stage processing can make it computationally intensive, requiring considerable hardware resources for training and inference. Despite these challenges, Mask R-CNN remains a highly influential model for instance, in segmentation tasks in medical imaging.

4.1.4 Transformer

While fully convolutional network-based models and U-Net-based architectures have achieved strong results in segmentation tasks, particularly in medical image segmentation, transformer-based approaches have also demonstrated high effectiveness in this domain. Transformer models initially introduced in NLP through the groundbreaking "Attention is All You Need"^[32] paper, introduced by Vaswani, A in 2017, have quickly been adapted to Computer Vision tasks. The transformer architecture leverages the self-attention mechanism to model global relationships between pixels in an image, making it particularly well-suited for segmentation tasks that demand high precision.

Swin U-Net^[4] combines the hierarchical design of U-Net with the Swin Transformer architecture, enabling the modeling of long-range dependencies and multi-scale feature representation. Unlike traditional CNNs, Swin U-Net utilizes shifted window attention mechanisms in both encoder and decoder, ensuring efficient computation and improved feature extraction. The architecture excels in tasks requiring precise boundary delineation and multi-resolution analysis, making it suitable for segmenting organs and tumors in CT and MRI scans. Swin U-Net has demonstrated superior performance compared to traditional U-Net models, especially in datasets requiring fine-grained segmentation.

Trans U-Net^[5] integrates the strengths of transformers and CNNs, utilizing a hybrid approach to leverage the spatial localization capabilities of CNNs and the global contextual understanding of transformers. In Trans U-Net, a pre-trained Vision Transformer (ViT) is employed as the encoder, while the decoder remains CNN-based with skip connections akin to the original U-Net. This hybrid design effectively captures both local and global information, making it particularly effective for medical imaging tasks like organ segmentation and tumor localization. Trans U-Net has set benchmarks on multiple medical imaging datasets, outperforming CNN-only architectures in terms of segmentation accuracy.

The UNETR^[14] model, short for U-Net with Transformers, is a pure transformer-based architecture specifically designed for 3D medical image segmentation. Unlike hybrid approaches, UNETR replaces convolutional encoders entirely with a transformer encoder,

4.2. Survival Analysis

enabling a direct mapping of volumetric data into latent representations. The decoder retains the hierarchical structure of U-Net to reconstruct the segmentation mask from coarse to fine levels. By directly processing 3D inputs and modeling long-range dependencies across spatial dimensions, UNETR achieves remarkable performance in 3D tasks, such as brain tumor segmentation and cardiac segmentation, often outperforming hybrid transformer-CNN models.

MedT (Medical Transformer)^[31] is a lightweight transformer architecture tailored for medical image segmentation. Unlike conventional transformer models, MedT employs gated axial attention mechanisms, which reduce computational complexity while maintaining the ability to model global and local interactions. The architecture is particularly effective on small medical datasets, as it incorporates techniques to avoid overfitting and ensures generalization. MedT’s simplicity and efficiency make it a preferred choice for resource-constrained environments or applications where computational overhead is a concern. It has shown competitive results in tasks like lesion segmentation and anatomical structure delineation.

4.2 Survival Analysis

Recent advancements in survival analysis have seen the adoption of deep learning techniques to address limitations of traditional models like Cox Proportional Hazards. This section reviews the following models: Cox PH model, DeepSurv, DeepHit, RNN SURV, and DeepConvSurv.

The Cox PH model is a seminal approach in survival analysis that assumes proportional hazards over time. It uses a linear combination of covariates to predict the hazard function. Despite its robustness, its inability to capture non-linear relationships and interactions in high-dimensional data limits its application to complex medical datasets.

DeepSurv^[24] extends the Cox PH model by incorporating deep neural networks to model non-linear relationships between covariates and survival outcomes. By optimizing a partial likelihood loss function, DeepSurv has demonstrated superior performance over Cox PH, particularly in personalized risk predictions and survival analysis in datasets with complex, high-dimensional features.

DeepHit^[25] is a deep learning framework tailored for competing risks and survival analysis. Unlike traditional models, it directly predicts the probability distribution of survival times using a neural network. DeepHit excels in handling multiple risks simultaneously and provides time-dependent survival probabilities, making it ideal for datasets with heterogeneous survival outcomes.

4.2. Survival Analysis

RNN-SURV^[12] applies Recurrent Neural Networks (RNNs) to handle longitudinal survival data with time-varying covariates. Its sequential nature allows it to capture temporal dependencies, offering robust predictions for dynamic patient data, such as electronic health records. RNN-SURV is particularly effective in modeling survival outcomes where covariates change over time.

DeepConvSurv^[34] utilizes CNNs to analyze medical imaging data, such as histopathological and radiological images, for survival analysis. By extracting spatial and texture features from images, DeepConvSurv bridges the gap between imaging biomarkers and survival predictions. This model has proven effective in integrating imaging and clinical data for precision oncology.

Chapter 5

Methodology

This chapter presents an overall architecture for the project and proposes solutions for segmentation tasks and survival analysis.

5.1 Overview of the Proposed Framework

The proposed methodology for predicting the survival of lung cancer patients using medical images and clinical data is structured into a systematic pipeline, ensuring clear and logical processing of the input data to generate meaningful predictions. The overall architecture consists of multiple sequential stages, each designed to process specific data types effectively.

Lung Tumor Segmentation: The preprocessed medical images are input into a segmentation model to localize lung tumors. In this study, we employ the nnU-Net^[19] architecture due to its proven effectiveness in medical image segmentation tasks that do not require extensive manual tuning. The output is a binary segmentation mask, in which tumor regions are delineated from surrounding tissues. This stage is critical as it enables the localization of the tumor, which forms the basis for feature extraction and survival analysis.

Dual Approaches for Feature Extraction and Clinical Integration: In this study, two different approaches are proposed for extracting tumor-related features from the segmentation mask and integrating them with clinical data, each representing a distinct modeling strategy.

- The primary approach, referred to as **Radiomics-Based Feature Extraction**, involves extracting handcrafted features that describe the tumor's shape, size, and intensity statistics (e.g., volume, surface area, elongation, roundness, intensity mean, standard deviation, etc.). These features are then combined in tabular form

5.2. Lung Tumor Segmentation

with clinical data such as age, gender, and cancer-related information for survival prediction.

- The alternative approach, referred to as **Deep Learning-Based Feature Fusion**, leverages deep learning to learn high-level feature representations. Specifically, it uses a Convolutional Neural Network (CNN) (e.g., ResNet^[16]) to encode the tumor region and an Multilayer Perceptron (MLP) to process clinical variables, with both feature vectors subsequently fused into a joint representation.

Survival Analysis: Regardless of the approach used for feature extraction, the resulting features serve as input to a SA model. This stage estimates the patient’s survival probability or risk over time. Both classical statistical models such as Cox PH or AFT, and more advanced models, such as DeepSurv are explored. These models learn the relationship between tumor/clinical features and patient outcomes, producing outputs such as risk scores, survival curves, or time-to-event predictions.

This framework allows for flexibility in model design while ensuring robust survival prediction. By combining image-derived tumor features with clinical variables and supporting both classical and deep learning-based analysis pathways, the proposed pipeline offers a practical and adaptable approach to survival prediction in lung cancer patients.

5.2 Lung Tumor Segmentation

Tumor segmentation in CT scans is a critical component of the proposed framework, as it directly affects the quality of extracted imaging features and, consequently, the survival prediction performance. As discussed in the Related Work section, many deep learning models have been proposed for tumor segmentation, each with its own strengths and limitations.

2D segmentation models such as U-Net, ResUNet++, or DUCK-Net have shown good performance on planar medical images; however, they are inherently limited when applied to 3D volumetric data such as CT or MRI scans. These models treat each slice independently, which results in the loss of inter-slice spatial context.

On the other hand, recent transformer-based models like UNETR and SwinUNet have shown promise in volumetric segmentation. However, these models are computationally expensive and require large-scale training data to generalize well. This makes them less practical for high-resolution images, particularly for lung CT scans where both image resolution and computational demand are high.

5.2. Lung Tumor Segmentation

Given these trade-offs, 3D convolutional architectures remain the most practical and effective approach for volumetric tumor segmentation in lung cancer CT scans. Among them, 3D U-Net^[6] is a well-established backbone that captures spatial context across all three dimensions. Building upon this foundation, the nnU-Net framework^[19] further automates and optimizes the segmentation pipeline by self-configuring its architecture, preprocessing, and training strategies based on the dataset properties.

nnU-Net was chosen for this project because it has consistently achieved strong performance across many medical image segmentation tasks and can automatically adapt to new datasets without manual tuning. In addition, nnU-Net has outperformed other methods in several international medical segmentation challenges, such as the Medical Segmentation Decathlon (MSD)^[1] and the Brain Tumor Segmentation (BraTS) challenge^[20]. These results demonstrate that nnU-Net is a reliable and well-validated framework for medical segmentation tasks, especially in complex datasets like lung CT scans.

5.2.1 Network's Architecture

The nnU-Net framework is not a fixed segmentation model but rather a self-configuring system that automatically adapts its architecture, training strategy, and inference pipeline based on the characteristics of a given dataset. The key innovation of nnU-Net lies in its ability to replace manual tuning with expert-knowledge-driven heuristics. To achieve this, it organizes configuration decisions into three main categories: fixed parameters, rule-based parameters, and empirical parameters. An overview of the overall nnU-Net architecture and its components is illustrated in Figure 5.2.1, highlighting the interaction between dataset fingerprinting, automatic configuration, and model training.

Fixed Parameters

Fixed parameters refer to design choices that remain constant across datasets. These decisions are grounded in best practices established in the medical image segmentation community. For instance, the model architecture is always derived from the U-Net template, featuring an encoder-decoder structure with skip connections, instance normalization, leaky ReLU activations, and deep supervision. The optimizer is Stochastic Gradient Descent (SGD) with Nesterov momentum set to 0.99, and the learning rate follows a polynomial decay schedule starting from 0.01. Data augmentation is applied extensively, including random rotations, Gaussian blur, brightness and contrast adjustments, and low-resolution simulations. The training process typically spans 1,000 epochs, each comprising 250 mini-batches with foreground oversampling to address class

5.2. Lung Tumor Segmentation

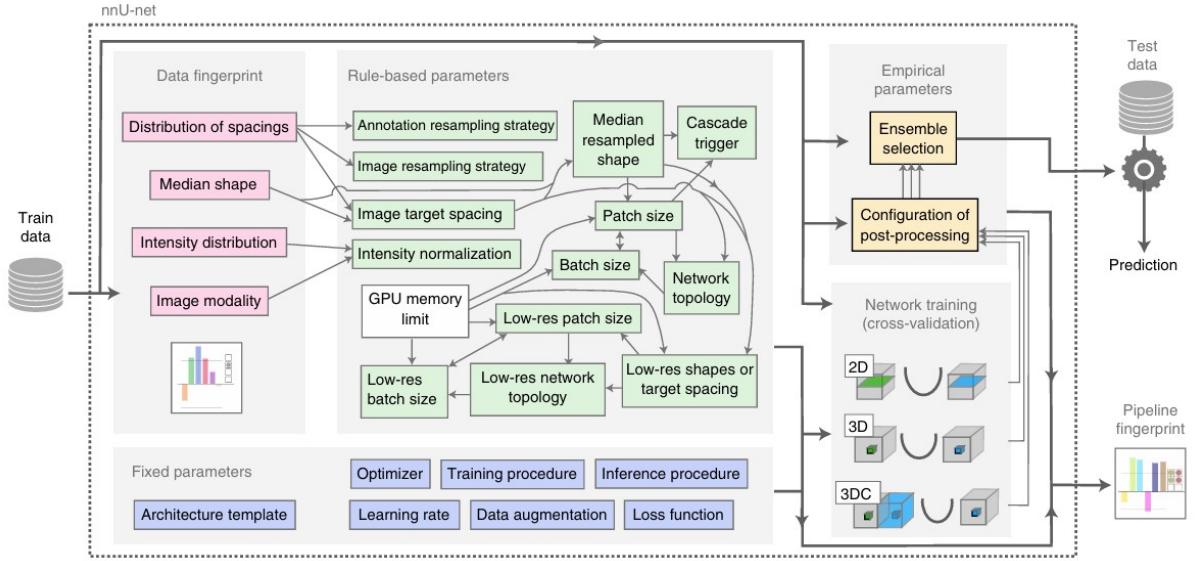


Figure 5.2.1: nnU-Net automated method configuration [19]

imbalance. During inference, nnU-Net uses a sliding window strategy with 50% patch overlap and Gaussian weighting to smooth predictions.

Rule-based Parameters

Rule-based parameters, in contrast, are dynamically derived from dataset-specific properties referred to as the "data fingerprint." This fingerprint includes information such as the median image shape, voxel spacing, and intensity distribution. Based on these statistics, nnU-Net configures several core components of the pipeline. The resampling strategy is selected according to the level of spacing heterogeneity across cases. For intensity normalization, CT images are globally z-scored based on the dataset, while MRI images are normalized on a per-case basis. The network architecture itself, including the number of downsampling operations and resolution levels, is automatically defined according to the patch size and median image shape. Additionally, when the patch size covers less than 12.5% of the median image volume, a two-stage cascade is triggered. An illustration of the two-stage cascade architecture is shown in Figure 5.2.2.

Empirical Parameters

Empirical parameters are those established based on experimental outcomes during model training. For example, nnU-Net evaluates whether applying post-processing operations, such as retaining only the largest connected component per class, improves cross-validation performance. If beneficial, these operations are incorporated into the

5.2. Lung Tumor Segmentation

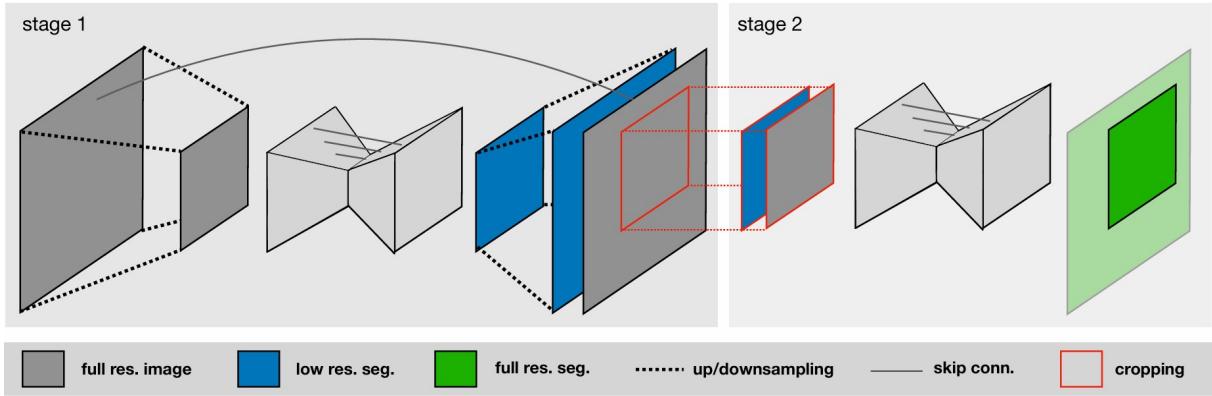


Figure 5.2.2: nnU-Net cascade architecture^[19]

final inference pipeline. Moreover, nnU-Net conducts automatic model selection and ensembling. Depending on the validation results, the best-performing configuration may be chosen from among 2D U-Net, 3D full-resolution U-Net, or the 3D cascade, or even ensemble their outputs to maximize segmentation accuracy.

Through this structured approach, nnU-Net combines domain expertise, data-specific adaptation, and performance-driven tuning to deliver a powerful and generalizable solution for biomedical image segmentation across diverse datasets.

5.2.2 Loss Function

A combination of dice loss and cross-entropy loss is employed to leverage the strengths of region-based and pixel-wise optimization objectives. This hybrid loss function helps improve the model's stability and segmentation performance, especially on imbalanced datasets where tumor regions are often much smaller than the background.

Dice Loss

Dice loss is derived from the Dice Similarity Coefficient, a commonly used metric in medical image segmentation to measure the overlap between the predicted segmentation and the ground truth. Let p_i be the predicted probability for voxel i , and g_i be the ground truth label (0 or 1) for voxel i .

$$\mathcal{L}_{Dice} = 1 - \frac{2 \times \sum_{i=1}^N p_i g_i}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N g_i^2}$$

Dice loss focuses on the overlap between the predicted segmentation and the ground truth. It directly optimizes for the segmentation quality by penalizing the model when there is a mismatch between the prediction and the ground truth, making it particularly effective for handling class imbalance.

Cross-Entropy Loss

Cross-entropy loss is a widely used loss function for classification tasks, including pixel-wise classification in segmentation. Let N be the number of voxels, C be the number of classes, $g_{i,c}$ be the ground truth label (one-hot encoded) for voxel i and class c , and $p_{i,c}$ be the predicted probability for voxel i belonging to class c .

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C g_{i,c} \log(p_{i,c})$$

Cross-entropy loss evaluates the pixel-level classification performance by comparing the predicted probability distribution to the true distribution. It encourages the model to output high probabilities for the correct class at each voxel, helping improve the local accuracy of the segmentation.

The final loss can be written as follows.

$$\mathcal{L}_{final} = \mathcal{L}_{Dice} + \mathcal{L}_{CE}$$

5.3 Radiomics-Based Feature Extraction

In this approach, radiomic features are extracted directly from segmented tumor masks generated by the nnU-Net model. These quantitative features describe both the shape and intensity distribution of the tumor region and are then combined with clinical data in a tabular form to build survival prediction models.

Figure 5.3.1 illustrates the complete workflow of this approach. After obtaining the binary segmentation mask from the nnU-Net model, we apply feature extraction techniques using the SimpleITK library. The extracted features are then merged with clinical attributes (e.g., age, gender, comorbidities), and the resulting tabular data are used as input to survival models such as Cox PH.

5.3.1 Feature Extraction using SimpleITK

SimpleITK is a simplified layer built on top of the Insight Segmentation and Registration Toolkit, which is one of the most widely used libraries for medical image processing and analysis. Designed for ease of use and rapid prototyping, SimpleITK provides a Pythonic interface for working with volumetric data such as CT, MRI, or PET scans. In the context of this study, SimpleITK is used to extract quantitative features from segmented tumor regions, which are essential for downstream survival prediction.

Specifically, we make use of two key components:

5.3. Radiomics-Based Feature Extraction

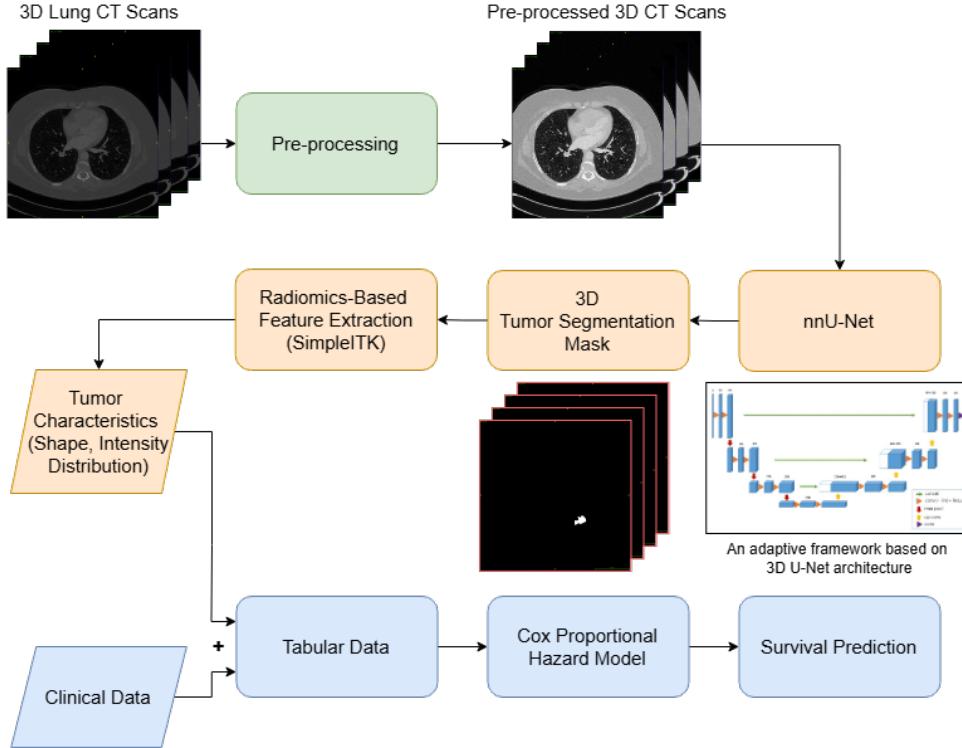


Figure 5.3.1: Radiomics-based feature extraction workflow

- **LabelShapeStatisticsImageFilter:** This module extracts geometric and morphological descriptors of the tumor, such as volume, surface area, and shape compactness.
- **LabelIntensityStatisticsImageFilter:** This module computes first-order intensity statistics (e.g., mean, standard deviation, minimum, and maximum) within the labeled tumor region, based on the original CT scan.

By leveraging these built-in filters, we can efficiently compute a set of interpretable, low-dimensional features that reflect both the size and internal variation of the tumor. These features serve as a compact representation of the tumor's characteristics and are valuable inputs to survival models such as Cox PH or AFT.

5.3.2 Selected Features and Their Significance

The goal of tumor feature extraction is to transform the segmentation masks into a compact set of quantitative descriptors that reflect the tumor's size, shape, and internal structure. In this study, a curated set of features was selected based on their clinical relevance, interpretability, and computational efficiency. These features fall into two main categories: morphological features (describing the tumor's shape and geometry) and

5.4. Deep Learning-Based Feature Fusion

statistical intensity features (describing the voxel intensity distribution within the tumor region). The selected features are summarized in Table 5.3.1.

Table 5.3.1: Selected radiomic features extracted from segmented tumor regions and their descriptions

Feature	Description
Volume	Total number of voxels in the tumor region
Mean	Average intensity value of voxels within the tumor.
Min	Minimum intensity value in the tumor region.
Max	Maximum intensity value in the tumor region.
Std	Standard deviation of voxel intensities, indicating intensity dispersion.
Median	The median intensity value of the tumor, robust to outliers.
Surface Area	The perimeter of the tumor's 3D shape.
Elongation	Degree to which the tumor shape is stretched or elongated.
Flatness	Measure of how flat the tumor is; useful for identifying disc-like shapes.
Roundness	Measure of how closely the shape resembles a perfect sphere.

These extracted features serve as concise yet informative representations of the tumor's appearance and geometry. By combining both intensity-based statistics and shape-related metrics, the resulting feature set captures essential aspects of tumor heterogeneity and progression. This representation not only facilitates model interpretability but also enables effective integration with clinical data for survival analysis tasks.

5.4 Deep Learning-Based Feature Fusion

In addition to the radiomics-based approach, this section introduces a deep learning-based framework designed to automatically learn feature representations from both imaging and clinical data. This method aims to overcome the limitations of handcrafted features by leveraging the representational power of convolutional and fully connected neural networks, and fusing them in a unified architecture to enhance the survival prediction performance.

Figure 5.4.1 illustrates the complete pipeline of this approach. The architecture consists of two parallel branches:

- **Image Branch:** processes 3D imaging data (CT scans, segmentation masks, and distance transforms) using a ResNet-based convolutional neural network to extract deep image features.

5.4. Deep Learning-Based Feature Fusion

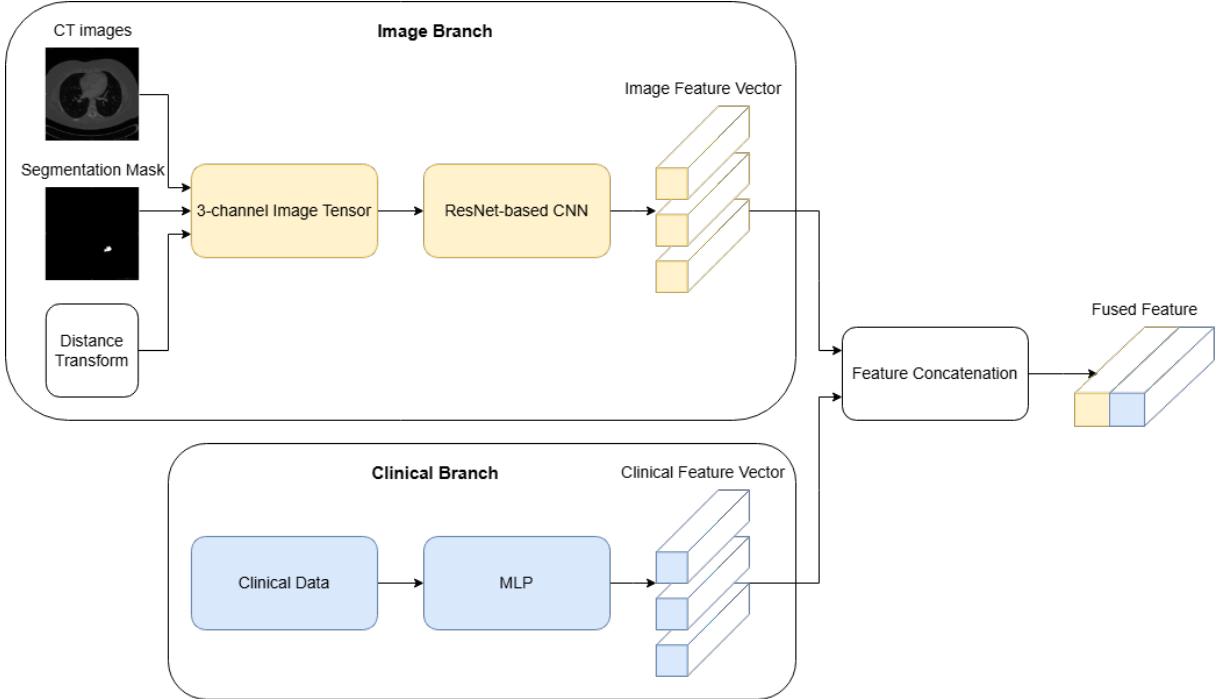


Figure 5.4.1: Deep learning-based feature fusion

- **Clinical Branch:** processes tabular clinical information using a MLP to extract clinical features.

The outputs of both branches are then fused through feature concatenation and passed into a survival analysis model to produce the final survival prediction. The key idea behind this design is to allow the model to jointly learn and integrate relevant patterns from both image and clinical modalities in an end-to-end fashion.

5.4.1 Image Branch: ResNet-Based Feature Extractor

To effectively leverage imaging information for survival prediction, we designed an image branch based on a CNN, specifically using the ResNet34 architecture^[16]. The image branch receives as input a 3-channel tensor for each patient, constructed from three modalities: (1) CT slice, (2) the corresponding segmentation mask, and (3) a distance transform of the mask. This multi-channel representation provides both raw anatomical information and spatial context regarding the tumor's location and proximity.

Slice Selection Strategy

CT scans of lung cancer patients typically contain hundreds of slices, while only a subset includes relevant tumor information. Feeding all slices into the network is

5.4. Deep Learning-Based Feature Fusion

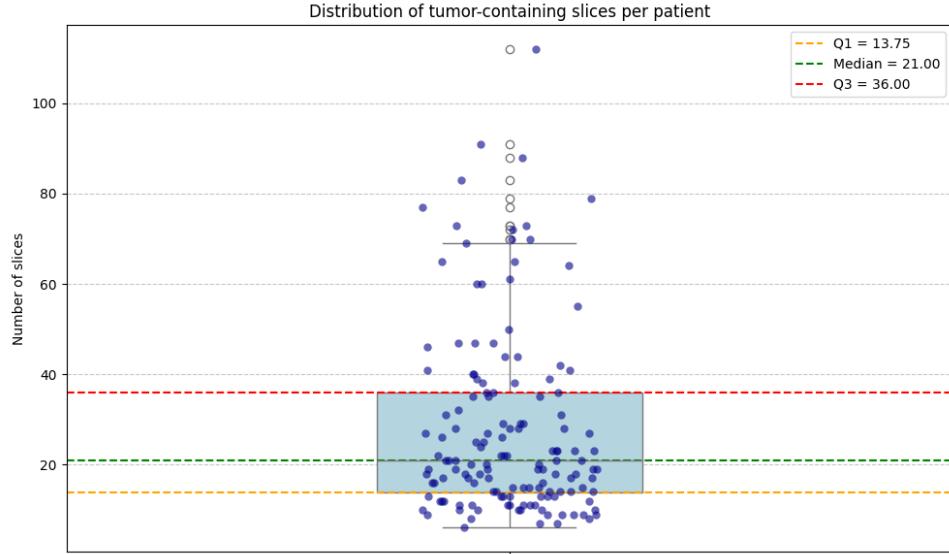


Figure 5.4.2: Distribution of patients by tumor slice count

computationally inefficient, introduces noise from non-informative regions, and may lead to overfitting. Therefore, it is critical to identify and extract the most relevant slices that contain the tumor.

To determine a consistent number of tumor-containing slices across patients, we first analyzed the distribution of slices in which the tumor was present using the segmentation masks. For each patient, the number of slices containing any non-zero tumor region was counted, and the results were visualized using a box plot (Figure 5.4.2). From this analysis, the first quartile (Q1) was found to be approximately 13.75 slices, the median (Q2) was 21 slices, and the third quartile (Q3) was 36 slices. This indicates that 25% of patients had 13 or fewer tumor-containing slices, half had 21 or fewer, and 75% had no more than 36 slices. The distribution was observed to be right-skewed, with a few patients having more than 80 slices containing tumor.

Based on this quartile analysis, we designed a slice selection strategy to ensure consistent input sizes across patients while preserving the most informative regions. Specifically, for patients with 21 or fewer tumor-containing slices, all slices were retained. For those with between 22 and 36 slices, a fixed number of 25 slices was selected. Finally, for patients with more than 36 slices, 36 slices were chosen. In every case, the selected slices were centered around the slice with the largest tumor area. If the selection window extended beyond the scan boundary, the strategy was adjusted symmetrically to maintain the target number of slices.

5.4. Deep Learning-Based Feature Fusion

Table 5.4.1: Slice selection strategy based on the number of tumor-containing slices

Tumor-containing slices	Number of slices selected
≤ 21 slices	Retain all slices
22-36 slices	Select 25 slices
> 36 slices	Select 36 slices

Image Tensor Construction

Once the relevant tumor-containing slices are selected, we construct a 3-channel tensor for each slice to serve as input for the image-based branch of the model. Each channel represents a complementary source of information:

- Channel 1: CT scan – the original grayscale CT slice, intensity-normalized. This provides anatomical and density-based context of tissues and structures.
- Channel 2: Tumor segmentation mask – a binary mask predicted by the nnU-Net model, indicating the precise location of the tumor. The mask is sparse and consists of only two pixel values (0 and 1), which alone may not be sufficiently informative.
- Channel 3: Distance transform of the mask – a continuous-valued map where each pixel encodes its distance to the nearest tumor boundary. This enhances spatial awareness around the tumor, capturing proximity information not present in the raw mask.

Using a multi-channel representation allows the model to simultaneously perceive complementary aspects of the tumor and its context. While the segmentation mask provides discrete boundaries, the distance transform highlights gradual spatial proximity, which can help the CNN better understand shape, spread, and surrounding tissue patterns. Meanwhile, the CT scan provides the raw intensity information that reflects actual tissue density and texture.

Overall, the chosen 3-channel setup captures a rich and diverse representation of the tumor’s appearance, location, and spatial context, while remaining computationally efficient and easy to integrate with standard CNN backbones such as ResNet34.

5.4.2 Clinical Branch: Multi-Layer Perceptron

To process the structured clinical data—including patient demographics, cancer-related information, and tumor attributes—we adopted a MLP architecture.

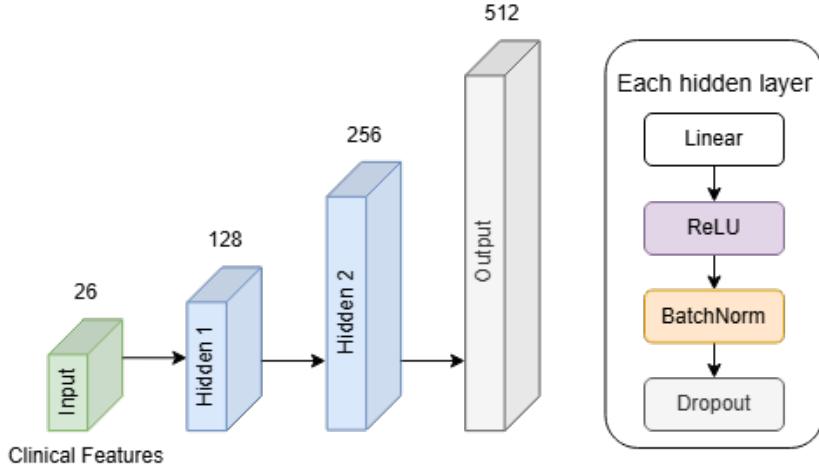


Figure 5.4.3: Architecture of the MLP used for extracting clinical feature vectors

MLPs are well-suited for handling tabular data due to their ability to model complex nonlinear relationships between input features and output representations.

In this study, the clinical input consisted of 26 features, including both categorical and continuous variables. The MLP was designed to map these features into a 512-dimensional latent feature vector, aligning with the feature dimension from the image branch for later fusion. The MLP architecture, as shown in Figure 5.4.3, comprises three fully connected layers with dimensions [128, 256, 512], each followed by ReLU activation, Batch Normalization, and Dropout ($p=0.3$) to enhance generalization and mitigate overfitting. The final output is a 512-dimensional clinical feature vector, which is used for feature fusion with the image branch.

The extracted clinical feature vector is subsequently concatenated with the image feature vector extracted from ResNet34, forming a multimodal representation used for survival prediction.

5.4.3 Feature Fusion via Concatenation

After obtaining feature vectors from the image and clinical branches, we concatenated them for joint representation learning. This fusion step enables the model to jointly reason over tumor morphology (captured from CT images) and patient-specific characteristics (derived from structured clinical data), both of which are crucial for accurate survival prediction in oncology.

Let:

- $f_{img} \in \mathbb{R}^{512}$: the feature vector extracted from the ResNet34 encoder after the global average pooling layer.

5.5. Survival Analysis Model

- $f_{cli} \in \mathbb{R}^{512}$: the feature vector produced by the MLP from clinical data.

The fusion vector is computed as:

$$f_{fusion} = \text{Concat}(f_{img}, f_{cli}) \in \mathbb{R}^{1024}$$

The resulting fused vector f_{fusion} serves as the unified input to the survival prediction model, which is described in the next section. By jointly encoding both anatomical and clinical information, the model is expected to better capture prognostic signals that may not be apparent from a single modality alone.

5.5 Survival Analysis Model

In this study, we explore various modeling strategies for survival prediction, ranging from classical statistical models to modern deep learning-based approaches. This section outlines the principles and motivations behind these methods, which will be compared later in the Experiments and Results section.

Kaplan-Meier

The Kaplan-Meier estimator is a non-parametric method used to estimate the survival function from time-to-event data. It is particularly useful in descriptive survival analysis, enabling visualization and comparison of survival curves across patient groups. However, Kaplan-Meier does not incorporate covariates (e.g., clinical or imaging features), and thus lacks the ability to make individualized survival predictions based on patient characteristics. This limitation makes it unsuitable as a predictive model for our task. Nonetheless, Kaplan-Meier remains useful for preliminary exploration and univariate analysis in survival datasets.

Parametric Models

Parametric survival models such as Weibull, log-normal, and log-logistic assume a specific probability distribution for the survival times. These models can yield interpretable survival curves and are computationally efficient. However, their performance heavily relies on the correctness of the assumed distribution, which may not hold across diverse patient populations. In particular, if the data distribution deviates significantly from the model assumption, the predictive accuracy can degrade substantially.

Moreover, while some parametric models can incorporate covariates (e.g., AFT models), the basic forms often model only the time-to-event and event status, making them less flexible in heterogeneous datasets like ours.

Accelerated Failure Time Models

AFT models describe survival time as a linear function of covariates in log-time, assuming a known survival distribution (e.g., Weibull, log-normal). They offer interpretable coefficients and are relatively simple to implement. However, their major limitation lies in the strong distributional assumptions. These assumptions can lead to poor model fit and reduced generalizability when applied to datasets with differing survival distributions. This lack of flexibility makes AFT models less robust in real-world clinical applications.

Cox Proportional Hazards Model (CoxPH)

The Cox Proportional Hazards model is a semi-parametric approach that models the hazard function as:

$$h(t) = h_0(t) \times \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_p X_p)$$

Here $h_0(t)$ is an unspecified baseline hazard, X is the vector of covariates, and β is the parameter vector. Cox PH avoids the need to specify the distribution of survival time, providing greater flexibility compared to AFT models. Additionally, it allows for interpretable hazard ratios and efficient handling of censored data. Due to its balance between flexibility and interpretability, CoxPH remains one of the most widely used models in clinical survival analysis.

Deep Learning-based Models: DeepSurv

DeepSurv is a deep neural network that extends the Cox PH framework by replacing the linear function with a neural network to capture complex nonlinear relationships between covariates and the hazard function. This model is particularly advantageous when dealing with high-dimensional or multi-modal data, such as image features combined with clinical variables.

Figure 5.5.1 shows the components of DeepSurv’s Architecture. First are fully connected layers, which take the input features X and transform them into higher-level representations. Each fully connected layer applies a linear transformation followed by a non-linear activation function (e.g., ReLU). Second is **Dropout Layers**^[30], these layers

5.5. Survival Analysis Model

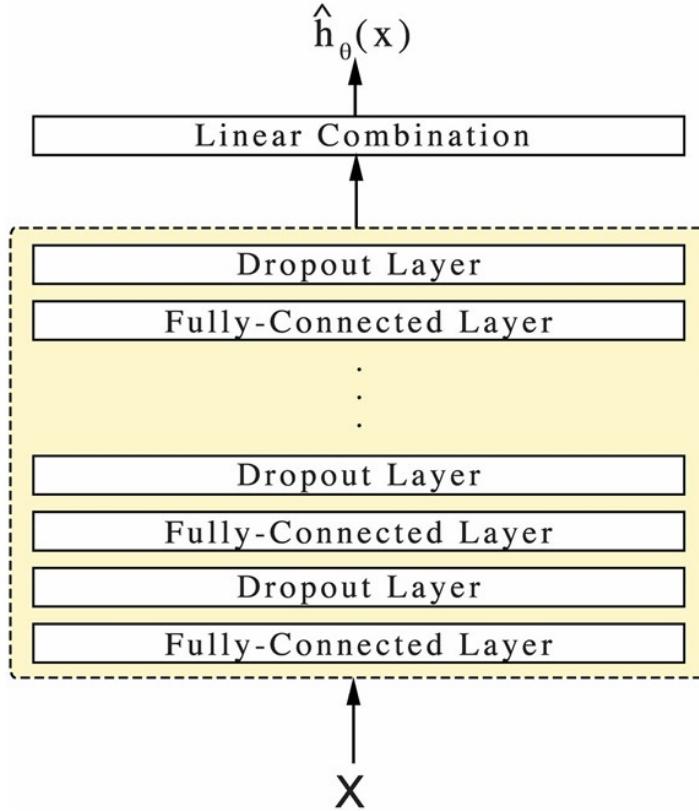


Figure 5.5.1: DeepSurv architecture^[24]

are used between fully connected layers to prevent overfitting by randomly deactivation a subset of neurons during training. This encourages the network to learn more robust patterns. Finally, a **Linear Combination Layer** is used to combine the features learned by the network into a single output, $\hat{h}_\theta(X)$, which represents the predicted risk score. However, the performance of DeepSurv is highly dependent on hyperparameter tuning, training stability, and data size.

In this work, we investigate and compare multiple survival analysis models, including AFT, Cox PH, and DeepSurv. While Cox PH is anticipated to offer a balance of performance and interpretability, especially given the nature of our clinical and imaging features, the final choice will be based on empirical results presented in the following section.

Chapter 6

Experiments and Results

This chapter focuses on evaluating the performance of deep learning models for lung tumor segmentation and survival analysis.

6.1 Dataset Analysis and Preprocessing

The Non-Small Cell Lung Cancer (NSCLC) Radiomics dataset was chosen as the primary data source. Its availability of CT scans, tumor segmentation masks, and clinical information makes it suitable for both segmentation and survival analysis tasks.

To ensure the data aligns with the model requirements, several preprocessing steps were performed, including image normalization, format conversion, and handling mismatches in slice count and dimensions. These steps help improve data consistency and optimize model performance.

6.1.1 Dataset

A radiogenomic dataset of NSCLC^[2] consists of medical imaging and molecular data for 211 subjects diagnosed with NSCLC. The data includes CT and PET/CT images, semantic annotations of tumor locations, segmentation maps of tumors in the CT scans, and results from various molecular profiling methods such as gene mutation analyses, gene expression microarrays, and Ribonucleic Acid (RNA) sequencing. Clinical data, including survival outcomes, are also included. The purpose of this dataset is to enable the discovery of relationships between tumor molecular and imaging features and to develop prognostic medical image biomarkers.

NSCLC Radiogenomics dataset is divided into 2 cohorts: (1) The R01 cohort, which consists of 162 subjects, including 69 from Stanford University School of Medicine and 93 from the Alto Veterans Affairs Healthcare System. Subjects were collected between

6.1. Dataset Analysis and Preprocessing

Table 6.1.1: Summary of NSCLC Radiogenomics's data types

Data Type	Number of subjects
Clinical Data	211
CT	211
CT Tumor Segmentations	144
CT Semantic Annotations	190
PET/CT	201
RNA-Seq	130
Gene expression Microarrays	26

August 7, 2008, and September 15, 2012. These individuals were selected from a pool of early-stage NSCLC patients who were scheduled for surgical treatment, with preoperative CT and PET/CT scans performed before their surgeries. Identifiers for this cohort follow the format R01-XXXXXX. (2) The AMC cohort, which includes an additional 49 subjects, was retrospectively gathered from Stanford University School of Medicine. Identifiers for this cohort follow the format AMC-XXXXXX.

Table 6.1.1 provides an overview of all data types in the NSCLC Radiogenomics dataset, including the number of subjects for each type. However, since the primary focus of this study is lung tumor segmentation, the concentration is placed solely on subjects with segmentation masks. Out of a total of 211 subjects, 144 subjects have tumor segmentation masks. Furthermore, these 144 subjects belong to the R01 cohort, allowing us to exclude AMC cohort from the dataset to optimize memory usage.

The CT images are stored in Digital Imaging and Communications in Medicine (DICOM) format. Since this is a retrospective dataset, different subjects were scanned using different scanners, scanning protocols, and scanning parameters:

- Slice thickness ranged from 0.625 to 3 mm (mean: 1.5 mm).
- X-ray tube current ranged from 124 to 699 mA (mean: 220 mA) at 80-140 kVp (mean: 120 kVp).

The scanning procedure was performed with the patient lying on their back, arms relaxed at their sides, from the apex of the lungs to the adrenal glands, all during a single breath-hold.

6.1. Dataset Analysis and Preprocessing

Table 6.1.2: Summary of demographic and clinical cohort characteristics

Feature	Number of subjects
Gender	
Female	36
Male	108
Smoking Status	
Current	30
Former	92
Nonsmoker	22
Histology	
Adenocarcinoma	112
NSCLC NOS (not otherwise specified)	3
Squamous cell carcinoma	29
Pathological T Stage	
T1a	38
T1b	30
T2a	40
T2b	9
T3	16
T4	5
Tis	6
Pathological N Stage	
N0	115
N1	12
N2	17
Pathological M Stage	
M0	140
M1a	1
M1b	3

According to the authors, initial segmentations for 144 subjects were generated using an unpublished automatic segmentation algorithm applied to axial CT image series. All of these segmentations were reviewed by a thoracic radiologist, who has over 5 years of experience, and edited as needed using ePAD. Final segmentations were then reviewed by an additional thoracic radiologist, and any disagreements regarding tumor boundaries were discussed and adjusted accordingly, with final approval by the thoracic radiologist. All segmentations are stored as DICOM Segmentation Objects^[22].

6.1. Dataset Analysis and Preprocessing

The NSCLC Radiogenomics dataset not only provides masks for segmentation but also includes valuable clinical data from patients. Table 6.1.2 summarizes the clinical data, such as age, gender, and pathological stage using R¹, which are crucial for analyzing the relationship between imaging features and patient outcomes.

6.1.2 Preprocessing

Preprocessing plays a crucial role in improving the efficiency and accuracy of deep learning models, particularly in the training process. For medical imaging, preprocessing is often a challenging task due to the complexity and variability of the data, including differences in image formats, sizes, resolutions, and intensities across datasets. Proper pre-processing can standardize the data and ensure compatibility with the model architecture, significantly speeding up training and improving performance.

In the proposed method mentioned above, most of the pre-processing is handled automatically by nnU-Net, which provides robust and efficient methods specifically tailored for medical image segmentation. However, it is still important to discuss some common preprocessing techniques and the ones employed by nnU-Net.

Format Conversion

The NSCLC Radiogenomics dataset mainly consists of CT scans provided in the DICOM format. While DICOM is a standard in medical imaging for storing metadata and images, it is not ideal for direct use in deep learning workflows due to its complexity, as images are often split into multiple files, and metadata handling can be inconsistent. Instead, Neuroimaging Informatics Technology Initiative (NIfTI) format is preferred for pre-processing in medical image analysis. NIfTI is a single-file format that preserves 3D spatial information, making it easier to handle and process compared to DICOM. Converting the raw DICOM images to NIfTI simplifies the workflow and ensures that the data is in a standardized format that is compatible with nnU-Net. Slicer3D² is a tool that can be used to convert DICOM files into NIfTI files.

Handling Variations in Slice Count and Dimensions

CT scans from different patients often vary in the number of slices (depth), in-plane resolution, and voxel spacing. For example, one patient's scan may have 300 slices, while another's has only 150 slices, leading to challenges in standardizing the data. nnU-Net

¹<https://www.r-project.org/>

²<https://www.slicer.org/>

6.1. Dataset Analysis and Preprocessing

addresses this by resampling all images to a consistent voxel spacing and normalizing their dimensions. This ensures that the input patch size for the model remains consistent, facilitating uniform training and reducing computational overhead.

Removing Redundant Slices

A typical CT scan of a patient might contain hundreds of slices, but only a fraction of these slices (e.g., 30-40 out of 300) might actually contain the tumor. Including all slices in the training process can lead to inefficiencies, as the model would spend unnecessary time processing slices that do not contribute meaningful information. Therefore, redundant slices that do not contain the region of interest (tumor) can be excluded to optimize the dataset. This not only reduces storage and computational requirements but also helps the model focus on learning features relevant to the task.

Intensity Normalization and Standardization

CT images typically represent voxel intensities in Hounsfield Units (HU), which can vary widely depending on the imaging protocol and patient characteristics. nnU-Net applies intensity normalization to standardize the HU values across all scans. This typically involves clipping the intensities to a predefined range (e.g., -1000 to 1000 HU) and scaling them to a normalized range (e.g., [0, 1]). This process ensures consistency across the dataset, enabling the model to learn effectively without being biased by intensity variations.

Other Common Preprocessing Techniques in Medical Imaging

Some widely used preprocessing techniques in medical imaging, which are also implemented by nnU-Net, include:

- Data Augmentation: Applying transformations such as rotation, scaling, flipping, and elastic deformations to artificially increase the diversity of the dataset.
- Cropping or Padding: Ensuring that all images have uniform dimensions by either cropping excess regions or padding smaller images.
- Segmentation Mask Alignment: Ensuring that the segmentation masks are properly aligned with the corresponding CT images after resizing or resampling.

By automating much of the preprocessing pipeline, nnU-Net simplifies the handling of complex medical imaging data, allowing researchers to focus on model development and

6.2. Experiment Setup

experimentation. The use of standardized formats like NIfTI, coupled with techniques like resampling, intensity normalization, and redundancy removal, ensures that the input data is optimized for deep learning, resulting in faster training and better segmentation outcomes.

6.2 Experiment Setup

6.2.1 Tumor Segmentation Experiments

This study employed the standard nnU-Net framework for tumor segmentation without architectural modifications. All experiments were conducted in a Jupyter Notebook environment running on a server equipped with Tesla T4 GPU acceleration.

The nnU-Net configuration was left unchanged, allowing the framework to automatically adapt its architecture and training pipeline to the properties of the input dataset. This self-configuring behavior is a core strength of nnU-Net, removing the need for manual architecture tuning and enabling reproducible results across different datasets.

The training process followed the recommended settings by the nnU-Net framework. Specifically, the model was optimized using SGD with Nesterov momentum set to 0.99. A polynomial learning rate schedule was applied with an initial value of 0.01. The loss function was a combination of Dice loss and cross-entropy loss, designed to balance pixel-wise classification accuracy with overlap quality. Training was performed for 1,000 epochs, each consisting of 250 minibatches, with a foreground oversampling strategy used to address class imbalance in the segmentation task.

To ensure robust evaluation, a 5-fold cross-validation strategy was adopted. The training data was divided into five equal parts, using four folds for training and the remaining fold for validation in each round. Performance metrics were averaged across all five folds to obtain a reliable estimation of the model's generalization capability.

6.2.2 Survival Analysis Experiments

For the Non-CT models, only clinical data from the NSCLC Radiogenomics dataset was used. This dataset includes 26 clinical features along with two additional columns representing the time-to-event and event status. In contrast, the CT-based models utilized an extended clinical dataset that incorporated 10 additional features extracted from pre-processed CT images, resulting in a total of 36 features plus the same two survival outcome columns.

Prior to training, the clinical data underwent preprocessing to ensure consistency. Missing values in the "Weight" column were imputed using the column mean. Categorical

6.2. Experiment Setup

variables were transformed using label encoding to convert them into a numerical format suitable for machine learning models.

Both Non-CT and CT models were trained and evaluated using an 80/20 train-test split. For the DeepSurv model, identical training configurations were applied in both Non-CT and CT settings, with 500 training epochs, a batch size of 32, and a learning rate of 0.01.

6.2.3 Evaluation Metrics

To evaluate the performance of the models in this project, several metrics are employed to provide a comprehensive analysis of their effectiveness. These metrics are tailored to the specific tasks of tumor segmentation and SA while also including standard metrics for segmentation.

Dice Coefficient

The Dice Coefficient^[10], also known as the Dice Similarity Index (DSI), is a widely used metric in medical image segmentation. It measures the overlap between the predicted segmentation mask and the ground truth. The formula for the Dice Coefficient is as follows:

$$\text{Dice Coefficient} = \frac{2 \cdot |A \cap B|}{|A| + |B|}$$

Where:

- A is the set of pixels (or voxels) in the predicted segmentation.
- B is the set of pixels (or voxels) in the ground truth segmentation.
- $|A \cap B|$ represents the number of overlapping pixels (or voxels).

The Dice Coefficient ranges from 0 to 1:

- 1 indicates perfect overlap.
- 0 indicates no overlap.

This metric is particularly suitable for medical image segmentation tasks, as it emphasizes the agreement between predictions and ground truth, even in imbalanced datasets where tumor regions are much smaller than the background.

6.2. Experiment Setup

Intersection over Union (IoU)

The Intersection over Union (IoU) is a standard metric used to evaluate the accuracy of segmentation models. It measures how well the predicted segmentation overlaps with the ground truth annotation. The IoU is defined as:

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

where:

- Area of Overlap is the number of pixels (or voxels) common to both the predicted segmentation and the ground truth.
- Area of Union is the total number of pixels present in either the predicted segmentation or the ground truth.

The IoU value ranges from 0 to 1:

- An IoU of 1 indicates a perfect overlap between the prediction and the ground truth.
- An IoU of 0 indicates no overlap at all.

This metric is critical for segmentation tasks as it directly evaluates the ability of the model to correctly delineate the target regions, providing a quantitative measure of segmentation quality.

Concordance Index (C-Index)

The Concordance Index (C-Index)^[13] is a standard metric used to evaluate the predictive accuracy of survival models. It measures how well the predicted risk scores correlate with the actual survival times. The C-Index is defined as:

$$C = \frac{\text{Number of concordant pairs}}{\text{Total number of comparable pairs}}$$

A pair of patients (i, j) is considered concordant if the patient with a shorter survival time has a higher risk score (or a lower predicted survival probability).

The C-Index ranges from 0 to 1:

- A C-Index of 1 indicates perfect prediction.
- A C-Index of 0.5 indicates a performance no better than random chance.

This metric is critical for survival analysis as it directly evaluates the ability of the model to rank patients based on their survival risk.

Other Complimentary Metrics

In addition to task-specific metrics, standard classification metrics are used to provide additional insights into the model’s performance:

Precision: Measures the proportion of true positives (correctly predicted tumor regions) out of all positive predictions.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

Recall (Sensitivity): Measures the proportion of true positives out of all actual positives (tumor regions in the ground truth).

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

These metrics are particularly useful for evaluating the segmentation model’s performance in identifying tumor regions and assessing its reliability in terms of false positives and false negatives.

6.3 Results

To evaluate the effectiveness of the proposed framework, we conducted a series of experiments on the publicly available NSCLC-Radiomics dataset. The experiments were designed to assess two primary tasks: (1) tumor segmentation using different deep learning architectures, and (2) survival prediction using clinical data, image-derived features, and their combinations. This section presents both quantitative and qualitative results, along with comparisons across baseline models and proposed enhancements.

6.3.1 Tumor Segmentation Performance

The first set of experiments focused on evaluating the performance of various neural network architectures for the lung tumor segmentation task. Specifically, we compared the standard 3D U-Net, an Attention U-Net variant, and the nnU-Net framework. All models were trained and evaluated on the same dataset under similar preprocessing conditions to ensure fairness.

We used the Dice Similarity Coefficient as the primary evaluation metric, which measures the overlap between predicted tumor regions and ground-truth annotations. The results are summarized in Table 6.3.1.

The nnU-Net model achieved the highest Dice score of 0.742, outperforming both the standard 3D U-Net and the Attention U-Net. The results demonstrate the effectiveness

6.3. Results

Table 6.3.1: Quantitative comparisons of segmentation performance of proposed model with other deep learning techniques

Model	Dice	IoU	Precision	Recall
3D UNet	0.604	0.472	0.721	0.596
Attention UNet	0.646	0.362	0.617	0.525
nnUNet_ResidualEncoder_3dfullres	0.742	0.624	0.851	0.746

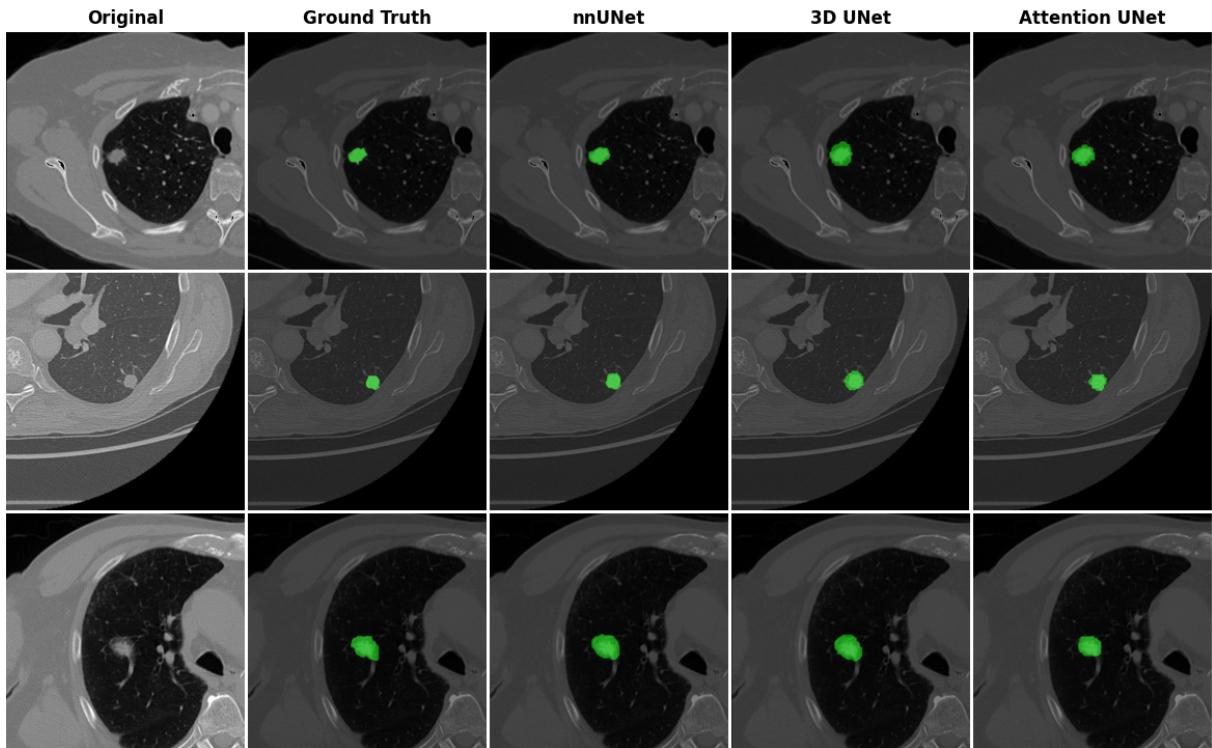


Figure 6.3.1: Comparison of tumor segmentation results across models

of nnU-Net’s automated architecture and training pipeline in segmenting complex tumor regions in 3D CT scans. While the Attention U-Net provided notable improvements over the baseline 3D U-Net, it still fell short of nnU-Net’s performance.

To provide visual insights into the segmentation quality, we include qualitative examples of predicted tumor masks from each model, as shown in Figure 6.3.1. Each example includes the original CT slice, ground truth annotation, and predicted masks from nnU-Net, 3D U-Net, and Attention U-Net, respectively.

As shown in Figure 6.3.1, the nnU-Net prediction most closely matches the ground truth, especially in terms of boundary precision and tumor shape preservation. In contrast, 3D U-Net and Attention U-Net models tend to under-segment.

6.3. Results

Table 6.3.2: C-index comparison of survival models using CT, clinical, and combined features

Model	CT Only	Clinical Only	Clinical + CT (Our Method)
CoxPH	0.621	0.756	0.769
AFT Weibull	0.535	0.714	0.746
AFT Log-Normal	0.519	0.751	0.767
AFT Log-Logistic	0.558	0.762	0.772
DeepSurv	0.504	0.613	0.628

6.3.2 Survival Prediction: CT vs Clinical vs Combined Models

In the second set of experiments, we assess the contribution of CT-derived imaging features to survival prediction by comparing three configurations: (1) models trained using only imaging features extracted from CT scans (CT Only), (2) models trained using only clinical variables (Clinical Only), and (3) models trained with both clinical and imaging features (Clinical + CT). Table 6.3.2 presents the comparative results across various survival analysis methods, including the classical Cox PH, AFT models with different distributional assumptions (Weibull, Log-Normal, Log-Logistic), and a deep learning-based model (DeepSurv).

The findings show that CT-only features yield lower C-Index values across all models, indicating that imaging data alone is not sufficiently predictive of patient survival. For instance, the best performance among CT-only models was achieved by the Cox PH model with a C-Index of 0.621, while DeepSurv reached only 0.504.

In contrast, models trained using only clinical variables demonstrated significantly better performance. Notably, the AFT model with a Log-Logistic distribution achieved a C-Index of 0.762, suggesting that clinical data provides strong prognostic information.

Most importantly, models that combine both clinical and imaging features consistently outperformed the other two setups. The AFT Log-Logistic model achieved the highest performance overall with a C-Index of 0.772, followed by the AFT Log-Normal model at 0.767 and Cox PH at 0.769. These results confirm our central hypothesis that radiological information extracted from CT scans, when combined with clinical data, offers complementary and meaningful prognostic value in predicting patient survival.

On the Performance of DeepSurv

Surprisingly, DeepSurv underperformed compared to traditional models, achieving a C-Index of only 0.504 when trained on CT features, 0.613 on only clinical data and 0.628

6.3. Results

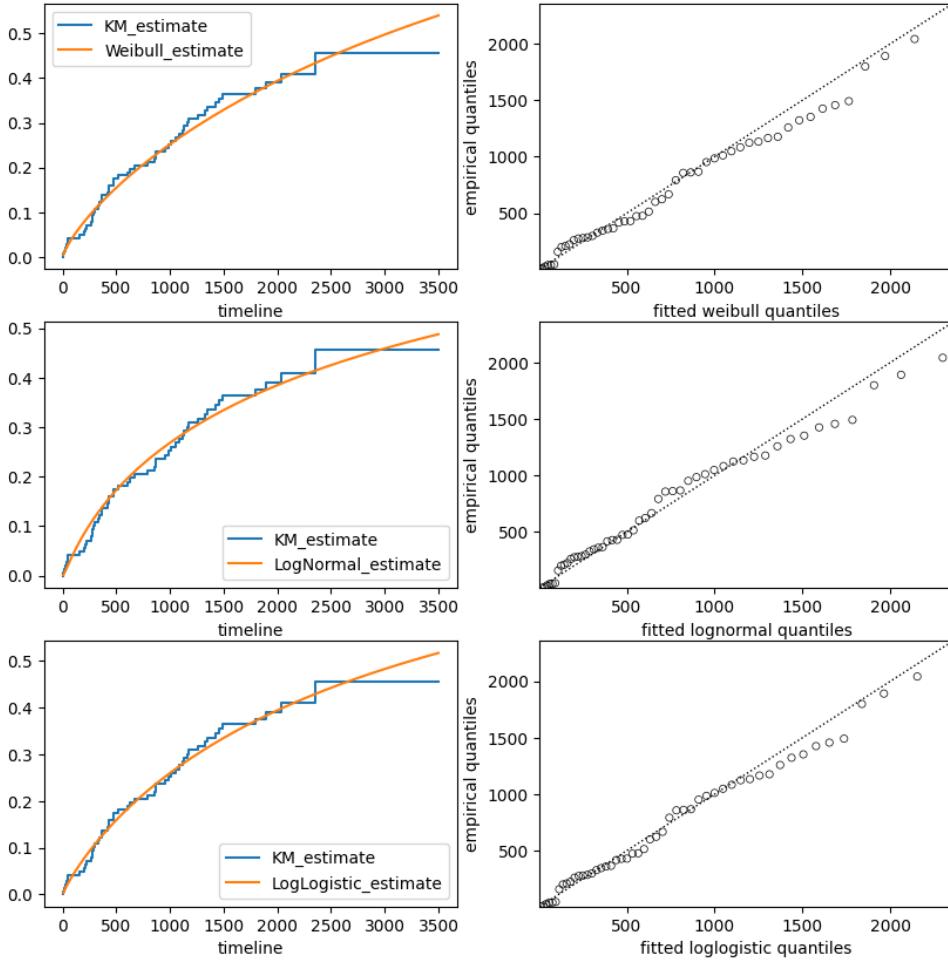


Figure 6.3.2: Comparison of distributional fits to the survival data

when additional CT-derived features were included. This result is notably lower than that of the Cox PH and AFT models, despite DeepSurv's theoretical advantage in capturing non-linear relationships. One likely reason is the limited dataset size ($N = 144$, with a censoring rate of 64%), which poses challenges for training deep learning models effectively. Furthermore, the model's relatively high complexity compared to the input dimensionality may have resulted in overfitting, further reducing generalization capability.

On Parametric Survival Models and Distribution Fitting

To better understand the data's temporal structure, we examined how well various parametric distributions fit the observed survival times. Figure 6.3.2 shows Kaplan-Meier estimates (blue) alongside fitted survival curves (orange) for Weibull, Log-Normal, and Log-Logistic models, as well as corresponding Q-Q plots.

These visualizations indicate that the dataset fits all three distributions reasonably

6.3. Results

well. The Q-Q plots reveal that the observed quantiles align closely with the theoretical quantiles, especially for Log-Logistic and Log-Normal distributions, suggesting that an AFT model based on these distributions could perform effectively on this dataset.

Indeed, the AFT models achieved solid results, particularly AFT Log-Logistic with a test C-Index of 0.558 (CT only), 0.762 (clinical only), and 0.772 (combined method). However, despite good empirical fits, we chose not to rely on AFT models as our primary modeling tool due to their inherent limitations.

Limitations of AFT Models

Although AFT models can offer good performance when the assumed distribution matches the data, they are inherently less flexible when generalizing to external datasets with unknown or variable survival time distributions. As Collett (2015) pointed out in “Modelling Survival Data in Medical Research”^[8], “the main disadvantage of the AFT model lies in the necessity of assuming a particular distribution for survival times, which, if incorrectly specified, can lead to biased regression estimates and misleading conclusions.”

This limitation makes AFT models less attractive for datasets where the underlying time-to-event distribution is uncertain or heterogeneous across subgroups. In contrast, the Cox PH model avoids this issue by not requiring a distributional assumption for the baseline hazard, offering more robustness and interpretability across diverse clinical settings.

Despite being a classical method, the Cox PH model demonstrated robust and competitive performance, especially when combined with radiomics features. Its flexibility, interpretability, and empirical strength continue to make it a strong baseline in survival modeling. This reinforces our choice to use Cox PH as the main evaluation baseline for CT-enhanced models.

6.3.3 Primary Method vs Proposed Enhancement

To evaluate the effectiveness of incorporating deep learning for CT-based survival prediction, we compared our primary method, which uses handcrafted radiomic features extracted from tumor segmentation masks, with a proposed enhancement that relies on deep feature fusion via neural networks. Table 6.3.3 summarizes the performance of the two approaches in terms of the C-Index.

From Table 6.3.3, it is evident that the primary method using radiomic features achieved slightly better results than the proposed deep learning approach. This outcome suggests that, under the current experimental conditions, handcrafted features remain

6.3. Results

Table 6.3.3: C-index comparison between primary method and proposed enhancement

Model	C-index
Radiomics + CoxPH (Primary Method)	0.769
CNN-Clinical Fusion + CoxPH (Proposed Enhancement)	0.730

more effective for survival prediction — likely due to their robustness when dealing with small datasets.

There are several reasons why the proposed enhancement has not yet surpassed the baseline. Most notably, the limited dataset size — 144 patients with high censorship (64%) — poses a significant challenge for deep neural networks, which typically require large-scale data to generalize effectively. The use of a ResNet18 backbone pretrained on natural images (ImageNet) likely exacerbates this issue, as it does not capture the unique spatial and intensity patterns of medical imaging.

In summary, although the proposed enhancement did not outperform the radiomics-based baseline, its ability to achieve a respectable C-Index underscores its potential. With access to larger datasets, survival-specific pretraining, and better-aligned training objectives, the deep fusion framework may evolve into a powerful alternative to conventional handcrafted pipelines, offering more scalable and automated survival prediction.

Chapter 7

Conclusion and Discussion

The content of this chapter presents the achievements made during the course of the project, as well as the challenges encountered.

7.1 Achieved Results

The primary goal of this study was to enhance survival prediction for lung cancer patients by incorporating medical imaging data alongside traditional clinical features. To this end, a two-stage pipeline was proposed. The first stage involved tumor segmentation using the nnU-Net framework to localize regions of interest in the CT images. In the second stage, features extracted from the segmented tumors were combined with clinical variables and used as input for survival analysis models.

Experimental results show that integrating CT-derived features with clinical data improved survival prediction performance across all models tested. Notably, the AFT Log-Logistic model achieved the best performance, with a C-index of 0.772 when using both CT and clinical data, compared to 0.762 when using clinical data alone. Similar improvements were also observed for other traditional models such as Cox PH and AFT Log-Normal. Although the DeepSurv model underperformed relative to classical methods, it still benefited marginally from the inclusion of CT features.

These findings confirm the effectiveness of incorporating image-based information into survival analysis, particularly in cases where radiomic features are available. While the improvements were not dramatic, they consistently demonstrated the value of CT-derived tumor characteristics as a complementary source of prognostic information.

7.2. Limitations

7.2 Limitations

While this study achieved promising results, it is important to recognize its limitations. The most significant constraint was the limited size of the dataset—only 144 patients, with 64% of the cases censored. Such a small and imbalanced sample restricts the statistical power of the models and limits their ability to generalize to broader populations.

Moreover, since the proposed enhancement was an exploratory idea developed under time constraints, only basic hyperparameter tuning and architectural adjustments were feasible. As a result, the deep learning models—especially DeepSurv—may not have reached their optimal performance.

Another practical limitation was the restricted computational resources available for training. Training deep learning models, particularly those involving medical images, is time-consuming and requires significant GPU capacity, which was not always accessible during this study.

Finally, due to the focus on a specific dataset (NSCLC Radiogenomics), the current findings may not fully generalize to other lung cancer cohorts or imaging modalities without further validation.

7.3 Future Works

Several directions could be pursued to improve and extend this study. First and foremost, increasing the dataset size would be highly beneficial. Access to larger and more diverse patient cohorts—especially those with lower censoring rates—would provide a stronger foundation for model training and evaluation. Additionally, validating the pipeline on external datasets would help assess its generalizability.

In terms of modeling, further hyperparameter optimization and architectural tuning—especially for DeepSurv and the dual-branch CNN-MLP fusion model—could unlock better performance. This includes exploring deeper networks, more advanced fusion strategies, and regularization techniques to mitigate overfitting on small datasets.

Lastly, future work may consider extending the current framework to support survival prediction for other cancer types, enabling broader clinical applicability.

References

- [1] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. The medical segmentation decathlon. *Nature communications*, 13(1):4128, 2022.
- [2] Shaimaa Bakr, Olivier Gevaert, Sebastian Echegaray, Kelsey Ayers, Mu Zhou, Majid Shafiq, Hong Zheng, Jalen Anthony Benson, Weiruo Zhang, Ann NC Leung, et al. A radiogenomic dataset of non-small cell lung cancer. *Scientific data*, 5(1):1–9, 2018.
- [3] Freddie Bray, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Isabelle Soerjomataram, and Ahmedin Jemal. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 74(3):229–263, 2024.
- [4] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.
- [5] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.
- [6] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.

REFERENCES

- [7] Taane G Clark, Michael J Bradburn, Sharon B Love, and Douglas G Altman. Survival analysis part i: basic concepts and first analyses. *British journal of cancer*, 89(2):232–238, 2003.
- [8] David Collett. *Modelling survival data in medical research*. Chapman and Hall/CRC, 2023.
- [9] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.
- [10] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [11] Razvan-Gabriel Dumitru, Darius Peteleaza, and Catalin Craciun. Using duck-net for polyp image segmentation. *Scientific reports*, 13(1):9803, 2023.
- [12] Eleonora Giunchiglia, Anton Nemchenko, and Mihaela van der Schaar. Rnn-surv: A deep recurrent model for survival analysis. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4–7, 2018, Proceedings, Part III 27*, pages 23–32. Springer, 2018.
- [13] Frank E Harrell, Robert M Califf, David B Pryor, Kerry L Lee, and Robert A Rosati. Evaluating the yield of medical tests. *Jama*, 247(18):2543–2546, 1982.
- [14] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 574–584, 2022.
- [15] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer, 2016.

REFERENCES

- [18] David W Hosmer Jr, Stanley Lemeshow, and Susanne May. *Applied survival analysis: regression modeling of time-to-event data*, volume 618. John Wiley & Sons, 2008.
- [19] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.
- [20] Fabian Isensee, Paul F Jäger, Peter M Full, Philipp Vollmuth, and Klaus H Maier-Hein. nnu-net for brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6*, pages 118–132. Springer, 2021.
- [21] Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 11–19, 2017.
- [22] Charles E Kahn, John A Carrino, Michael J Flynn, Donald J Peck, and Steven C Horii. Dicom and radiology: past, present, and future. *Journal of the American College of Radiology*, 4(9):652–657, 2007.
- [23] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.
- [24] Jared L Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, 18:1–12, 2018.
- [25] Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.
- [26] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [27] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Matthias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz,

REFERENCES

- et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
- [28] Richard Peto, MCetal Pike, Philip Armitage, Norman E Breslow, DR Cox, SV Howard, N Mantel, K McPherson, J Peto, and PG Smith. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. ii. analysis and examples. *British journal of cancer*, 35(1):1–39, 1977.
- [29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.
- [30] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [31] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, and Vishal M Patel. Medical transformer: Gated axial-attention for medical image segmentation. In *Medical image computing and computer assisted intervention–MICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, proceedings, part I 24*, pages 36–46. Springer, 2021.
- [32] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- [33] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.
- [34] Xinliang Zhu, Jiawen Yao, and Junzhou Huang. Deep convolutional neural network for survival analysis with pathological images. In *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pages 544–547. IEEE, 2016.