IMPERIAL COLLEGE LONDON

DEPARTMENT OF PHYSICS

# Digistain Prognostic Score Development

## Infrared Spectroscopy to Guide Chemotherapy Decisions

**Zamzam Al-Khalili**

Submitted for the Degree of

*Doctor of Philosophy*

Summer 2024

*to*

*my late father*

*my beloved mother*

*my little bundles of joy, Aisha and Ahmed*

*my husband, who walked every step of this journey with me*

# Acknowledgments

I wish to express my sincere appreciation to Prof. Chris Phillips for granting me the opportunity to work under his supervision, and for his unwavering support throughout my rather unusual journey.

I am deeply indebted to Dr. Hemmel Amrania for his thought-provoking contributions and invaluable guidance, which played the biggest role in shaping and enriching this thesis.

I would also like to acknowledge Dr. William Hart for generously sharing his expertise on statistical analysis, Ms. Christina Angelou for providing the essential spectral data for this research, and Mr. George Greaves for always stepping in to help, whether it is locating missing keys or ensuring compliance with safety regulations.

Last but certainly not least, I want to extend my gratitude to my family, especially my beloved sisters and wonderful nieces. Thank you for being my pillars of strength and sources of joy!

# Originality Declaration

The work presented in this thesis is my own, and all else is appropriately referenced or credited.

# Conflict of Interest Declaration

My supervisor, Prof. Chris Phillips, is the Chief Scientific Officer of Digistain Limited, while Dr. Hemmel Amrania, a former postdoctoral researcher under his supervision, is the Chief Executive Officer.

Discussions with both of them provided invaluable insights significantly enhancing my understanding of the subject matter. However, neither their roles nor our discussions have influenced the results, interpretations, or conclusions presented in this work.

# Copyright Declaration

# Abstract

Digistain is a mid-infrared technique that measures absorption at wavelengths corresponding to phosphate moieties in the nucleus and amide moieties in the cytoplasm. It produces an objective metric, named Digistain Index (DI), that inherently correlates with Nuclear-to-Cytoplasmic Ratio (NCR), a well-recognized measure of solid malignancies including breast cancer. The aim of this thesis is to develop a prognostic score that evaluates cancer risk and guides adjuvant chemotherapy treatment by incorporating DI with histological and pathological factors.

A cohort of 1184 breast cancer patients, who did not receive adjuvant chemotherapy, was selected for this study. Spectral measurements were conducted using a Fourier-Transform Infrared (FTIR) spectrometer on tissue samples collected from the identified patient cohort. Then, the DI was extracted from the spectral data using a computational code. Subsequently, the Digistain Prognostic Score (DPS), incorporating DI with tumour grade and lymph node status, was developed and validated using survival analysis tools, mainly Cox proportional hazards regression and Kaplan-Meier curves, in addition to Receiver Operating Characteristic (ROC) curves.

The resulting DPS showed statistically significant prognostic power and stratification capability for Overall Survival (OS) and Recurrence-Free Survival (RFS) at 5- and 10- years. The prognostic power was evidenced by the strong hazard ratios ($> 2$) and significant p-values ($\leq 0.0001$) obtained from Cox proportional hazards regression, in addition to the consistently high Area Under the Curve (AUC) values from ROC curves (0.7 - 0.8). The stratification capability was demonstrated by the clear distinction between high- and low- risk groups represented by Kaplen-Meier curves.

These results suggest that DPS holds substantial clinical utility as it facilitates individualized patient care by guiding treatment plans, particularly in regards to adjuvant chemotherapy.

# Contents


**I Introduction and Background** **25**

**1 Introduction and Motivation** **27**

1.1 Standard Practices in Breast Cancer Management . . . . . 27

1.2 Chemotherapy Overtreatment . . . . . 27

1.3 The Need for Alternative Prognostic Models . . . . . 28

1.4 Digistain: A Promising Alternative . . . . . 28

1.5 Thesis Framework . . . . . 29

**2 Breast Cancer: Conventional Practices Lead to Chemotherapy Overtreatment** **31**

2.1 Breast Cancer Heterogeneity . . . . . 31

2.2 Breast Cancer Epidemiology . . . . . 31

2.3 Breast Cancer Etiology . . . . . 32

2.4 Carcinogenesis: From a Normal Cell to a Cancer Cell . . . . . 32

2.5 Diagnostic Practices . . . . . 33

2.6 Nuclear-to-Cytoplasmic Ratio (NCR): A Morphological Proxy of Malignancy . . . . . 34

2.7 Histopathological Features . . . . . 35

2.7.1 Grade . . . . . 35

2.7.2 Histological Special Types . . . . . 36

2.7.3 Metastasis . . . . . 37

2.7.4 Lymph Node Status . . . . . 37

2.7.5 Tumor Size . . . . . 37

2.7.6 Stage . . . . . 37

2.8 Molecular Biomarkers . . . . . 37

2.8.1 Hormone Receptor Status . . . . . 37

2.8.2 HER2 Status . . . 38
2.8.3 Ki-67 . . . 38
2.9 Prognostic Factors and Prognostic Models . . . 38
2.10 Treatment Options . . . 39
2.11 Adjuvant Chemotherapy: Dilemma and Ramifications . . . 39
2.11.1 The Dilemma . . . 39
2.11.2 Prognostic Models to Guide Decision-Making . . . 40
2.11.3 Chemotherapy Overtreatment: The Catastrophic Result . . . 40
2.12 Conclusion . . . 41

**3 Gene Expression Profiling to Guide Adjuvant Chemotherapy Treatment 43**
3.1 Gene Expression . . . 43
3.2 Exploration of Cancer Biology Through Gene Expression Profiling . . . 44
3.3 Molecular Sub-Typing of Breast Cancer Based on Gene Expression Profiling . . . 45
3.4 Commercially Available Gene Expression Profiling Tests for Breast Cancer . . . 46
3.4.1 MammaPrint . . . 46
3.4.2 Oncotype DX . . . 46
3.4.3 EndoPredict . . . 46
3.5 Early Evidence of Gene Expression Profiling Efficacy . . . 47
3.6 Limitations and Challenges . . . 47
3.6.1 High Cost . . . 47
3.6.2 Turnaround Time . . . 47
3.7 Conclusion . . . 47

**4 Statistical Tools for Prognostic Model Development and Evaluation 49**
4.1 Survival Analysis . . . 49
4.1.1 Survival Endpoints . . . 50
4.1.2 Censored Data . . . 50
4.1.3 Survival Function . . . 51
4.1.4 Hazard Function . . . 51
4.1.5 Cumulative Hazard Function . . . 52
4.1.6 Kaplan-Meier Survival Curve Estimator . . . 52
4.1.7 Log-Rank Test . . . 55
4.1.8 Cox Proportional Hazards Regression . . . 58

4.2 Performance Metrics . . . 59
4.2.1 Sensitivity, Specificity and Accuracy . . . 60
4.2.2 Predictive Values . . . 61
4.2.3 Receiver Operator Characteristic Curves . . . 62
4.3 Conclusion . . . 65

**II Digistain as a Prognostic Model** **67**

**5 The Digistain Technique** **69**
5.1 Overview of Infrared Absorption Spectroscopy . . . 69
5.1.1 Light . . . 69
5.1.2 Matter . . . 70
5.1.3 Forms of Interaction . . . 70
5.1.4 Infrared Absorption Spectroscopy to Probe Biological Samples . . . 71
5.1.5 Transmittance and Absorbance . . . 72
5.1.6 Beer-Lambert Law . . . 72
5.1.7 Fourier Transform Infrared Spectrometers . . . 73
5.2 Spectroscopic Techniques in Pathology . . . 74
5.2.1 Fourier Transform Infrared Techniques . . . 75
5.2.2 Other Spectroscopic Techniques . . . 75
5.3 Principle of Operation of Digistain . . . 76
5.4 The Digistain Index . . . 77
5.5 Workflow of Digistain . . . 77
5.5.1 Tissue Sample Sourcing . . . 77
5.5.2 Region of Interest (ROI) Selection . . . 77
5.5.3 De-Paraffinization . . . 78
5.5.4 Absorption Measurements . . . 78
5.5.5 DI Calculation . . . 78
5.6 The Historical Evolution of Digistain . . . 78
5.6.1 Digistain as a Free Standing Spectrometer . . . 78
5.6.2 Digistain as an FTIR-Based Setup . . . 82
5.7 Strengths of Digistain . . . 83
5.8 Weaknesses of Digistain . . . 83

5.9 Conclusion . . . 83

**6 The Digistain Index: Calculation and Validation 85**

6.1 Digistain Index Calculation . . . 85

6.1.1 Spectral Peaks Characterisation . . . 86

6.1.2 Baseline Correction . . . 87

6.1.3 Peak-to-base Normalization . . . 89

6.1.4 Area Integration . . . 90

6.1.5 Digistain Index Formulation . . . 91

6.2 Digistain Index Validation . . . 91

6.2.1 Testing Area Integration Segment . . . 91

6.2.2 Testing Digistain Index Calculation Segment . . . 93

6.2.3 Testing Overall Code Artifacts and Errors . . . 93

6.3 Conclusion . . . 94

**7 The Digistain Prognostic Score: Development and Validation 95**

7.1 Study Cohort . . . 95

7.1.1 Sourcing, Preparation and Measurement of Samples . . . 95

7.1.2 Clinicopathological Characteristics of Patients . . . 96

7.1.3 Calculation and Distribution of Digistain Indices . . . 96

7.2 Digistain Prognostic Score Development . . . 96

7.2.1 Impact of Grade, Lymph Node Status and DI on Prognosis . . . 97

7.2.2 Combined Impact of Grade, Lymph Node Status and DI on Prognosis . . . 99

7.2.3 Formulation of Digistain Prognostic Score Equation . . . 100

7.3 Digistain Prognostic Score Validation . . . 101

7.3.1 Calculation and Distribution of Digistain Prognostic Scores . . . 101

7.3.2 Cox Regression: Assessing Risk Prediction . . . 101

7.3.3 ROC: Assessing Overall Performance . . . 102

7.3.4 Kaplan Muier: Assessing Risk Stratification . . . 103

7.3.5 Negative Predictive Values . . . 105

7.4 Conclusion . . . 105

**III Concluding Remarks and Future Work** **107**

**8 Concluding Remarks** **109**

**9 Future Work** **111**

**References** **113**

**Appendix A: Reproduction Permissions** **119**

# List of Figures


2.1 Incidence and mortality of breast cancer in North America and Sub-Saharan Africa . . . . . 32
2.2 Division of mother cells to daughter cells . . . . . 33
2.3 The Hematoxylin + Eosin (H+E) staining process . . . . . 34
2.4 Breast cancer histological grades . . . . . 36

3.1 Gene expression: transcription and translation . . . . . 44
3.2 Heatmap for gene expression profiling . . . . . 45

4.1 Example of survival analysis data . . . . . 51
4.2 Example of Kaplan-Meier curves . . . . . 54
4.3 $\chi^2$ distributions with varying degrees of freedom: 1df, 5 df and 10 df . . . . . 57
4.4 Example of Receiver Operating Characteristic (ROC) curve . . . . . 63
4.5 Area Under Curve (AUC) of an ROC curve to asses overall usefulness of a clinical test . . . . 63
4.6 Youden Index to find optimal cut-point of a clinical test . . . . . 65

5.1 The electromagnetic spectrum . . . . . 70
5.2 The ground and excited states of an atom . . . . . 70
5.3 Main light-matter forms of interaction . . . . . 71
5.4 The infrared region of the electromagnetic spectrum. . . . . 71
5.5 Absorbance and transmittance . . . . . 72
5.6 Example of Beer-Lambert law . . . . . 73
5.7 Fourier Transform Infrared (FTIR) spectrometer . . . . . 74
5.8 Absorbance spectrum of a breast cancer tissue sample . . . . . 76
5.9 The U-shaped Digistain prototype . . . . . 79
5.10 Correspondence between filters used in Digistain and absorbance spectrum of a tissue sample 79

5.11 Images of a breast biopsy section. . . . . 80
5.12 Correlation between Digistain Indices (DI) and cancer grades . . . . . 81
5.13 The 1D Digistain prototype . . . . . 81

6.1 Spectral peaks characterised by height and size . . . . . 86
6.2 Absorbance spectra affected by different baseline behaviors . . . . . 88
6.3 Absorbance spectrum highlighting areas subtracted to correct for baseline fluctuations . . . . 89
6.4 Absorbance spectrum highlighting areas used to normalise absorbance peaks . . . . . 90
6.5 Base areas used in testing the area integration segment of the code . . . . . 92
6.6 Artificial absorbance spectra used in testing the overall robustness of the code . . . . . 94

7.1 Histogram of Digistain Index (DI) values . . . . . 96
7.2 Forest plots of multivariate analysis, incorporating grade, stage, and DI as covariates. . . . . 100
7.3 ROC curves to test prognostic ability of DPS . . . . . 102
7.4 Kaplan-Meier curves to test stratification power of DPS . . . . . 104

# List of Tables


2.1 Prognostic categories of the Nottingham Prognostic Index (NPI) . . . . . 40

3.1 Main molecular subtypes of breast cancer . . . . . 46

4.1 Contingency table for use in a logrank test . . . . . 56
4.2 Possible outcomes of a clinical test based on actual values and predicted values . . . . . 60
4.3 Summary of sensitivity, specificity, PPV and NPV . . . . . 62
4.4 Discriminatory power of a clinical test based on AUC . . . . . 64

6.1 Expected areas and computed areas show perfect agreement . . . . . 92
6.2 Expected DI and computed DI show perfect agreement . . . . . 93

7.1 Cox model for evaluating the ability of grade to predict survival . . . . . 97
7.2 Cox model for evaluating the ability of lymph node status to predict survival . . . . . 98
7.3 Cox model for evaluating the ability of DI to predict survival . . . . . 98
7.4 Cox model for evaluating the ability of grade, lymph node status and DI to predict survival . 99
7.5 Cox model for evaluating the ability of DPS to predict clinical survival . . . . . 101
7.6 Negative Predictive Values (NPV) based on DPS high and low risk classification . . . . . 105

# Part I

# Introduction and Background

# Chapter 1

# Introduction and Motivation

Breast cancer is the most prevalent type of cancer among women. Not only does it lead in terms of cancer diagnoses but it also accounts for the highest number of cancer-related deaths. In 2020 alone, breast cancer affected 2.3 million women and claimed the lives of over 685,000.[1]

This thesis aims to develop a new prognostic model to inform treatment decisions regarding adjuvant chemotherapy. This chapter explores the motivation behind it and outlines its structure.

## 1.1 Standard Practices in Breast Cancer Management

Breast cancer diagnosis begins with a biopsy of the suspected tumor, followed by tissue examination. The diagnostic process focuses on histopathological factors and molecular biomarkers. Key histopathological factors include tumor grade, lymph node status, tumor size, and tumor stage. Molecular biomarkers include Hormone Receptor (HR) status and Human Epidermal Growth Factor Receptor 2 (HER2) status. These factors play crucial roles in prognosis and are instrumental in guiding treatment decisions.

The Nottingham Prognostic Index (NPI), which incorporates tumor grade, tumor size, and lymph node status, is widely regarded as the gold standard for breast cancer prognosis.

## 1.2 Chemotherapy Overtreatment

The decision to administer adjuvant chemotherapy is one of great significance and complexity. While chemotherapy can significantly enhance the survival rate of a patient, it also carries the risk of severe side

effects such as cardiac toxicity and neurological complications.[2][3]

The decision often relies on risk assessment using the NPI, which incorporates tumor grade. However, tumor grade depends on qualitative features, making it subjective and resulting in approximately 30% inter-observer variability.[4] This subjectivity can lead to inconsistent risk assessments and ultimately result in chemotherapy overtreatment, which impacts patients quality of life and imposes substantial burdens on healthcare systems.[3]

Studies suggest that a staggering 85% of women who undergo chemotherapy, based on the calculation of their individual risk profiles using conventional prognostic factors like NPI, do not derive any benefit from the treatment.[5] Furthermore, a report by the National Confidential Enquiry into Patient Outcome and Death (NCEPOD), an independent body that reviews clinical practices across the UK, investigated the deaths of 600 patients who received adjuvant chemotherapy. The investigation revealed that 43% of these patients battled adverse treatment effects, and 27% died because of chemotherapy.[6]

## 1.3 The Need for Alternative Prognostic Models

Chemotherapy overtreatment underscores the urgent need to explore prognostic models that are more objective and reliable than the NPI. In recent years, Gene Expression Profiling (GEP) has garnered attention for providing a more personalised approach to evaluating cancer risk compared to conventional prognostic models. However, GEP tests come with high costs and long turnaround times, limiting their widespread application and effectiveness.[7]

## 1.4 Digistain: A Promising Alternative

Digistain, a spectroscopic technique pioneered at Imperial College, holds immense promise in addressing this challenge. It operates in the mid-infrared region of the spectrum, offering a non-invasive, cost-effective, and time-efficient solution to analyze tissue samples. It produces an objective metric, named Digistain Index (DI), that inherently correlates with Nuclear-to-Cytoplasmic Ratio (NCR), a well-recognized measure of solid malignancies including breast cancer.

This thesis seeks to unlock the full potential of Digistain in revolutionizing cancer management by using it to develop an new prognostic model, termed the Digistain Prognostic Score (DPS), that is designed to inform treatment decisions and mitigate chemotherapy overtreatment.

To achieve this ambitious objective, we embark on a retrospective study that explores the survival of over

a thousand breast cancer patients who did not undergo chemotherapy treatment. The first step is to optimise DI, and the second step is to integrate DI with other prognostic factors to produce the new prognostic model.

## 1.5 Thesis Framework

The first part of this thesis provides the necessary background information. Chapter 2 provides an overview of breast cancer, briefly discussing its heterogeneity, epidemiology and etiology and then covering its diagnosis, prognosis, and management, which leads to the pressing issue of chemotherapy overtreatment. Chapter 3 discusses GEP as a potential solution, along with the challenges hindering its widespread application. Chapter 4 reviews essential statistical tools required to effectively utilize Digistain in producing a new prognostic model.

Armed with an understanding of both the problem and the needed solution, as well as the necessary statistical tools, the second part of this thesis delves into Digistain. Chapter 5 introduces Digistain, highlighting its scientific underpinnings, inherent strengths, and previous research findings. Chapter 6 focuses on the first step of creating a prognostic model from the Digistain technique: converting spectra measurements to DI. Chapter 7 focuses on the second step: converting DI to DPS.

The final part of this thesis includes concluding remarks and future work. Chapter 8 presents the concluding remarks on the thesis. Chapter 9 outlines a plan for future work.

# Chapter 2

# Breast Cancer: Conventional Practices Lead to Chemotherapy Overtreatment

This chapter provides an overview of breast cancer, briefly discussing its heterogeneity, epidemiology and etiology before delving into its initiation, diagnosis, prognosis, and treatment options. It concludes with a discussion on the pressing issue of chemotherapy overtreatment.

## 2.1 Breast Cancer Heterogeneity

Breast cancer is an umbrella term that encompasses all cancers originating in the breast tissue despite their heterogeneity. These cancers exhibit diversity in histological and molecular characteristics, disease progression patterns, and responses to treatment.[8] Some studies go as far as suggesting that breast cancer is a collection of different diseases that simply affect the same organ.[9] This heterogeneity poses significant challenges in estimating personalised prognosis and choosing appropriate treatment plans.

## 2.2 Breast Cancer Epidemiology

Breast cancer *incidence*, the number or rate of new cases, and *mortality*, the number or rate of deaths, vary by geographical location. The statistics taken between 2014 and 2019 reveal survival rates over 85% in high-

income countries, which reduce to less than 60% in poorer regions of the world, as summarised in figure 2.1. This drastic difference is linked to disparity in access to diagnostic and treatment facilities, where people in high-income countries benefit from increased awareness, earlier detection, and more effective interventions.[10]

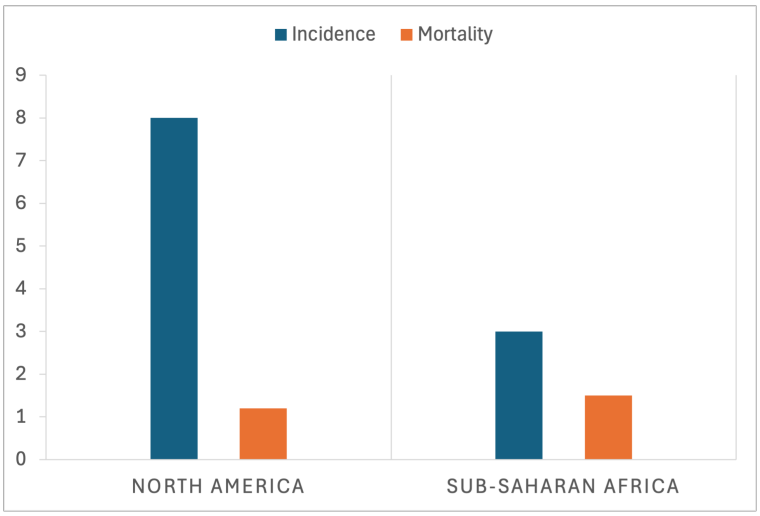



*Figure. 2.1:* Estimated cumulative risk of incidence (being diagnosed) and mortality (dying) of breast cancer in North America and Sub-Saharan Africa, highlighting the higher mortality-to-incidence ratio in Africa, likely due to disparities in access to diagnostic and treatment facilities. (Containing data from[10])

## 2.3 Breast Cancer Etiology

Several factors contribute to the increased risk of developing breast cancer. Women who have been diagnosed with cancer in one breast are more likely to develop cancer in the other breast. Moreover, women who have a family history of breast cancer, whether on the paternal or maternal side, are also at a higher risk of developing breast cancer. This is especially true if the family member developed cancer at an early age, had bilateral breast cancer, or if multiple family members were affected. These genetic tendencies often involve mutations in the BReast CAncer 1 (BRCA1) and BReast CAncer 2 (BRCA2) genes, which can be passed down from one generation to the next. Additionally, hormonal fluctuations, such as those experienced during menopause or through hormone therapy, can also heighten risk. Lifestyle choices, including lack of movement, poor dietary habits, and alcohol consumption, along with environmental factors like exposure to radiation, may further increase the risk of developing breast cancer.[11]

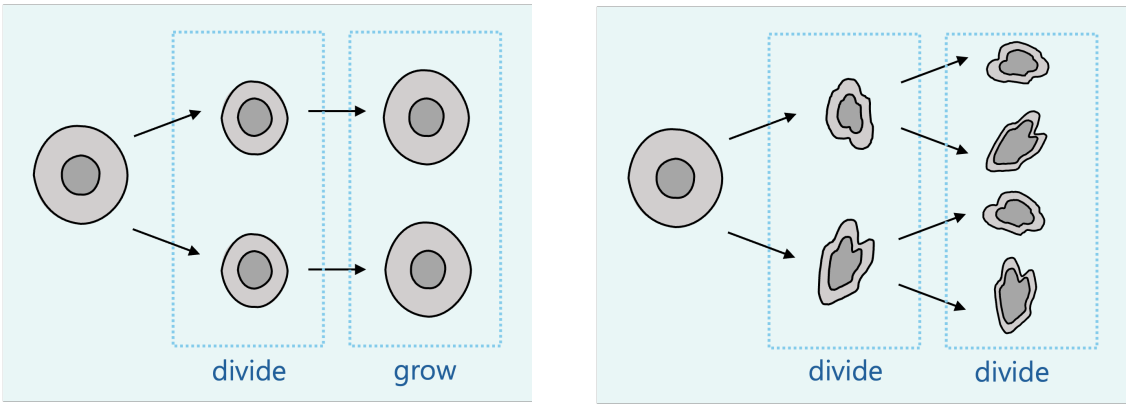
## 2.4 Carcinogenesis: From a Normal Cell to a Cancer Cell

The human cell consists mainly of a *nucleus*, which governs the cell and houses most of its genetic material (DNA), surrounded by a gel-like substance called *cytoplasm*. In the division of a healthy cell into daughter

cells, its nucleus is replicated while its cytoplasm is shared. Subsequently, the daughter cells undergo growth and division.[12]

Mutations in DNA can occur spontaneously or due to external factors such as radiation, viruses, bacteria, heat, and chemicals. These mutations have the potential to convert healthy cells into cancerous ones, a process known as *carcinogenesis*, particularly when mutation levels are elevated.[11]

Cancer induces changes in mitotic activity (i.e. rate of cell division) and pleomorphic features (i.e. size and shape of cells), resulting in smaller, irregularly-shaped cells with heightened nuclear content, as illustrated in figure 2.2.[12]




(a) (b)

*Figure. 2.2:* Division of mother cells to daughter cells. (a) Healthy: During cell division, nuclear content is cloned while cytoplasm is shared. Daughter cells grow and divide, resulting in cells of uniform size and shape. (b) Cancerous: Premature cell division leads to irregularly shaped cells with a higher than normal nuclear to cytoplasmic ratio.

## 2.5 Diagnostic Practices

Cancer diagnosis usually starts with a biopsy, which involves taking a tissue sample from the suspected tumor. This can be done using Fine Needle Aspiration (FNA), needle biopsy or surgical biopsy. The sample is immersed in formaldehyde solution in order to preserve it. Then, it is immersed several times in alcohol in order to dehydrate it. Tissue samples preserved and prepared using these steps are referred to as Formalin -Fixed Paraffin-Embedded (FFPE) tissue samples. Thin slices are cut from the FFPE tissue samples and placed on glass slides. Then, they are de-paraffinated and chemically stained using two dyes: Hematoxylin (H), which dyes the nuclear part of the cells blue, and Eosin (E), which dyes the cytoplasmic part pink. This process is known as H+E Staining and it is illustrated in figure 2.3. The stained slides are examined under a standard

optical microscope, allowing histopathological features to be extracted. Furthermore, immunohistochemistry tests are conducted on the biopsy samples, enabling the extraction of molecular metrics.[13]

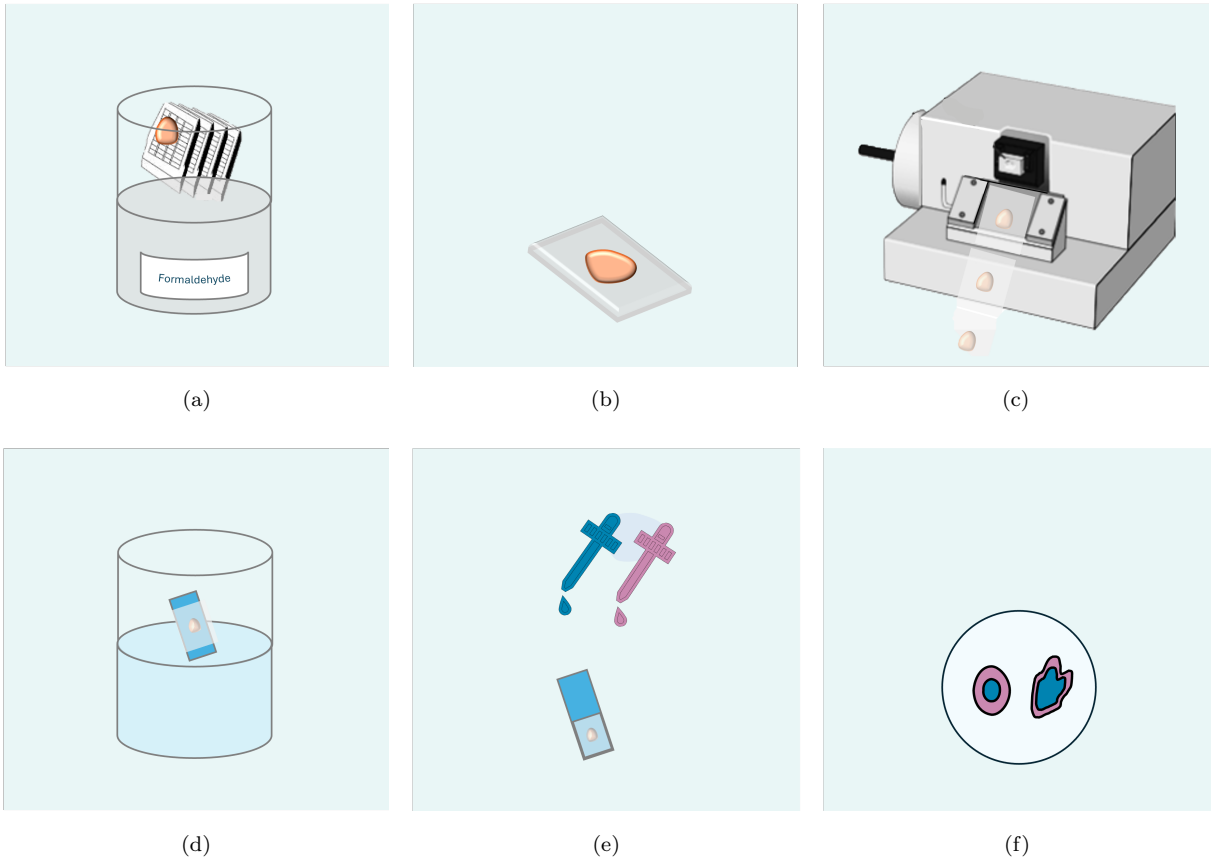



*Figure. 2.3:* The Hematoxylin + Eosin (H+E) staining process: (a) Fix the sample in formaldehyde. (b) Embed the fixed sample in paraffin. (c) Section the embedded sample. (d) Mount the sectioned sample on a slide. (e) Stain the mounted sample with hematoxylin and eosin. (f) Observe the stained sample under a microscope.

## 2.6 Nuclear-to-Cytoplasmic Ratio (NCR): A Morphological Proxy of Malignancy

The Nuclear-to-Cytoplasmic Ratio (NCR) is defined as the ratio of the size of the nucleus to the size of the cytoplasm within a cell. It reflects cellular morphology, with higher NCR values often correlating with malignancy and more aggressive cancer behavior. NCR serves as a significant proxy for elevated tumor DNA content across various solid cancers, including breast cancer.[14]

Traditionally, NCR is determined through microscopic measurements after H+E staining, where the dimensions of the nucleus and cytoplasm are manually assessed. However, these measurements are performed

on thin tissue sections, offering only two-dimensional representations that may not accurately reflect the three-dimensional structure of cells. Furthermore, this approach is prone to significant inter-observer variability, leading to inconsistencies in results.[15]

Flow cytometry has also been employed to measure NCR by analyzing the relative fluorescence intensity of nuclear and cytoplasmic components, provided the cells are appropriately stained. While effective, this method involves complex sample preparation and analysis, limiting its practicality in routine pathology workflows.[15][16]

More recently, the field has shifted toward digital image analysis, which utilizes specialized software to accurately measure nuclear and cytoplasmic areas. This technique enhances reliability, reduces observer bias, and provides a more standardized approach to NCR assessment.[15]

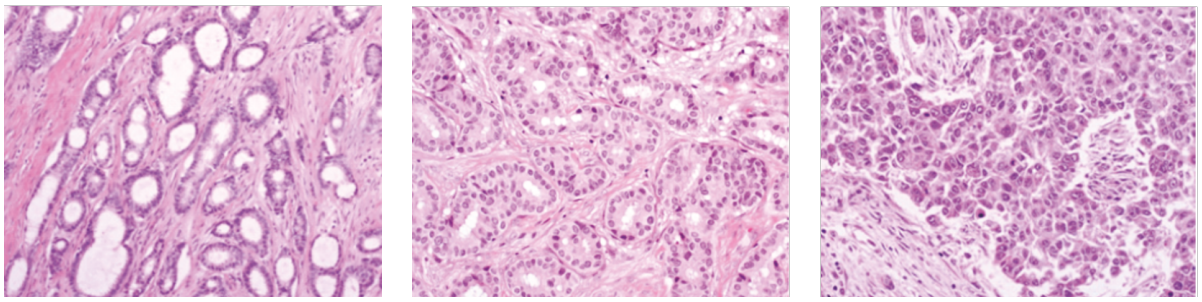
## 2.7 Histopathological Features

Histological features refer to how cancer cells appear under a microscope, while pathological features describe how cancer cells spread. Combined, these are known as histopathological features, which form the traditional basis of cancer diagnosis and prognosis.

### 2.7.1 Grade

The grade of a cancer reflects the degree of abnormality in the appearance and behavior of cancer cells compared to normal cells. In breast cancer, grading adheres to the Nottingham Grading System (NGS), which evaluates three morphological features: tubule formation (i.e. structure of cells), nuclear pleomorphism (i.e. size and shape of nucleus), and mitotic activity (i.e. rate of cell division). Each characteristic receives a score ranging from 1 to 3, yielding a total score between 3 and 9. Increasing scores correspond to greater levels of abnormality. Grade 1 encompasses scores from 3 to 5, while grade 2 comprises scores 6 and 7, and finally grade 3 includes scores 8 and 9.[17][18] Grades 1, 2 and 3 are respectively referred to as well-differentiated, moderately differentiated and poorly differentiated. Figure 2.4 shows an example for each grade.

Breast cancer is graded by histopathologists mainly based on qualitative morphological factors, leading to subjective grades. A study involving over 24,000 samples graded by 732 histopathologists from various institutions found that the probability of different pathologists grading the same tumor differently can reach as high as 30%.[14]



(a) (b) (c)

*Figure. 2.4:* Breast Cancer graded based on tubule formation, nuclear pleomorphism and mitotic activity as (a) grade 1. (b) grade 2. (c) grade 3. (Reproduced from[17]. No permission required. See Appendix A.)

### 2.7.2 Histological Special Types

Breast cancers are labeled based on the exact tissue of origin. Most breast cancers originate in the milk producing glands, known as *lobules*, or in the milk passages, known as *ducts*. These breast cancers are referred to as *lobular cancers* and *ductal cancers*, respectively. Finally, a small percentage of breast cancers originate in the fatty and fibrous connective tissues within the breast.[11] Ductal and lobular carcinomas exhibit distinct histological features when viewed under a microscope, allowing pathologists to differentiate them based on their microscopic appearance.

Breast cancers that remain contained within their tissues of origin are referred to as *non-invasive breast cancers* or *breast cancers in-situ*. In contrast, cancers that spread into surrounding tissues are labeled as *invasive breast cancers*. Finally, cancers that spread to other organs or to lymph nodes are categorized as *metastatic breast cancers*.[11]

Moreover, breast cancers can be further classified, based on their growth patterns, into histological special types.[9] According to the World Health Organization (WHO), there are at least 21 distinct histological special types. However, 75% of breast cancers still lack enough known features to merit classification into any histological special type.[19]

The most frequently occurring type of breast cancer is Invasive Ductal Carcinoma of No Special Type (IDC-NST), which originates in the milk ducts but spreads to the fatty tissues of the breast and potentially other parts of the body.[9] This is followed by Invasive Lobular Carcinoma (ILC). Non-invasive cancers constitute a small percentage, with Ductal Carcinoma In Situ (DCIS) representing the majority of them.[11]

### 2.7.3 Metastasis

Metastasis is the spread of cancer cells from the primary tumour site to other organs of the body, either through the blood system or lymph system.

### 2.7.4 Lymph Node Status

Lymph nodes are integral components of the immune system. They are distributed throughout the human body. When breast cancer metastasizes, one of the initial sites of involvement is often the lymph nodes located in the underarms. If cancer cells are detected in one or more of these lymph nodes, the condition is classified as Lymph Node positive (LN-positive).[20][21]

### 2.7.5 Tumor Size

The size of a tumor is defined by its longest dimension. Larger tumor sizes often correlate with a higher likelihood of lymph node involvement, as the increased size increases the probability of cancer cells spreading to nearby lymph nodes.[20]

### 2.7.6 Stage

The stage of a given cancer is a measure of its spread beyond the primary tumor site. Breast cancer usually follows a TNM staging system, where T refers to tumour size, N refers to number of positive lymph nodes, and M refers to metastasis status. The TNM classifies cancers from 0 to IV. Higher stages indicate a greater extent of spreading. Stage 0 refers to a non-invasive breast cancer that has not spread to any lymph nodes, stages I to III refer to invasive breast cancers, while stage IV refers to a metastatic breast cancer that has spread to other organs or to lymph nodes distant from the breast.[22]

## 2.8 Molecular Biomarkers

Molecular biomarkers are genes, proteins, or hormones found in the tumor that can influence its behavior.

### 2.8.1 Hormone Receptor Status

Hormone Receptors (HR) are specialized proteins that bind to hormones. The main types of hormone receptors found in breast cancer are Estrogen Receptors (ER) and Progesterone Receptors (PR). Breast cancers that have either one or both of these receptors are labeled as HR-positive.[23] These cancers require

hormones in order to grow and proliferate. HR-positive breast cancers constitute approximately 60% to 70% of newly diagnosed breast cancers.[20]

### 2.8.2 HER2 Status

The Human Epidermal growth factor Receptor 2 (HER2) is a protein that promotes cell growth. Breast cancers with normal HER2 levels are said to be HER2-negative while those with higher-than-normal HER2 levels are said to be HER2-positive. Excessive levels of HER2 correlate with higher proliferation, and accordingly more aggressive breast cancer.[23] HER2-positive breast cancers constitute approximately 15% to 20% of newly diagnosed breast cancers.[20]

### 2.8.3 Ki-67

Ki-67 is a protein that cells produce during their growth and division processes. It is one of the most commonly used measures of cell proliferation rate, and it is reported as a percentage. High Ki-67 levels indicate a high proliferation rate, suggesting more aggressive cancer with a higher risk of spreading to other tissues or organs.[23]

## 2.9 Prognostic Factors and Prognostic Models

Prognosis is the prediction of breast cancer outcome, which can be local recurrence, distant recurrence, disease-specific death, or overall death. Prognosis helps patients foster a sense of preparedness, helps researchers select suitable candidates for studies and clinical trials and helps monitoring organizations compare the performance of different healthcare institutions. Most importantly, prognosis helps clinicians tailor individualized treatment regimens.[24]

A *prognostic factor* is any variable linked to subsequent clinical outcomes. It can stratify patients into different categories and predict overall outcome. For example, the risk of death within 5 years for lymph node-negative breast cancers is 17%, which increases to 27% for cases with 1-3 positive nodes, to 54% for cases with 4-12 positive nodes, and further to 72% for cases with more than 13 positive nodes.[25] This example highlights the profound effect lymph node status has on overall survival, confirming its role as a prognostic factor.

Prognostic factors are often not enough to predict individualized outcomes, despite their adequacy for assessing overall outcomes. This has prompted the development of *prognostic models*, also referred to as *prognostic scores*, which combine various prognostic factors to provide a more comprehensive assessment.[26]

## 2.10 Treatment Options

Breast cancer patients often undergo surgical procedures to remove the cancerous tissue. The surgery can range from breast-conserving surgery, where only the cancerous lesion and a small amount of surrounding tissue is removed, to radical mastectomy, where the entire breast along with all surrounding lymph nodes and chest wall muscles are removed. The choice of surgery depends on several factors, including the size and number of tumors within the breast.[27]

Before surgery, some patients may receive *neoadjuvant therapy*, aiming to halt tumor growth and decrease its size, making the surgery less invasive and more effective. After surgery, patients typically receive *adjuvant therapy*, which aims to kill any remaining cancer cells and reduce the risk of recurrence.[27]

Regardless of whether the treatment is administered before or after surgery, the type of treatment is guided by clinical, histopatholigical and molecular factors. The main types of treatment are: radiation therapy, chemotherapy, endocrine therapy and HER2-targeted therapy.[27] Radiation therapy uses high energy x-rays while chemotherapy uses cytotoxic drugs in order to destroy cancer cells. Endocrine therapy blocks the hormones that fuel cancer growth, and hence it is effective for ER-positive and PR-positive cancers only. Finally, HER2-targeted therapy, most commonly trastuzumab, specifically targets the HER2 protein on cancer cells making it effective for HER2-positive cancers.[27]

Despite extensive efforts to standardize breast cancer care, treatment plans continue to show significant variability. A study carried out by the Multidisciplinary Application of Genomics in Clinical practice (MAGIC) used data collected from 672 HR-positive, HER-negative breast cancer patients to assess differences in treatment plans advised by 911 experienced clinicians from 52 countries. The study showed discordance in treatment plans in at least 15% of patients. These are mainly the patients older than 50, with a grade 1 or grade 2, ER-positive tumour that is less than 3 cm in size.[28]

## 2.11 Adjuvant Chemotherapy: Dilemma and Ramifications

### 2.11.1 The Dilemma

The administration of chemotherapy to a patient is a difficult decision, but abstaining from using is equally difficult. On one hand, chemotherapy may substantially increase the survival rate of a patient and decrease the recurrence rate of a cancer. On the other hand, chemotherapy may cause drastic side effects including cardiac toxicity and neurological complications.[2 3]

### 2.11.2 Prognostic Models to Guide Decision-Making

The decision to include adjuvant chemotherapy is based on the individual risk profile. Typically, a 20% risk of distant recurrence within 10 years is used as the threshold for recommending chemotherapy.[25] Therefore, accurate risk prediction is critical for guiding treatment decisions and ensuring the best possible outcomes for patients.

Risk estimation relies on prognostic factors or prognostic models, with one of the most commonly used models being the Nottingham Prognostic Index (NPI). It predicts prognosis following surgery by incorporating tumor size, lymoh node status, and histological grade.[29] It is formulated as

$$NPI = 0.2S + N + G \tag{2.1}$$

where,

$S$: size, in cm

$N$: lymoh node status, 1 for 0 nodes, 2 for 1-3 nodes, and 3 for $> 3$ nodes

$G$: grade

The NPI is the gold standard in breast cancer prognosis assessment. A higher NPI reflects a worse prognosis, indicating the need for more aggressive treatment such as chemotherapy.[29] The ranges of NPI scores, reflecting different prognosis categories, are summarised in table 2.1.

*Table 2.1:* Prognostic categories of the Nottingham Prognostic Index (NPI).[30]

| Area Under Curve (AUC) | Index | Prognosis |
|---|---|---|
| I | $\leq 2.4$ | Excellent |
| II | $> 2.4$ but $\leq 3.4$ | Good |
| III | $> 3.4$ but $\leq 5.4$ | Moderate |
| IV | $> 5.4$ | Poor |

### 2.11.3 Chemotherapy Overtreatment: The Catastrophic Result

The decision to administer adjuvant chemotherapy depends on risk prognostication, predominantly through the NPI, which incorporates grade, and we already know from section 2.7.1 that grade is subjective as it depends on morphological features including: tubule formation, nuclear pleomorphism and mitotic activity.

This ultimately results in chemotherapy overtreatment.

Studies suggest that a staggering 85% of women who undergo chemotherapy, based on the calculation of their individual risk profiles using conventional prognostic factors like NPI, do not derive any benefit from the treatment.[5]

Furthermore, a report by the National Confidential Enquiry into Patient Outcome and Death (NCEPOD), an independent body that reviews clinical practices across the UK, investigated the deaths of 600 patients who received adjuvant chemotherapy. The investigation revealed that 43% of these patients battled adverse treatment effects, and 27% died because of chemotherapy.[6]

## 2.12 Conclusion

Current breast cancer diagnostic practices focus on histopathological and molecular factors, which play crucial roles in prognosis and guiding treatment decisions. However, despite their significance, determining whether to administer adjuvant chemotherapy remains challenging, often resulting in overtreatment. This overtreatment impacts patients quality of life and imposes substantial burdens on healthcare systems. To address this issue, researchers and clinicians are exploring additional or alternative prognostic markers that are more objective and reliable. A common approach is gene expression profiling, which we discuss next.

# Chapter 3

# Gene Expression Profiling to Guide Adjuvant Chemotherapy Treatment

This chapter offers an overview of Gene Expression Profiling (GEP), and its potential to reduce chemotherapy overtreatment. It starts by explaining how GEP works, and its role in subtyping breast cancer. Then, it discusses commercially available tests like MammaPrint, Oncotype DX, and EndoPredict, and their role in providing personalised prognostic information. Finally, it addresses the limitations and challenges of these tests.

## 3.1 Gene Expression

DNA is made up of functional units known as genes. Each gene contains instructions for producing a functional product, which is often a protein. Gene expression occurs in two stages: *transcription* and *translation*. During transcription, the gene is copied into messenger RNA (mRNA). Subsequently, the mRNA is translated into a protein.

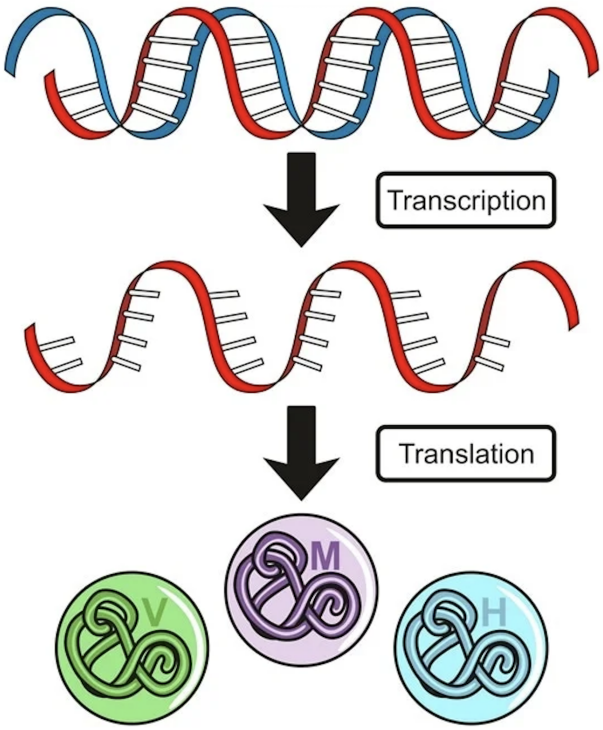



*Figure. 3.1:* Gene expression happens at two stages. Transcription: DNA is copied into mRNA. Translation: mRNA is translated into a protein. (Reproduced from[31]. No permission required. See Appendix A.)

## 3.2 Exploration of Cancer Biology Through Gene Expression Profiling

The biology of a tumor can be explored at three phases: DNA, mRNA, or protein. Cancer cells often have mutations in various genes, but not all mutated genes are active or expressed at significant levels. Hence, in the past two decades, there has been a shift in focus to determine the molecular profile of tumors by measuring gene expression levels of specific molecules at the mRNA phase.

GEP, also widely known as MultiGene Assay (MGA), is conducted on samples of tumor tissue obtained from biopsies or surgical procedures. It aims to measure gene expression levels for specific sets of genes that are present in cancer cells, thus providing insight into the unique genetic characteristics of each individual cancer. This information can be invaluable for tailoring personalised treatment approaches and predicting the responsiveness of the cancer to targeted therapies.

## 3.3 Molecular Sub-Typing of Breast Cancer Based on Gene Expression Profiling

ER, PR, HER2, and Ki-67 are widely used biomarkers. Individually, they all have well established prognostic power. The natural progression is to merge these four biomarkers in order to leverage all the information they possess. This is achieved by measuring the expression levels of these genes. The results are often visualized using heatmaps, where each gene is presented as a row and each patient as a column, as illustrated in figure 3.2.

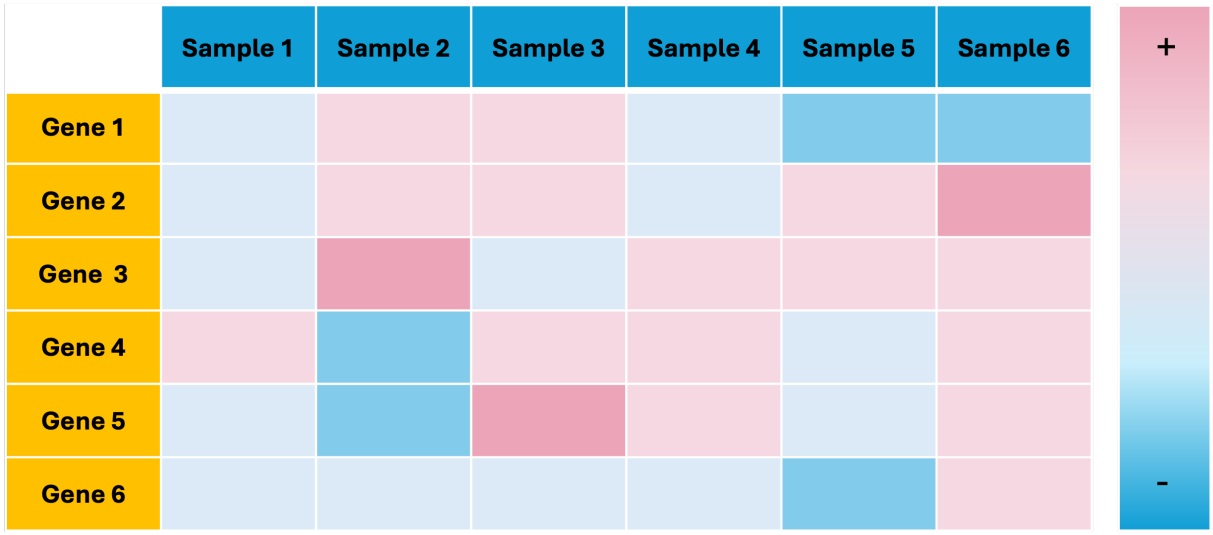



*Figure. 3.2:* Heatmap for gene expression profiling. Rows represent individual genes, while columns represent patient samples. The color intensity indicates the level of gene expression, with clusters highlighting groups of patients with similar gene expression profiles.

Studies have shown that breast cancers classified based on ER, PR, HER2, and Ki-67 gene expression fall into four main molecular subtypes: Luminal A, Luminal B, HER2-enriched, and Triple Negative. Each molecular subtype exhibits different patterns of gene expression, as summarized in table 3.1, and responds differently to various treatment modalities.

*Table 3.1:* Main molecular subtypes of breast cancer based on gene expression profiling of four molecular biomarkers.[8]

| Luminal A | Luminal B | | HER2+ | TN |
|---|---|---|---|---|
| | HER2 - | HER2 + | | |
| ER + | ER + | ER + | ER - | ER - |
| PR + | PR + | PR -/+ | PR - | PR - |
| HER2 - | HER2 - | HER2 + | HER2 + | HER2 - |
| Ki-67 Low | Ki-67 High | Ki-67 High/Low | Ki-67 High | Ki-67 High |

## 3.4 Commercially Available Gene Expression Profiling Tests for Breast Cancer

### 3.4.1 MammaPrint

MammaPrint is the first GEP test to be introduced for breast cancer. It measures the expression of a 70-gene panel to stratify patients into high risk and low risk of distant metastases in a 5-year interval. It can be used regardless of ER-status.[5]

### 3.4.2 Oncotype DX

Oncotype DX is the most commonly used GEP test worldwide. It examines a set of 21 genes, including 16 cancer-related genes and 5 reference genes. It evaluates the expression levels of each gene and uses these weighed values to calculate a Recurrence Score (RS), ranging from 0 to 100. Based on this RS, patients are categorized as low-risk, intermediate-risk, or high-risk for cancer recurrence. It is primarily utilized for ER-positive, HER2-negative, early-stage breast cancers. It is recommended for tumors smaller than 5 centimeters.[5 32]

### 3.4.3 EndoPredict

EndoPredict is one of the most recent GEP tests. It measures the expression of 8 cancer-related genes and 4 reference genes. It is primarily used for ER-positive, HER2-negative breast cancers.[5]

## 3.5 Early Evidence of Gene Expression Profiling Efficacy

A study on 75 women with early breast cancer revealed that the use of GEP aided in making treatment decisions for 28% of patients for whom clinicians were undecided. Additionally, these tests spared approximately 25% of patients from undergoing unnecessary adjuvant chemotherapy.[3]

Another study specifically targeting women with ER+ HER- breast cancer, who were routinely due to receive adjuvant chemotherapy treatment, evaluated Oncotype DX for 67 patients to determine overall survival at a 10-year interval and guide chemotherapy decisions. The study revealed that in at least 65% of the patients, chemotherapy may cause more harm than good as it would not improve their overall survival.[33]

## 3.6 Limitations and Challenges

### 3.6.1 High Cost

GEP tests commonly rely on expensive reagents, significantly increasing the cost per test. The high cost can limit the availability of these tests in regions with limited healthcare budgets or where individuals lack adequate insurance coverage. This disparity in access to healthcare resources between high-income and low-income countries may potentially widen the gap in survival rates between these regions.

### 3.6.2 Turnaround Time

GEP tests are based on complex laboratory workflows, which must be conducted in specialized central laboratories. Typically, it takes 10-14 days for these tests to be processed once the samples arrive at the lab. However, the logistics involved in sending the samples to these central labs can extend the turnaround time considerably, which is a major issue as timely access to diagnostic results and subsequent treatments is crucial for patient outcomes.[34]

## 3.7 Conclusion

Gene expression consists of two steps: transcription, from gene to mRNA, and translation, from mRNA to protein. GEP focuses on determining the molecular profile of a tumor by measuring the expression of genes at the mRNA phase. This profiling can help address the problem of chemotherapy overtreatment in two ways.

First, it categorizes breast cancer into distinct molecular subtypes based on the expression of established

biomarkers like ER, PR, HER2, and Ki-67. Each subtype exhibits unique biological characteristics and responses to treatment, aiding in treatment decisions.

Second, specialized tests based on GEP, such as MammaPrint, Oncotype DX, and EndoPredict, measure the expression of more genes, providing personalised prognostic information that further informs treatment decisions. By leveraging these tests, a significant number of patients, particularly those with ER-positive, HER2-negative breast cancer, can be spared unnecessary chemotherapy. However, these tests rely on expensive reagents and complex laboratory workflows, leading to relatively high costs and long turnaround times. Therefore, the remainder of this thesis aims to explore an alternative approach.

# Chapter 4

# Statistical Tools for Prognostic Model Development and Evaluation

The aim of this chapter is to present the statistical concepts that will be used throughout our analysis. First, we review the survival analysis tools that are necessary for developing a prognostic model. Next, we review performance metrics that can be used to assess the accuracy and reliability of the prognostic model we create. Detailed derivations are beyond the scope of this thesis.

## 4.1 Survival Analysis

Survival analysis is a branch of statistics that is heavily used in clinical research. The focus of survival analysis is to explore the *time-to-event*, a continuous variable measured from a predefined starting point such as the initiation of treatment to the occurrence of a specific outcome, and possibly to link it to variables that explain it or predict it.[35]

Survival analysis captures the dynamic nature of events by incorporating varying entry times and study durations. It utilizes all available data, including information from participants who do not experience the event, providing a comprehensive understanding of the survival data and a robust method for their assessment and interpretation.[35]

Survival analysis is often used to assess the impact of a treatment or intervention on the survival of participants, and to establish prognostic models, which help categorize patients based on their prognosis,

guiding clinical decisions about treatment strategies and overall patient management.[35]

In this section, the terms *survival endpoints* and *censored data* are reviewed. Then, the *survival function*, the *hazard function* and the *cumulative hazard function* are defined. Finally, prominent statistical techniques including *Kaplan-Meier Curves*, to describe survival distributions, *logrank test*, to compare survival distributions, *Cox proportional hazards regression*, to explain the effect of prognostic factors on survival outcomes, are explored.

### 4.1.1 Survival Endpoints

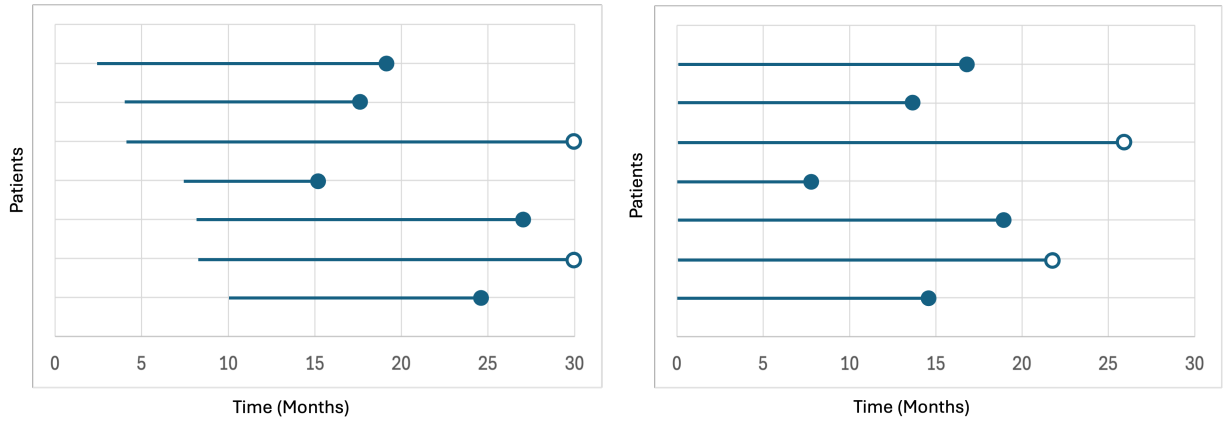
Survival endpoints include *Overall Survival* (OS) and *Recurrence-Free Survival* (RFS). These are vital metrics that play significant roles in assessing clinical outcomes. OS and RFS adhere to the *Standardized Definitions for Efficacy End Points* (STEEP) criteria, which were established in 2007 and updated in 2021.

OS ends in death from breast cancer, death from another cause, or death from unknown cause. It is considered the gold standard in oncology trials as it provides valuable information about the overall impact of various factors on the survival of individuals.

RFS ends in local recurrence, distant recurrence, death from breast cancer, death from another cause, or death from unknown cause. It is often used in early-stage cancer trials as it evaluates the effectiveness of treatments in delaying or preventing recurrence. The rates are commonly expressed as metrics over a span of five years (5Y) or ten years (10Y).

### 4.1.2 Censored Data

Clinical outcomes for many patients often remain unobserved throughout the duration of the study. Rather than being excluded from the study, these patients are classified as *censored*, or more precisely as *right-censored*. In most clinical studies, censored data often arise due to administrative reasons, encompassing both study and follow-up termination. This form of censoring, known as *administrative censoring*, is statistically independent of event times and does not introduce systematic bias into the analysis. Consequently, administrative censoring allows for robust and meaningful interpretations of survival data.[35] Figure 4.1 illustrates an example of time-to-event data generated from the timeline of patient enrollment to either event occurrence (i.e. uncensored) or follow-up termination (i.e. censored).




(a) (b)

*Figure. 4.1:* Example of survival analysis data, with censored events represented as empty dots, showing: (a) Timeline of patient enrollment to follow-up termination or event occurrence. (b) Time-to-event.

### 4.1.3 Survival Function

The *survival function* is the probability that an individual survives death or recurrence beyond a certain time. It is complementary to the *Cumulative Distribution Function* (CDF), which is the probability that the event occurs on or before this time.[36] Mathematically, the survival function is given by

$$S(t) = P(t_i > t) = 1 - P(t_i \leq t) \tag{4.1}$$

where,

$t_i$: unique event time

$P(t_i > t)$: probability that the time until the event $t_i$ is greater than t

$P(t_i \leq t)$: probability that the time until the event $t_i$ is less than or equal to t (i.e. CDF)

At the start of the study, the survival probability is 1. As time progresses, the survival probability decreases, reflecting the occurrence of events.

### 4.1.4 Hazard Function

The *hazard function* is the instantaneous rate of event occurrence at a certain time. Mathematically, it is expressed as the conditional probability that the event occurs in the short time interval, $t_i + \Delta t$, given that it has not occurred before over the width of the interval, $\Delta t$.[36] Taking the limit as the width of the interval goes down to zero, results in

$$h(t) = lim_{\Delta t \to 0} \frac{P(t \leq t_i < t + \Delta t | t_i \geq t)}{\Delta t}. \tag{4.2}$$

The relationship between the hazard function and the survival function can be derived from their definitions and basic calculus principles. The hazard function is determined to be equal to the partial derivative with respect to time of the logarithm of the survival function.[36] This is given by

$$h(t) = -\frac{\partial log[S(t)]}{\partial t}. \tag{4.3}$$

Over time, the hazard function can be constant, increase, decrease, or take some other pattern.[36]

### 4.1.5 Cumulative Hazard Function

The *cumulative hazard function*, which serves as a a measure of the total risk accumulated up to a specific time, is found by integrating the hazard function from the start of the study up to to this time. The result shows that the cumulative hazard function exhibits a reverse logarithmic relationship with the survival function.[36] This is given by

$$H(t) = \int_0^t h(u)du = -log[S(t)]. \tag{4.4}$$

Similarly, the survival function can be found from the hazard function by

$$S(t) = e^{-H(t)}. \tag{4.5}$$

While the survival function represents the probability of surviving beyond a particular time, the cumulative hazard function characterizes the total hazard encountered up to that point.[36]

### 4.1.6 Kaplan-Meier Survival Curve Estimator

#### The Estimator

The estimation of *survival distribution*, which represents the temporal distribution of event occurrences in a study, is a fundamental component of survival analysis. The process involves capturing the probability of events happening at different time points in order to gain a comprehensive understanding of the progression

of the studied phenomenon. Accurate estimation is crucial for drawing meaningful insights from the data.[35]

Ensuring accurate estimation necessitates the appropriate handling of censored cases. If censored cases are treated as events, it leads to estimates of survivals that are shorter than their actual durations. On the other hand, omitting censored cases results in a loss of data, a significant concern, especially in cancer studies where there is a high proportion of unobserved events.[35]

To address these challenges, the *Kaplan-Meier estimator* is often employed. This estimator considers all patients, including those with censored data, without treating unobserved events as actual events. It provides an unbiased estimate of the survival distribution, making it a valuable tool in survival analysis, especially in scenarios with censored data.[35]

The Kaplan-Meier estimates the survival at time $t_i$ at which at least one event has occurred by

$$S = \Pi_i \left(1 - \frac{d_i}{n_i}\right) \tag{4.6}$$

where $\Pi$ denotes the product of terms, $d_i$ is the number of individuals who experience an event by time $t_i$ and $n_i$ is the number of individuals who are neither censored nor had an event by time $t_i$ and hence remain *at risk.*[35]

**Resulting Graph**

The Kaplan-Meier estimator results in a staircase-like curve when visually represented as illustrated in figure 4.2. This characteristic arises from the stepwise nature of the estimator, which decreases by a step when an event occurs.[35]

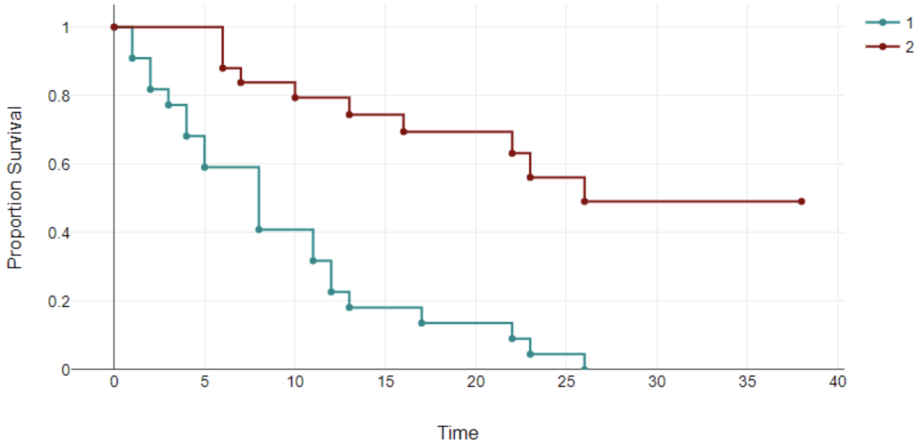



*Figure. 4.2:* Example of Kaplan-Meier curves: Curve 1 exhibits a steeper decline compared to curve 2, indicative of a higher frequency of event occurrences in group 1. Curve 2 does not reach zero, signaling the presence of censored data in group 2. This visual representation provides insights into the differences in survival experiences between the two groups. (Reproduced with permission from[37]. See Appendix A.)

Despite its simplistic appearance, the Kaplan-Meier curve is a powerful and widely used tool for analyzing survival data and estimating survival probabilities in clinical studies. At any specific point in time, the survival proportion or survival probability can be obtained directly from the graph by reading the y-axis value corresponding to that specific time on x-axis. The steepness of a Kaplan-Meier curve reflects the event rate; a steeper Kaplan-Meier curve indicates a higher event rate and a worse prognosis, while a flatter curve signifies a lower event rate and a better prognosis. It is noteworthy that the Kaplan-Meier estimator does not reach zero in the presence of censored data. This signifies that, at the end of the study or observation period, some individuals are still at risk of experiencing the event, but their outcomes are unknown due to censoring.[35]

**Interpretation of Results**

In studies assessing the efficacy of a treatment, patients are typically divided into a control group and a treatment group. The goal is to keep the curve for the treatment group as flat as possible, indicating a lower event rate and, consequently, a better prognosis. Conversely, in studies focused on developing a prognostic model, the aim is to create groups based on the model that are as distinct as possible, ensuring a notable separation between the curves representing different groups.

In the illustrative graph provided in figure 4.2, two distinct groups are represented through Kaplan-Meier curves. The blue curve exhibits a steeper decline, indicating a quicker decrease in survival probability, while

the red curve maintains a flatter trajectory. Upon reaching the 10-year mark, the blue group displays a mere 40% survival probability, whereas the red curve demonstrates a more favorable scenario with an 80% survival probability. Notably, the red curve does not reach zero by the end of the study, highlighting that, for many cases in this group, the event of interest did not occur. In interpreting these curves, the context matters. For instance, if the blue and red curves represent a control group (blue curve) and a treatment group (red curve) in a study evaluating a new cancer drug, then the graph strongly implies the effectiveness of the treatment as evidenced by the higher survival probability of the treatment group. On the other hand, if the curves test the capability of a prognostic model to categorize patients based on their risk profile, then the presence of what appears to be distinct low-risk (blue curve) and high-risk (red curve) groups, implies that the model has valuable discriminatory power and underscores its potential clinical utility in aiding personalised risk assessments and treatment decisions.

### 4.1.7 Log-Rank Test

#### The Test

The *logrank test*, alternatively referred to as the *Mantel-Cox test*, serves as a valuable tool for comparing survival experiences between clincal groups. It is widely embraced in medical research to assess not only the efficacy of diverse treatments but also to evaluate the stratification potential of prognostic models.[36]

It is crucial to ensure that any observed statistical differences are not solely attributed to disparities in group sizes or event occurrences. The comparison must be fair and robust to yield meaningful results. This fairness is achieved by implying a statistical principle, known as *conditioning*. By accounting for variations in the number of individuals at risk at any given time, as well as differences in the total number of events observed, the logrank test adjusts for potential biases and enables accurate comparisons of survival outcomes across different groups, ultimately enhancing the validity and reliability of the findings.[35]

At the core of its calculations, the logrank test uses *contingency tables*, such as 4.1. The contingency table summarises the numbers of events and number of patients at risk within each group, with a separate contingency table crafted for each unique time event.

*Table 4.1:* Contingency table summarizing the number of events and patients at risk in two groups at $t_i$ for use in a logrank test.[36]

| | Group 1 | Group 2 | Total |
|---|---|---|---|
| Event | $d_{1i}$ | $d_{2i}$ | $d_i$ |
| Non-Event | $n_{1i} - d_{1i}$ | $n_{2i} - d_{2i}$ | $n_i - di$ |
| At Risk | $n_{1i}$ | $n_{2i}$ | $n_i$ |

The null hypothesis of the logrank test states that the two groups have identical distribution curves. This means that the proportion of number of events (i.e. $d_i$) to total number at risk (i.e. $n_i$) in both groups must be equal for each and every unique event, $i$.[36]

The expected number of events in each group is this constant ratio ($d_i/n_i$) multiplied by total number of individuals at risk in the respective group.[36] For instance, for group $g$ and event $i$, the expected number of events $e_{gi}$ is determined by

$$e_{gi} = \frac{d_i}{n_i} n_{gi}. \tag{4.7}$$

When an event, $i$, occurs, the patient is not replaced signifying that individuals who have experienced this event are no longer considered at risk, thereby affecting the probabilities of subsequent occurrences.[36]

**The Resulting Test Statistic and Significance Level**

The log-rank test generates a test statistic, often denoted as the $Z^2$ *statistic.*[35] The $Z^2$ statistic can be approximated as the more common $\chi^2$ statistic, which is calculated as the sum of differences between the observed, $d_{gi}$, and expected numbers, $e_{gi}$, of events divided by the sum of expected numbers, $e_{gi}$.[38] That is

$$\chi^2 = \sum_g \frac{(\sum_i d_{gi} - \sum_i e_{gi})^2}{\sum_i e_{gi}}. \tag{4.8}$$

The *degrees of freedom* of a $\chi^2$ is determined as the number of groups being compared minus one. The associated *p-value* represents the area under the probability density curve to the right of the observed $\chi^2$ value, indicating the likelihood of obtaining a result as extreme as, or more extreme than, the observed result. Larger differences between observed and expected numbers yield larger $\chi^2$ values, leading to smaller p-values and suggesting a more significant departure from expected outcomes.[35]

The $\chi^2$ and corresponding significance level can be read from statistical tables for different degrees of freedom. Figure 4.3 shows $\chi^2$ distributions with varying degrees of freedom, and the value of $\chi^2$ corresponding to a significance level of 5% for each of them.

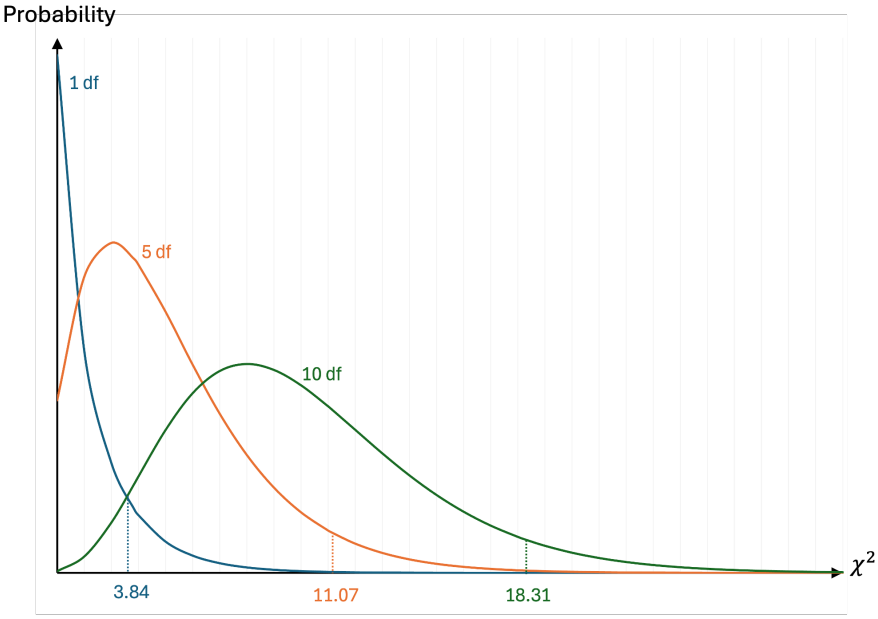



*Figure. 4.3:* $\chi^2$ distributions with varying degrees of freedom: 1df, 5 df and 10 df. The graph shows $\chi^2$ values corresponding to a significance level of 5%. For a test comparing two groups (i.e. 1 df) to be statistically significant, it must produce a $\chi^2$ higher than 3.84.

### Interpretation of Results

In the context of treatment efficacy studies, a small p-value signifies a statistically significant difference in survival between the control and treatment groups, supporting the efficacy of the treatment under investigation. Similarly, in prognostic model stratification studies, a small p-value indicates a significant difference in survival between high-risk and low-risk groups. This statistical significance provides evidence of the ability of the model to effectively stratify patients into distinct survival groups based on their respective risk profiles.

In most clinical studies, a p-value is deemed significant if it falls below the conventional threshold of 5% or 0.05. This widely accepted significance level is a standard criterion in statistical analysis, indicating a low probability that observed results occurred by chance alone and affirming the meaningfulness of the results within the context of the study.

In most clinical trials, the focus is to compare two groups, equating to one degree of freedom. The significance level is typically set at 5% (i.e. 0.05), which corresponds to a $\chi^2$ value of 3.84. Logrank tests that result in $\chi^2$ values higher than 3.84, and accordingly p-values less than 0.05, indicate a low probability that

observed results occurred by chance alone and lead us to conclude that the compared groups are statistically different.[36]

### 4.1.8 Cox Proportional Hazards Regression

#### The Model

The *Cox proportional hazards regression* is a statistical technique that aims to assess the relationship between one or more covariates and the occurrence of a clinical outcome such as death or recurrence after a given time interval. The covariate can be a continuous variable such as age or a categorical variable such as sex. For a number of covariates ($x_1$, $x_2$, ....,$x_n$) with corresponding regression coefficients ($\beta_1$, $\beta_2$, ....,$\beta_n$), the hazard is found as the product of the *baseline hazard*, which is the hazard if all covariates were set to zero, and a positive constant expressed as a log-linear function of the covariates, as follows

$$h(t|x) = h_0(t) \sum_{p}^{n} e^{\beta_p x_p}. \tag{4.9}$$

The baseline hazard, $h_0$, is time-dependent while the covariates term, $b_p x_p$, is time independent.[39]

#### The Resulting Hazard Ratio

For one covariate, say $x_1$, at value $a$, the hazard is

$$h(t|a) = h_0(t) e^{\beta a}. \tag{4.10}$$

For the same covariate, at value $b$, the hazard is

$$h(t|b) = h_0(t) e^{\beta b}. \tag{4.11}$$

Dividing

$$\frac{h(t|a)}{h(t|b)} = e^{\beta(a-b)}. \tag{4.12}$$

If the change in covariate is equal to 1, then the ratio is defined as the *hazard ratio* ($HR$)[39]. $HR$ is equal to the exponentiated regression coefficient, as given by

$$HR = e^{\beta}. \tag{4.13}$$

**Interpretation of Results**

A hazard ratio of 1 indicates no effect on risk, while a hazard ratio greater than 1 indicates an increase in risk and a hazard ratio less than 1 indicates a decrease in risk.[40] For example, a HR of 1.12 indicates that the risk associated with the covariate increases by 12% for every unit increase in the covariate, while a HR of 0.95 indicates that the risk decreases by 5% for each unit increase in covariate.[39] In cancer studies, the former covariate is referred to as a bad prognostic factor, and the latter covariate is referred to as a good prognostic factor.[40]

**Goodness of Fit**

The goodness-of-fit of a Cox proportional hazards model can be tested using the *log-likelihood ratio test*, also known as the *G squared test*. The null hypothesis of the log-likelihood ratio test states that a simpler model with no covariates exists. A p-value of 0.05 or less leads us to reject the null hypothesis and to conclude that the proposed Cox proportional hazards model is statistically acceptable.[38]

## 4.2 Performance Metrics

Clinical tests serve as invaluable tools for healthcare professionals. They aid in making informative decisions, initiating timely treatment interventions, and ultimately improving patient outcomes. These tests are categorized based on their intended aim into diagnostic and prognostic tests. Diagnostic tests are designed to identify individuals who have a particular disease or condition. Conversely, prognostic tests are focused on identifying patients at high risk of developing certain outcomes, such as disease progression or adverse events.

In the context of diagnostic testing, True Positives (TP) occur when a test correctly identifies diseased individuals as positive, while True Negatives (TN) occur when healthy individuals are correctly identified as negative. However, due to inherent inaccuracies, diagnostic tests may also produce False Positives (FP), incorrectly identifying healthy individuals as diseased, or False Negatives (FN), failing to detect the disease in individuals who are actually diseased. These outcomes are summarised in table 4.2. These terms are equally applicable to prognostic tests, where the emphasis shifts to identifying individuals at risk rather than those currently affected by the condition.

*Table 4.2:* Possible outcomes of a clinical test based on actual values and predicted values.

| | | Predicted Value | |
|---|---|---|---|
| | | Positive | Negative |
| Actual Value | Positive | True Positive (TP) | False Negative (FN) |
| | Negative | False Positive (FP) | True Negative (TN) |

### 4.2.1 Sensitivity, Specificity and Accuracy

#### Sensitivity

The sensitivity of a clinical test quantifies its ability to correctly identify those patients with the disease. By definition, it is equal to the True Positive Rate (TPR). It is calculated as the proportion of true positives to all diseased individuals.[36] The sensitivity is given by

$$\text{sensitivity} = TPR = \frac{TP}{TP + FN}. \tag{4.14}$$

A test characterized by high sensitivity indicates a small number of false negatives, meaning it is highly effective at correctly identifying individuals with the disease.

#### Specificity

In contrast, the specificity of a clinical test quantifies its ability to correctly identify those patients without the disease. It is equal as 1 minus the False Positive Rate (FPR). It is calculated as the proportion of true negatives to all healthy individuals.[36] The specificity is given by

$$\text{specificity} = 1 - FPR = \frac{TN}{TN + FP}. \tag{4.15}$$

A test with high specificity indicates a small number of false positives, meaning any positive test outcome is likely to be a true positive.

#### Sensitivity and Specificity Trade-Off

A perfect clinical test ideally possess 100% sensitivity and 100% specificity, effectively identifying all individuals with the disease while accurately excluding those without it. However, achieving such perfection

is inherently unattainable in practice. In reality, clinical tests exhibit a trade-off between sensitivity and specificity. Some tests have high sensitivity, effectively capturing a large proportion of true positive cases but may have lower specificity, leading to an increased likelihood of false positives. Conversely, other tests have high specificity, minimizing false positive results but potentially missing some true positive cases due to lower sensitivity. These complementary characteristics may be leveraged through a two-step diagnostic approach. Initially, a high-sensitivity test is used as a *screening test* on the whole population where the primary objective is to rule out the presence of the disease, ensuring that individuals who truly have the condition are not overlooked. Subsequently, a high specificity test is used as *confirmatory test* to confirm the diagnosis of individuals with a positive screening.[36]

#### Accuracy

The accuracy of a clinical test quantifies the proportion of overall true test outcomes to overall tests.[36] It is calculated as

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{4.16}$$

A test with high accuracy indicates a small number of false positives and false negatives.

### 4.2.2 Predictive Values

#### Positive Predictive Value

Positive Predictive Value (PPV) assesses the reliability of a positive test result in correctly identifying individuals with the disease or condition of interest. It quantifies the probability of an individual having the condition of interest (TP) among those with a positive test result (TP + FP).[36] The $PPV$, is given by

$$PPV = \frac{TP}{TP + FP}. \tag{4.17}$$

#### Negative Predictive Value

Similarly, Negative Predictive Value (NPV) assesses the likelihood that a negative test outcome accurately indicates the absence of the condition in an individual. It provides a measure of how likely it is that a positive test outcome accurately reflects the presence of the condition in an individual. It quantifies the probability of an individual not having the condition of interest (TN) among those with a negative test result (TN +

FN).[36] The $NPV$, is given by

$$NPV = \frac{TN}{TN + FN}. \tag{4.18}$$

**Sensitivity and Specificity vs. PPV and NPV**

To easily see the difference between sensitivity and specificity, on one hand, and PPV and NPV, on the other hand, they are summarised in table 4.3

*Table 4.3:* Sensitivity and specificity are read horizontally, while PPV and NPV are read vertically, as the ratio of the first variable to the total of the first and second variables.[36]

| | | Predicted Value | | |
|---|---|---|---|---|
| | | Positive | Negative | |
| Actual Value | Positive | True Positive (TP) | False Negative (FN) | Sensitivity |
| | Negative | False Positive (FP) | True Negative (TN) | Specificity |
| | | PPV | NPV | |

### 4.2.3 Receiver Operator Characteristic Curves

**ROC Curve**

A *Receiver Operator Characteristic* (ROC) curve is generated by plotting the sensitivity versus 1-specificity at varying thresholds as shown in figure 4.4. It is used to assess the overall diagnostic performance of a clinical test or a prognostic model, and to compare the performance of two or more tests. It is also used to select an optimal cut-off value for determining the presence or absence of a disease.[36]

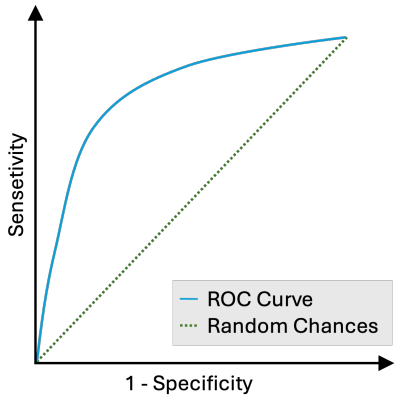



*Figure. 4.4:* Example of a Receiver Operating Characteristic (ROC) curve. It is used to assess the overall diagnostic performance of a clinical test or a prognostic model and to select its optimal cut-point.

### AUC

The overall diagnostic performance of a clinical test hinges on its capacity to make binary predictions effectively. This evaluative criterion is commonly assessed using the *Area Under the Curve* (AUC) of an ROC curve as illustrated in figure 4.5.

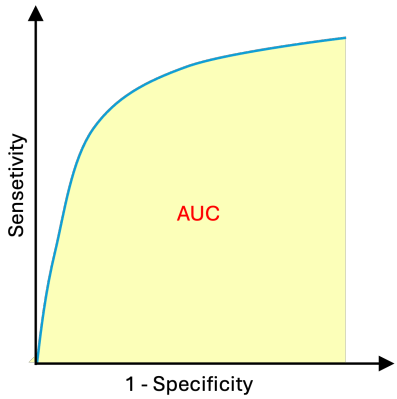



*Figure. 4.5:* The shaded area in known as the Area Under the Curve (AUC) of an ROC curve. It is used to assess the overall usefulness of a clinical test. A high AUC indicates a strong ability to separate classes, while an AUC of 0.5 or less indicates no separation capacity.

A high AUC value close to 1 signifies robust discrimination between classes, indicating a strong ability to distinguish between positive and negative cases. An AUC value of less than 0.5 suggests that the test has limited discriminatory power, essentially indicating chance-level performance where positive and negative cases are classified randomly.[36] The levels of discriminatory levels are summarised in table 4.4.

*Table 4.4:* Discriminatory power of a clinical test based on AUC.[36]

| Area Under Curve (AUC) | Discrimantory Power |
|---|---|
| 0.9 - 1 | Excellent |
| 0.8 - 0.9 | Very Good |
| 0.7 - 0.8 | Good |
| 0.6 - 0.7 | Acceptable |
| 0.5 - 0.6 | Bad |
| $\leq 0.5$ | Not Useful |

When assessing a single clinical test, an AUC higher than 0.6 is required, indicating a reasonable discriminatory ability between positive and negative cases. Additionally, when comparing tests, the one with higher AUC is preferred, signifying superior discriminatory power and effectiveness in distinguishing groups.[36]

**Cut-Points**

When using a test with a continuous variable to classify patients into categories, a cut-point must be chosen. This cut-point serves as a threshold for distinguishing between the two groups. However, since no test is perfect, there will always be some overlap between the groups, meaning that individuals near the cut-point may be misclassified.

The ROC curve is a graphical representation of the trade-off between sensitivity and specificity at different cut-points. The optimal cut-point should be selected to find a compromise between sensitivity and specificity, tailored to the specific clinical application and the consequences of false positives and false negatives.[41] For instance, as discussed in 4.2.1, screening tests require higher sensitivity to ensure that most individuals with the condition are identified, while confirmatory tests require higher specificity to ensure that those identified as having the condition are truly positive.

There are different methods to determine the optimal cut point. One of the most common is to maximise the Youden Index, which is the distance between the ROC curve and the random chances line, as illustrated in figure 4.6.[41] The Youden Index is expressed in terms of sensitivity, $Se$, and specificity, $Sp$, as

$$\text{Youden Index} = Se - (1 - Sp) = Se + Sp - 1. \tag{4.19}$$

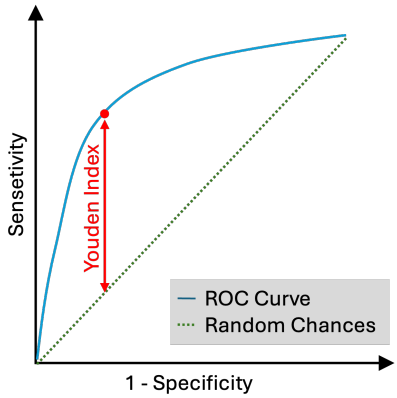



*Figure. 4.6:* The Youden Index is calculated at the maximum distance between the ROC curve and the random chances line. It is used to find the optimal cut-point of a clinical test.

This method finds the cut-point that maximizes the sum of sensitivity and specificity, effectively treating them with equal importance.[41] In many applications, either sensitivity or specificity is more important, and alternative methods are utilised.

## 4.3 Conclusion

Survival analysis techniques are essential in developing and assessing prognostic models. Cox proportional hazards regression is commonly employed to predict survival or to construct prognostic models based on given covariates. Kaplan-Meier curves are instrumental in visually representing survival probabilities over time, while the log-rank test serves as a valuable tool for comparing survival between groups.

Moreover, the performance of prognostic models can be evaluated using various metrics, including sensitivity, specificity, accuracy, positive predictive value, and negative predictive value. Lastly, ROC curves are employed to illustrate the intrinsic trade-off between sensitivity and specificity, with the AUC providing a measure of the overall diagnostic performance of the model across various thresholds.

# Part II

# Digistain as a Prognostic Model

# Chapter 5

# The Digistain Technique

This chapter serves as an introduction to Digistain, a mid-infrared technique we intend to use for developing a prognostic model to inform treatment decisions and mitigate chemotherapy overtreatment. It begins with an overview of the fundamental principles of infrared absorption spectroscopy and its applications in pathology. Then, it explores the principle of operation, index formulation, flow of work, historical evolution, strengths, and weaknesses of Digistain.

## 5.1 Overview of Infrared Absorption Spectroscopy

The study of how light and matter interact forms an important branch of science known as *spectroscopy*. In this section we review the basic principles of spectroscopy.

### 5.1.1 Light

The electromagnetic spectrum is not limited to visible light; it includes a range of regions characterized by different wavelengths and frequencies as illustrated in figure 5.1. As wavelength becomes longer along the electromagnetic spectrum, frequency decreases and so does energy.

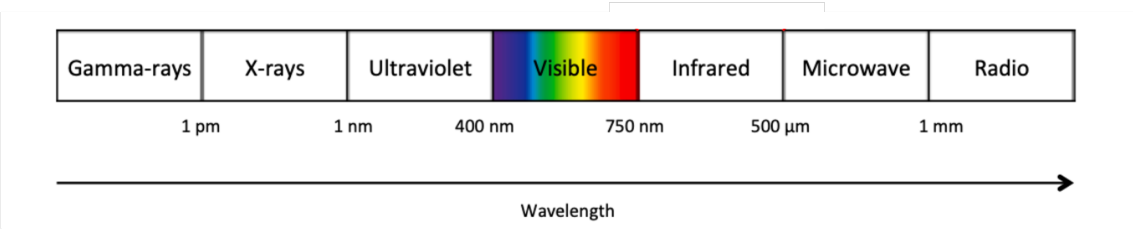



*Figure. 5.1:* The electromagnetic spectrum consists of different regions. Increase in wavelength corresponds to decrease in frequency and electromagnetic energy.

### 5.1.2 Matter

The atom is the building block of matter. It consists of a nucleus surrounded by electrons. These electrons occupy defined orbitals or energy levels around the nucleus. When electrons occupy the lowest energy level, the atom is in its ground state. Conversely, when electrons occupy higher energy levels, the atom is in an excited state. These states are illustrated in figure 5.2. An atom can only exist in one of these states, transitioning distinctly between them rather than existing in intermediate states.[42]

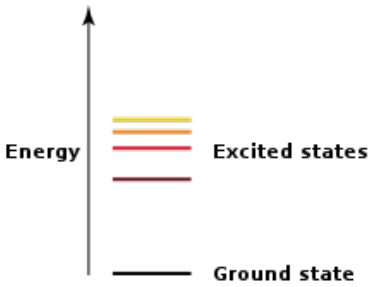
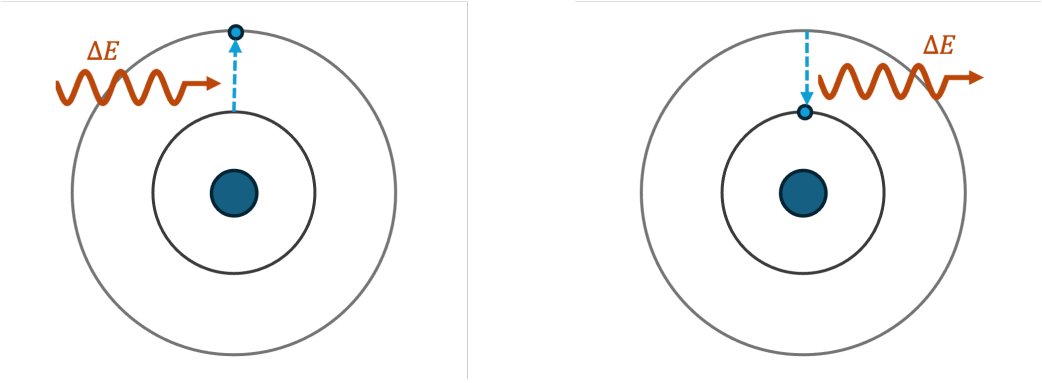



*Figure. 5.2:* The ground and excited states of an atom are determined by its electronic configuration. An atom can occupy either the ground state, characterized by the lowest energy configuration, or an excited state with higher energy. It cannot exist in between these discrete energy levels.

Atoms come together to form molecules, which also have defined energy levels. However, because molecules are composed of multiple atoms, their energy levels are more complex. These levels can be *electronic*, governed by the arrangement of electrons within their orbits; *vibrational*, relating to the stretching and bending of interatomic bonds; and *rotational*, indicative of molecular rotation around its axis.[42]

### 5.1.3 Forms of Interaction

The interaction between light and matter can take different forms, but mainly: *absorption*, where atoms or molecules absorb light, *emission*, where they emit light; and *scattering*, where light is redirected in different directions by existing particles.[42] Absorption and emission are illustrated in figure 5.3.




(a) (b)

*Figure. 5.3:* Main light-matter forms of interaction: (a) Absorption: The atom absorbs light of energy corresponding to the difference in energy between its two states. (b) Emission: The atom emits light of energy corresponding to the difference in energy between its two states.

### 5.1.4 Infrared Absorption Spectroscopy to Probe Biological Samples

The absorption of light by a molecule can result in various molecular processes, including electronic excitation, vibration, or rotation, depending on the energy of the light. Infrared absorption spectroscopy focuses on the absorption of light in the infrared region. The infrared region spans the range from 750 nm to 500 $\mu$m and is further subdivided into Near-Infrared (NIR), Mid-Infrared (MIR) and Far-Infrared (FIR) as shown in figure 5.4.[42]

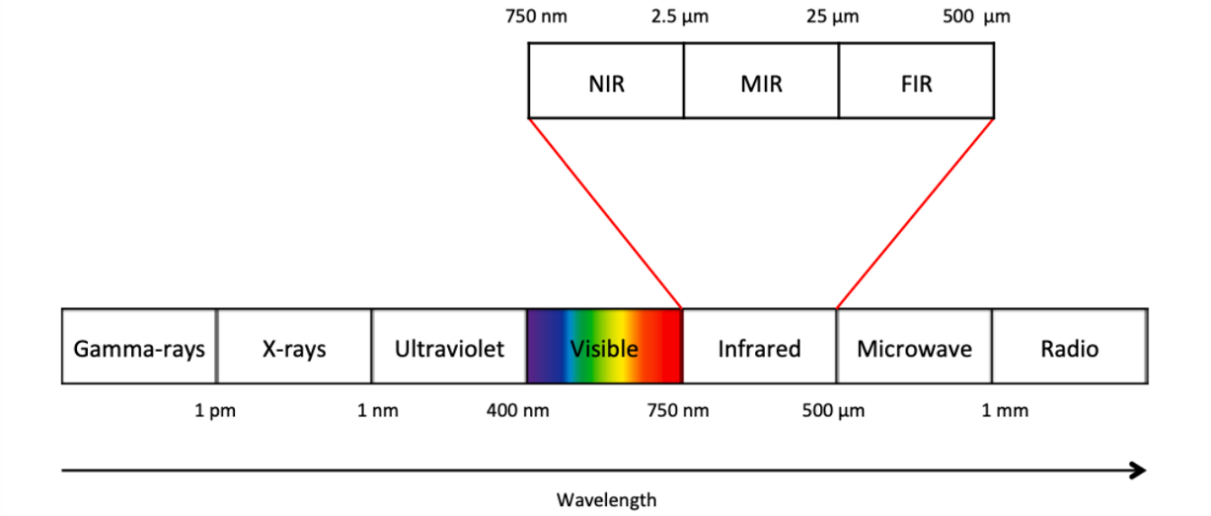



*Figure. 5.4:* The infrared region of the electromagnetic spectrum stretches from 750 nm to 500 $\mu$m and is subdivided into Near-Infrared (NIR), Mid-Infrared (MIR) and Far-Infrared (FIR).

When infrared light is absorbed by a molecule, its energy is often not sufficient to cause electronic excitation but it induces change in molecular vibrations. This is why infrared spectroscopy is often referred to as *vibration spectroscopy*. Molecules absorb infrared light only when the frequency matches one of the frequencies of their fundamental modes of vibration. Consequently, each molecule absorbs light at signature frequencies, resulting in a unique fingerprint that facilitates its identification when spectra of unknown samples are acquired.[42]

Biological samples are commonly studied using light in the infrared region in general, and mid infrared region in particular. This is because infrared light has frequencies that match the vibration modes of molecules, and allow for their identification without damaging the sample.

### 5.1.5 Transmittance and Absorbance

When light of intensity, $I_0$ , is incident on the sample. Part of the light is absorbed by the sample, and the remaining light, $I$ , is transmitted through the sample as illustrated in figure 5.5.

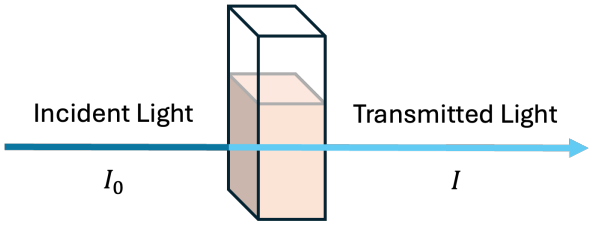



*Figure. 5.5:* Light of intensity, $I_0$ , hits the sample. Some is absorbed and the remaining light, $I_0$ , is transmitted.

The transmittance, $T$ , is the ratio of transmitted intensity, $I$ , to incident intensity, $I_0$.[43] It is given by

$$T = \frac{I}{I_0}. \tag{5.1}$$

The absorbance, $A$, is defined as the negative logarithm of the transmittance.[43] It is expressed as

$$A = -log_{10}T. \tag{5.2}$$

### 5.1.6 Beer-Lambert Law

The Beer-Lambert law states that the absorbance of light by a substance is directly proportional to the concentration of the absorbing molecules present in the sample.[43] It is written as

$$A = \epsilon c l \tag{5.3}$$

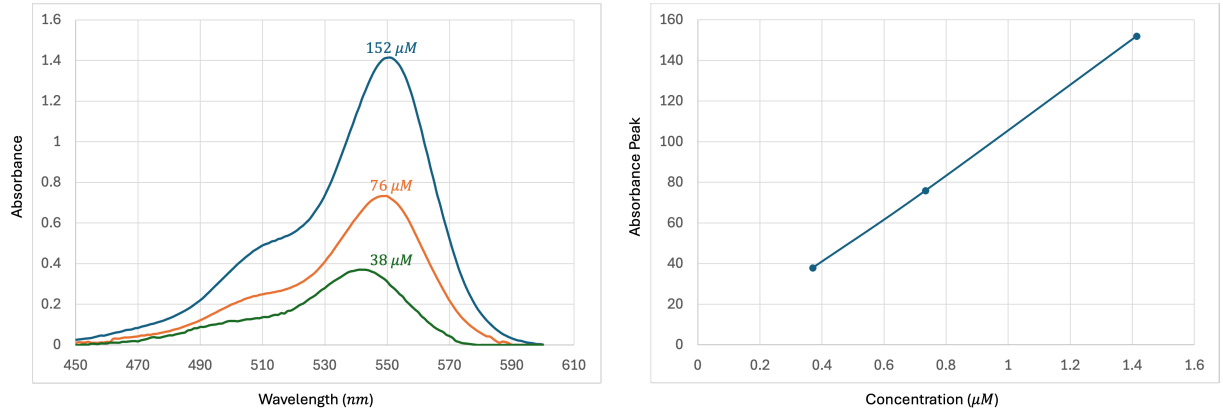
where,

$\epsilon$: absorption coefficient

$c$: concentration

$l$: path length

Therefore, for samples of consistent thickness, a higher absorbance at any given wavelength suggests a higher concentration of absorbing molecules within the sample. Figure 5.6 illustrates this relationship by plotting absorbance as a function of wavelength for different concentrations. It shows that the highest absorbance is observed for the highest concentration and vice versa. Furthermore, when the absorbance peaks are plotted against concentration, a straight line is observed, demonstrating the linear relationship between absorbance and concentration as described by the Beer-Lambert law.[43]




(a) (b)

*Figure. 5.6:* Example of Beer-Lambert Law: (a) Absorbance as a function of wavelength at different concentrations shows the proportional relationship between absorbance and concentration. (b) Absorbance peaks as a function of concentration results in a straight line as described by Beer-Lambert law. (Containing data from[43])

### 5.1.7 Fourier Transform Infrared Spectrometers

*Fourier Transform Infrared* (FTIR) spectrometers are commonly used to acquire transmission measurements. Radiation from the source goes through a beam splitter. One beam reflects from a fixed mirror while the second beam reflects from a movable mirror. The reflected beams travel back to the beam splitter and then

to the sample. The position of the movable mirror governs the path difference between the two beams, which in turn governs the wavelengths that interfere constructively when they recombine. For each position of the movable mirror, the transmission is acquired. A plot of transmitted radiation versus the position of the mirror, known as an *interferogram*, is recorded. The interferogram is then loaded to a computer and Fourier transformed to get a plot of transmitted radiation versus wavelength, frequency or wavenumber, known as a *transmission spectrum*, or a plot of absorbed radiation versus wavelength, frequency or wavenumber, known as an *absorption spectrum*. FTIR spectrometers have the ability to acquire spectra in a relatively short amount of time with high signal-to-noise ratios.[44]

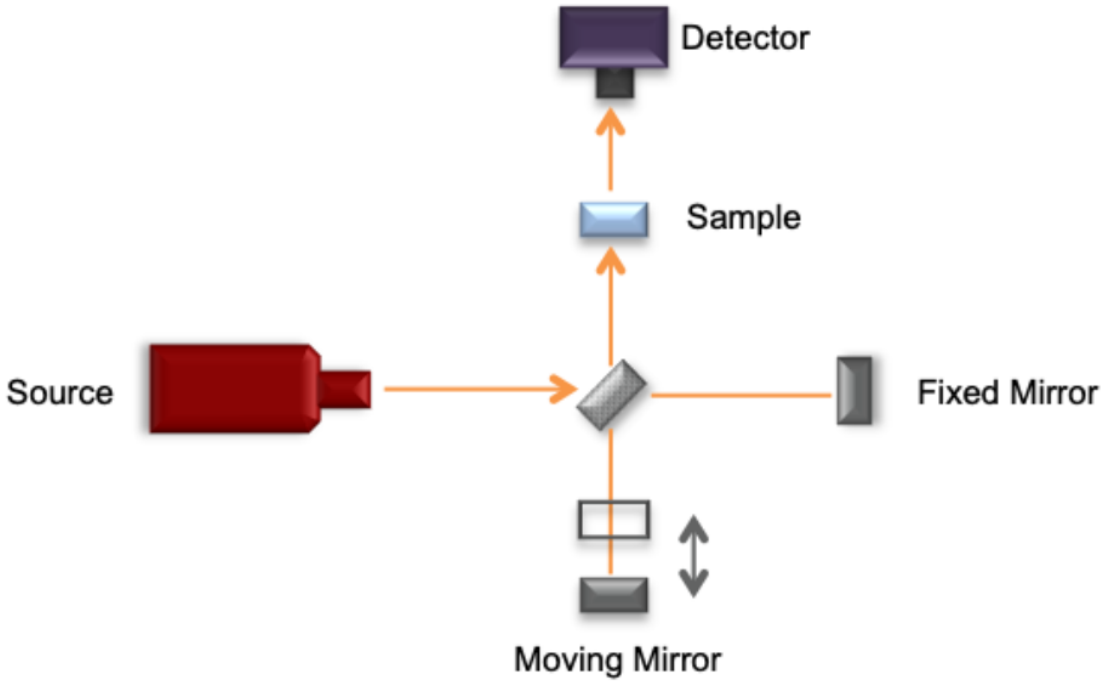



*Figure. 5.7:* Fourier Transform Infrared (FTIR) spectrometer: A ploychromaic light source is directed onto a beam splitter. One beam reflects from a fixed mirror while the second beam reflects from a moving mirror. The two beams travel different distances and hence their interference is destructive except at a certain wavelength. Changing the position of the moving mirror results in a different wavelength. The spectrum is scanned at every position.

## 5.2 Spectroscopic Techniques in Pathology

Spectroscopic techniques, particularly FTIR and Raman spectroscopy, are valuable tools in cancer research due to their ability to provide detailed molecular information about tissue samples. These techniques enable researchers to detect and analyze structural and chemical changes associated with various diseases, including cancer. In this section, these techniques are broadly discussed.

### 5.2.1 Fourier Transform Infrared Techniques

FTIR spectroscopy is widely used to differentiate healthy tissue from cancerous tissue or distinguish between various cancer subtypes. Some research groups focus on tissue samples for their ability to directly capture local biochemical changes in diseased areas. Other groups explore serum samples as a less invasive alternative, aiming to detect disease-specific biochemical changes reflected in blood composition. This approach has the potential for early and non-invasive diagnostics.[45]

To interpret the complex spectral data generated by FTIR, researchers often rely on advanced computational techniques, including Artificial Intelligence (AI). AI is employed to identify relevant peaks in the spectra and correlate them with specific biochemical changes. Machine learning models are trained and tested on spectral datasets, enabling the development of algorithms that can distinguish between different sample types with high accuracy. The combination of FTIR and AI-driven analysis is opening new avenues for pathology research, providing the potential for rapid, reliable, and non-invasive diagnostic techniques.[45][46]

Alternatively, some researchers focus on well-characterized spectral peaks associated with known biochemical compounds, such as proteins, lipids, or nucleic acids. This targeted approach relies on prior biochemical knowledge and avoids the need for AI-driven pattern recognition. By concentrating on established markers, these studies provide interpretable results, often focusing on specific peaks indicative of pathological changes. The Digistain technique employed in this thesis follows this approach, targeting predefined biochemical features to deliver reliable and interpretable results.[47]

### 5.2.2 Other Spectroscopic Techniques

Raman spectroscopy, similar to FTIR spectroscopy, provides a molecular fingerprint of compositions through its spectrum. However, unlike FTIR, Raman relies on scattering rather than absorption. A key disadvantage of Raman spectroscopy is its low cross-sections, which restrict analysis to small areas. Given the heterogeneity of cancer, this limitation makes Raman spectroscopy less favorable than FTIR-based techniques.[48] Despite these challenges, ongoing studies are increasingly combining Raman spectroscopy with deep learning algorithms to classify cancer types. Some research focuses on identifying biomarkers that can differentiate between various cancer forms. Once validated, these biomarkers could potentially guide treatment decisions. However, to date, no biomarkers identified through Raman spectroscopy have been clinically registered.[49][50]

Another, though less commonly used, technique is thermography, or thermal imaging. It involves capturing temperature-based images of tissue samples. Cancerous cells and their surrounding environments exhibit increased thermal activity due to elevated metabolic rates and enhanced vascular circulation. This method

helps distinguish cancerous from non-cancerous tumors and is particularly valuable for early detection. However, its application remains limited, with no established role in treatment decisions, including chemotherapy planning.[51 52]

## 5.3 Principle of Operation of Digistain

The NCR serves as a significant proxy for elevated tumor DNA content in various solid cancers, as discussed in section 2.6. Digistain provides an approach for determining NCR by leveraging the inherent biochemical differences between the nucleus and cytoplasm. The nucleus has a high concentration of phosphate moieties which have an absorption peak at 1234 $cm^{-1}$ (8.1 $\mu m$), while the cytoplasm has a high concentration of amide moieties which have an absorption peak at 1650 $cm^{-1}$ (6.1$\mu m$), as evident in the example given in figure 5.8. The ratio of absorbances is, by Beer-Lambert Law, proportional to the ratio of concentrations, and accordingly serves as a measure of NCR. This realization forms the foundation of the Digistain methodology, offering a reliable, quantifiable, and observer-independent measure of malignancy.[47 14]

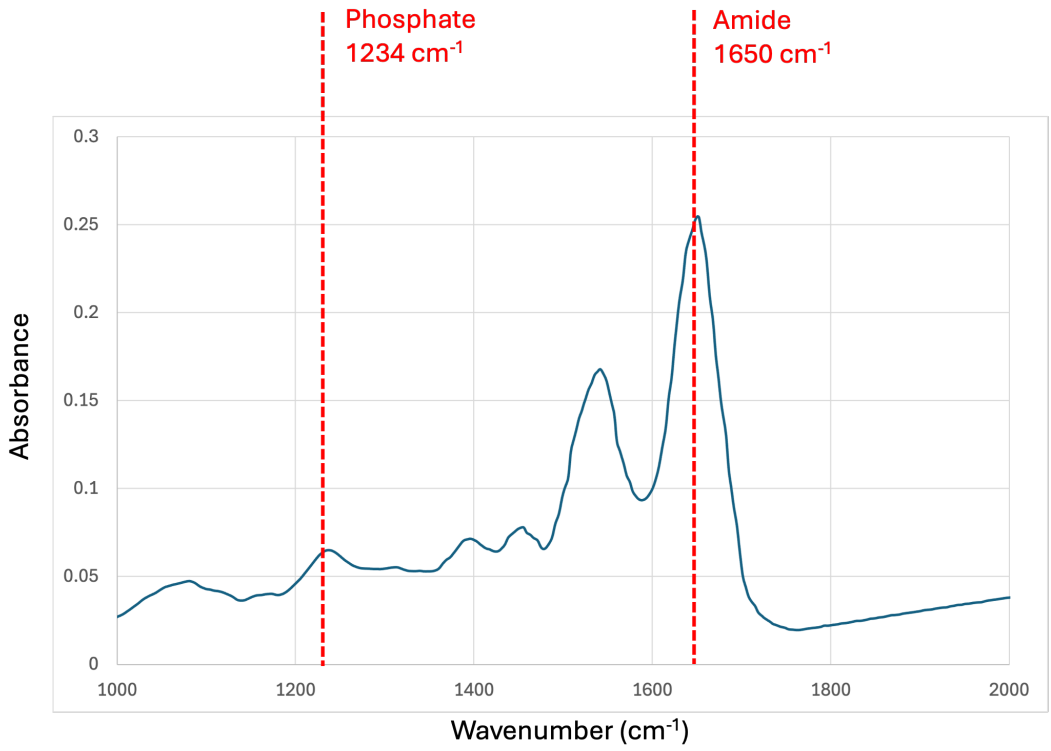



*Figure. 5.8:* Absorbance spectrum of a breast cancer tissue sample, exhibiting a peak at 1234 $cm^1$ caused by the phosphate moieties in the nucleus and another at 1650 $cm^1$ caused by the amide moieties in the cytoplasm.

## 5.4 The Digistain Index

The Digistain Index (DI) is defined as the ratio of amide absorbance to phosphate absorbance, making it inversely proportional to NCR, as given by

$$DI = \frac{A_{AP}}{A_{PP}} \tag{5.4}$$

where,

$A_{AP}$: Absorbance at Amide Peak (i.e. 1650 $cm^{-1}$)

$A_{PP}$: Absorbance at Phosphate Peak (i.e. 1234 $cm^{-1}$)

## 5.5 Workflow of Digistain

The tissue samples used in the Digistain study are sourced from tissue banks, where they undergo careful selection and treatment based on the specific requirements of the study. Upon receipt of the samples, the region of interest is selected with the assistance of a histopathologist. The slides are then de-paraffinized, and absorption measurements are taken. Finally, the DI is calculated. In this section, these steps are discussed.

### 5.5.1 Tissue Sample Sourcing

Breast cancer samples are typically preserved as FFPE blocks. For the Digistain procedure, two adjacent sections, each approximately 4 µm thick, are cut from each FFPE block. The first section is mounted on a standard glass slide and stained with hematoxylin and eosin dyes. The second section is mounted on a calcium fluoride (CaF) slide without staining to preserve its chemical integrity for infrared analysis.[47 14]

### 5.5.2 Region of Interest (ROI) Selection

After the slides are received, the H+E-stained slide is either physically sent to a histopathologist or scanned and transmitted digitally. The histopathologist selects a Region of Interest (ROI), typically focusing on the most cancerous portion of the tissue, and marks it for further analysis. This marked ROI is then used to guide the measurements on the unstained slide. The size of the selected region can vary depending on the characteristics of the tissue sample.

### 5.5.3 De-Paraffinization

The unstained slide undergoes a de-paraffinization process to remove the paraffin embedding medium and prepare the tissue for analysis. This process involves sequential heating and immersion in solvent baths to ensure that the tissue is properly prepared for infrared measurements.

### 5.5.4 Absorption Measurements

Absorption measurements are taken according to the specific version of the Digistain technology used, as detailed in the next section.

### 5.5.5 DI Calculation

External codes, which are specifically written for this purpose, are employed to read the absorption data and calculate the DI.

## 5.6 The Historical Evolution of Digistain

Digistain was initially developed as a free-standing spectrometer, with promising early results. However, over time, technical challenges arose due to the replacement of critical components and the sensitive nature of the measurements. At the outset of this thesis, the Digistain project existed as a collection of individual components that required careful alignment and integration. Initial efforts focused on improving both its hardware and software. However, it soon became clear that transitioning to an FTIR setup would provide a more practical and scalable solution, enabling faster and more efficient patient applications compared to engineering a bespoke hardware setup. This section provides an overview of the evolution of Digistain but does not include detailed discussions on circuit designs or programming code developed during the improvement phase, as these aspects are not directly relevant to the focus of this thesis.

### 5.6.1 Digistain as a Free Standing Spectrometer

#### Setup

The first Digistain prototype was constructed as a U-shaped spectrometer, consisting of an infrared source, a shutter, a filter wheel, a sample slide holder, and an infrared camera. A schematic diagram of the Digistain prototype is shown in figure 5.9.

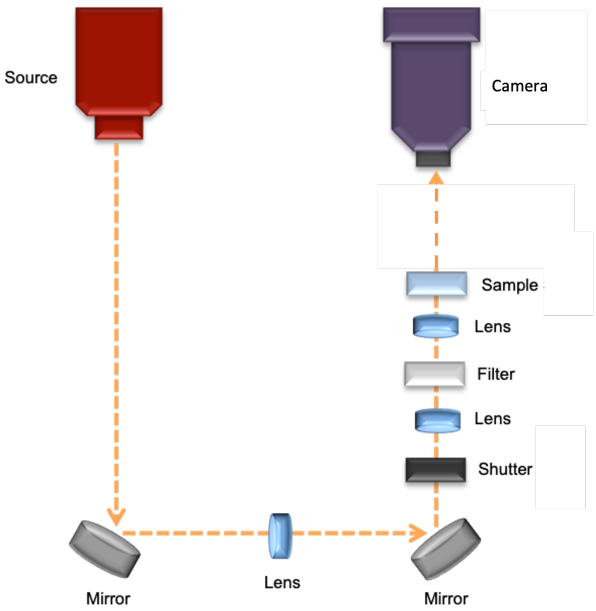
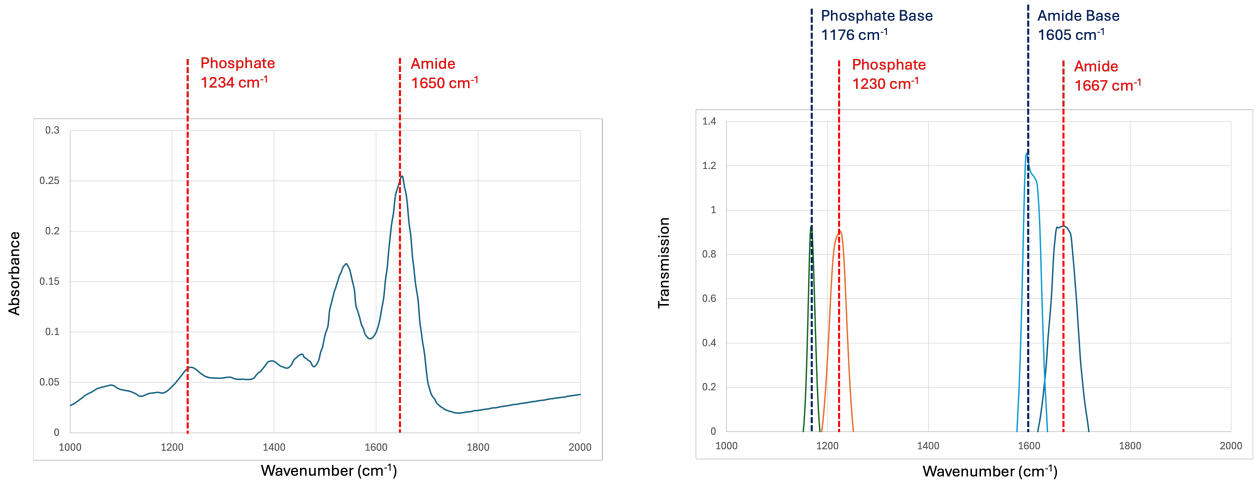



*Figure. 5.9:* The U-shaped Digistain prototype, consisting of an infrared source, a shutter, a filter wheel, a sample slide holder, an infrared camera in addition to extra mirrors and lenses.

The filter wheel featured four bespoke infrared bandpass filters: two designed for the amide peak and phosphate peak, and two corresponding to the amide base and phosphate base. The latter two filters were necessary to correct for possible contributions from other moieties. The filters are shown in figure 5.10.

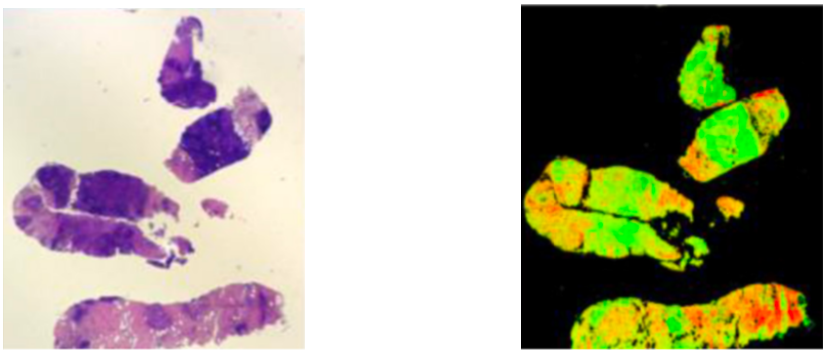



(a) (b)

*Figure. 5.10:* Correspondence between filters used in Digistain and absorbance spectrum of a human biopsy tissue sample. (a) Absorbance spectrum of a human biopsy tissue sample. (b) Bespoke infrared bandpass filter transmission spectra.

### Measurements

Images were captured using each of the four filters. For each filter, an image was taken with the shutter open and another with the shutter closed, allowing correction for background contributions. Additionally, images were captured with the sample in place and with the sample removed to account for non-uniform light distribution across the image. This process resulted in a total of 16 images, which were processed to produce a spatial map of Digistain indices. The map was digitally stained to mimic H+E staining without the need for any physical staining. The resulting image encoded both the morphological and chemical information of the sample, allowing for direct comparison with an adjacent H+E-stained slide, as shown in figure 5.11. From this color-coded image, any region could be selected, and the average DI within it could be calculated.[14]



(a) (b)

*Figure. 5.11:* Images of a breast biopsy section: (a) Visible light image of a section stained with hematoxylin and eosin dyes; (b) False-colored image of an adjacent section digitally stained as part of the Digistain protocol. (Reproduced from[14]. No permission required. See Appendix A.)

### Pilot Study

A pilot study of 71 breast cancer tissue samples was carried out in 2015. The samples were chosen at random from FFPE blocks collected between 1999 and 2003. They were of female patients, aged 30 - 84 at the time of the biopsy, and who had a follow-up 5 years later.[14]

A highly significant correlation was found between DI values and tumor grades, with a p-value of 0.0007. DI scores of 0.58 $\pm$ 0.08 , 0.61 $\pm$ 0.07 and 0.68 $\pm$ 0.09 corresponded to grade 1, 2 and 3 respectively as illustrated by the box-plot in figure 5.12.[14]

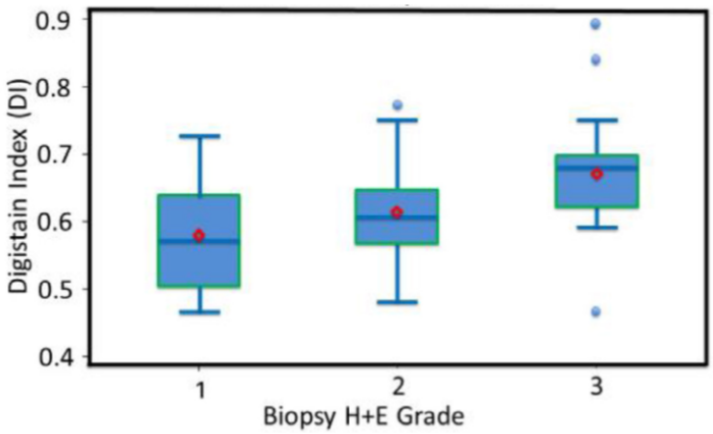



*Figure. 5.12:* A box plot illustrating the relationship between Digistain Indices (DI) and tumor grades (Grade 1, Grade 2, and Grade 3) for a sample of 71 patients. The plot shows a significant correlation between DI and tumor grade, with higher DI values associated with higher grades. Statistical analysis reveals a significant correlation (p-value = 0.0007), highlighting the potential of DI as a diagnostic marker for tumor progression. (Reproduced from[14]. No permission required. See Appendix A.)

### Technical Challenges

The Digistain spectrometer experienced non-uniform illumination in addition to issues with detector sensitivity. These combined factors resulted in a low Signal-to-Noise Ratio (SNR), which in turn led to a lack of reproducibility and compromised the accuracy and reliability of the measurements.

### Hardware Modifications

The components were reconfigured into a 1D arrangement as illustrated in figure 5.13. This reorganization increased the power density at the sample plane by at least fivefold. A chopper wheel was introduced as a replacement for the shutter, allowing for faster measurements, improved SNR, and enhanced image quality. The chopper wheel was positioned immediately after the light source to prevent overheating of the other components. Its output was used to trigger the camera, with both its frequency and time delay optimized for improved performance. Bespoke circuits were designed and built specifically to implement these modifications.

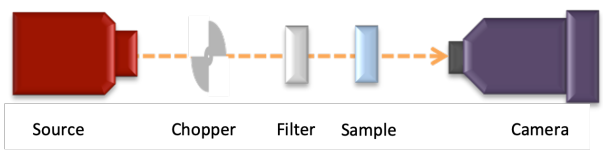



*Figure. 5.13:* The 1D Digistain prototype, consisting of an infrared source, a chopper wheel, a filter wheel, a sample slide holder, and an infrared camera.

#### Software Modifications

Various sequencing methods for capturing the 16 images were tested to determine the most effective configuration. Camera defects, such as bad pixels and pixel value fluctuations, were corrected. A MATLAB code was developed to implement these modifications.

#### Abandonment

Despite these modifications, the Digistain spectrometer continued to face significant technical challenges that made it difficult to use and almost impossible to replicate. This limitation posed a major hindrance to its potential use in routine pathology workflows. Furthermore, while the ability to compare Digistain images to H+E-stained slides was convenient, it did not offer a sufficient additional benefit to justify the complexity of the system.

### 5.6.2 Digistain as an FTIR-Based Setup

#### Setup

An alternative method for determining the DI involved the use of a benchtop FTIR spectrometer. This approach allowed for the simultaneous capture of absorption at all frequencies, eliminating the need for filters and multiple images. Although it did not enable the construction of spatially resolved images, it offered the advantages of being significantly faster and more replicable.

#### Optimization of ROI Size

The spectra from the FTIR represented single-channel values rather than pixel-by-pixel data. This limitation meant that the location and size of the ROI could not be chosen later, requiring them to be determined before capturing the spectra.

Larger ROIs helped maximize SNR, but for some samples, selecting a larger ROI was not feasible due to significant variations in tissue distribution, which could lead to capturing empty areas of the slide. The ROI size was chosen to strike a balance between maximizing SNR and addressing the practical limitations imposed by the sample characteristics. Based on these considerations, an ROI size of 400 $\times$ 400 µm was selected.

#### Measurements

Two spectra were captured: one from the sample and one from the background.

## 5.7 Strengths of Digistain

Digistain provides an objective numerical value that is inversely proportional to NCR, making it a valuable tool for both diagnostic and prognostic purposes.

Like other spectroscopic techniques, Digistain is non-phototoxic, utilizing low radiation intensities and photon energies that pose no harm to samples, patients, or operators. Additionally, it is label-free, relying solely on the absorption by phosphate and amide moieties, eliminating the need for dyes or additives.

Compared to GEP tests, Digistain does not require expensive reagents and features a relatively simple workflow, removing the need for central laboratories and offering the potential for decentralized use. The raw cost of a Digistain test is significantly lower, not exceeding £50 per test, in contrast to GEP tests, which typically cost up to £800.

Lastly, Digistain is user-independent, as any technician following the protocol can consistently produce the same results, eliminating subjectivity. Furthermore, its procedures integrate seamlessly with existing diagnostic protocols, utilizing sample slides prepared in the standard manner without the need for additional tests or biopsies. This enhances its practicality and accessibility in clinical and research settings.[47] [14]

## 5.8 Weaknesses of Digistain

Like most NCR techniques, Digistain relies on thin tissue sections, which offer two-dimensional representations of the sample. While this approach is effective, it may not fully capture the three-dimensional structure of the cells. Additionally, the selection of the ROI is performed by the histopathologist, which can introduce some variability depending on the area chosen for analysis. However, recent studies have empirically shown that this variability is minimal.

## 5.9 Conclusion

Digistain, a mid-infrared absorption spectroscopic technique, operates by measuring the ratio of phosphate absorption, which reflects nuclear content, to amide absorption, which reflects cytoplasmic content. This ratio is defined as DI. It is inversely proportional to NCR, a well-established indicator of malignancy. The Digistain technique is objective, non-phototoxic, label-free, user-independent, and cost-effective. Given these strengths, developing a prognostic model based on Digistain is highly attractive and holds significant potential for clinical application.

# Chapter 6

# The Digistain Index: Calculation and Validation

The basic definition of DI, as outlined in section , does not change. However, the method of extraction varies significantly depending on the setup used.

Digistain transitioned from a standalone spectrometer, which produced spatial maps of DI by capturing 16 images, to an FTIR-based setup that generates DI within a selected ROI using only two spectra as discussed in section 5.6.

This thesis focuses exclusively on the FTIR-based setup. This chapter addresses the first step of creating a prognostic model from the Digistain technique: converting FTIR measurements to DI. It starts by presenting the computational code written to calculate DI. Then, it shifts to the validation of DI, investigating potential errors that may arise during the computational process and describing the methods used to test for them.

## 6.1 Digistain Index Calculation

A MATLAB code was written to calculate DI. It comprised of five major steps. First, the two peaks of interest were defined. Second, baseline was subtracted to correct for baseline fluctuations. Third, peak areas were normalized to correct for contributions of interfering moieties. Fourth, areas were integrated to obtain a final measure for each peak of interest. Finally, the ratio of the resulting areas was calculated to obtain DI. The rationale and methodology behind these different segments of the code are presented in this section.

### 6.1.1 Spectral Peaks Characterisation

#### Rationale

Spectral peaks can be characterised by either *height*, which involves a single absorbance measurement, or *area*, which involves the integration or summation of absorbance measurements within a specified interval, as illustrated in figure 6.1.

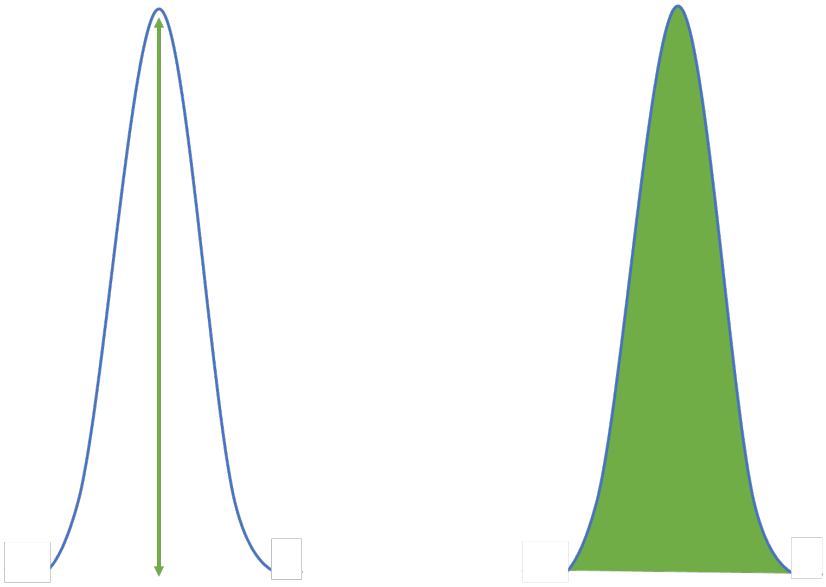



*Figure. 6.1:* Diagrams showing spectral peaks characterised by: (a) height and (b) size.

Area measurements are less likely to be affected by occasional errors caused by instrument malfunctions or noise spikes as they are cumulative measurements rather than singular measurements. This makes them more robust than height measurements.[53]

Additionally, for the same peak, the area value is always larger than the height value with the noise being almost at an equal level throughout the measurement interval. This causes area measurements to boast larger Signal-to-Noise Ratios (SNR) in comparison to height measurements.[53]

The increased robustness and higher SNR makes area measurements more preferable in sensitive applications in general and clinical applications in particular.

#### Implementation

The absorbance values needed to find the DI were defined as areas, with the amide peak ranging from 1587 $cm^{-1}$ to 1705 $cm^{-1}$, and the phosphate peak ranging from 1183 $cm^{-1}$ to 1273 $cm^{-1}$. These ranges were initially selected to align with the bandpass filters used in the free standing spectrometer but were subsequently optimized to enhance the extraction and utilisation of the molecular components of interest.

### 6.1.2 Baseline Correction

#### Rationale

In spectral measurements, the baseline represents the intensity of the infrared radiation absorbed when no sample is present. Ideally, this baseline should remain constant throughout the measurement process. However, various factors such as instrument drift or environmental conditions can introduce fluctuations in the baseline. These fluctuations often obscure the true signal of interest, leading to inaccuracies in data interpretation and analysis.

To mitigate this issue, a common approach is to apply what is known as a *baseline correction* before proceeding with the analysis. The primary goal of the correction is to retain the prominent peaks in the spectrum while flattening out the rest of the spectrum.[54] This ensures that the measured absorbance accurately reflects the molecular vibrations present in the sample.

As observed in figure 6.2, the baseline can have a a linear, curved, sloped, or irregular shape, and it is necessary to choose a suitable correction to effectively minimize the influence of baseline fluctuations.

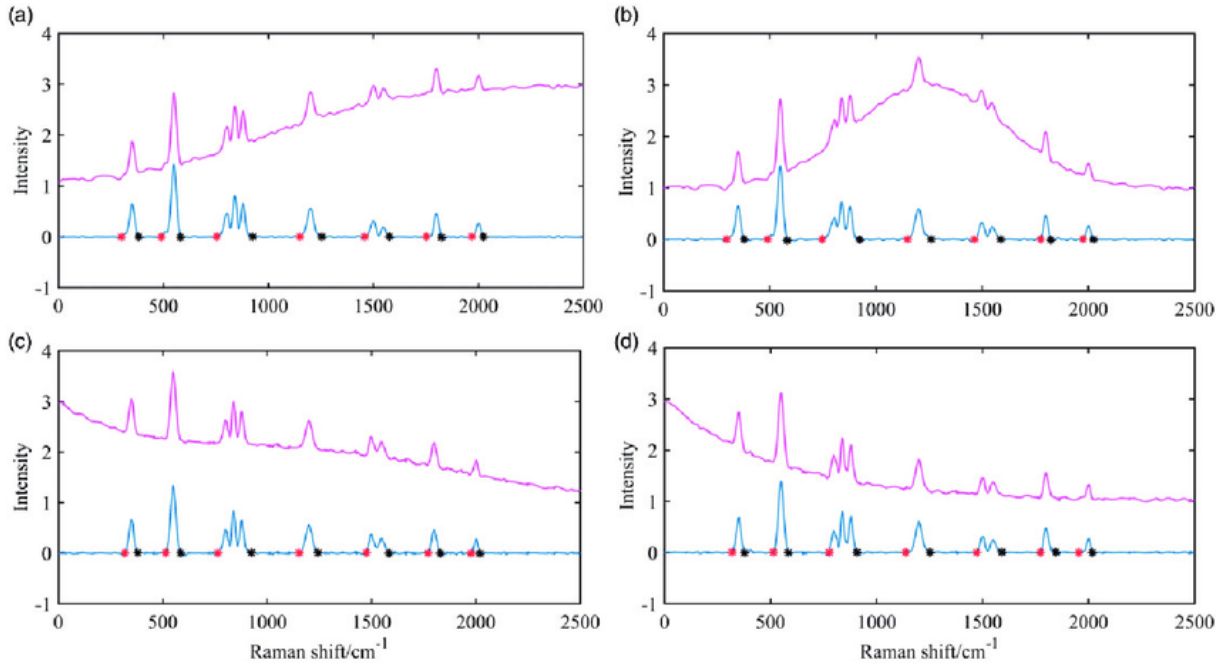



*Figure. 6.2:* Absorbance spectra affected by different baseline behaviors (pink), and then baseline-corrected spectra retaining the peaks of interest only (blue). (Reproduced from[55]. No permission required. See Appendix A.)

### Implementation

For each peak, the endpoint with the lowest absorbance was used as reference. This was equivalent to the right endpoint, with wavenumber 1705 $cm^{-1}$, for the amide peak and the left endpoint, with wavenumber 1183 $cm^{-1}$, for the phosphate peak. Then, a horizontal line was drawn across the reference for each peak, and the area below the line was subtracted as illustrated in figure 6.3. This facilitated an accurate and reliable analysis of the spectral data. By adjusting the baseline to bring the amide and phosphate peaks to the same level within each sample, accurate DI calculation was ensured. Additionally, by effectively aligning the baselines across all samples, comparisons between DI values of different samples were permitted.

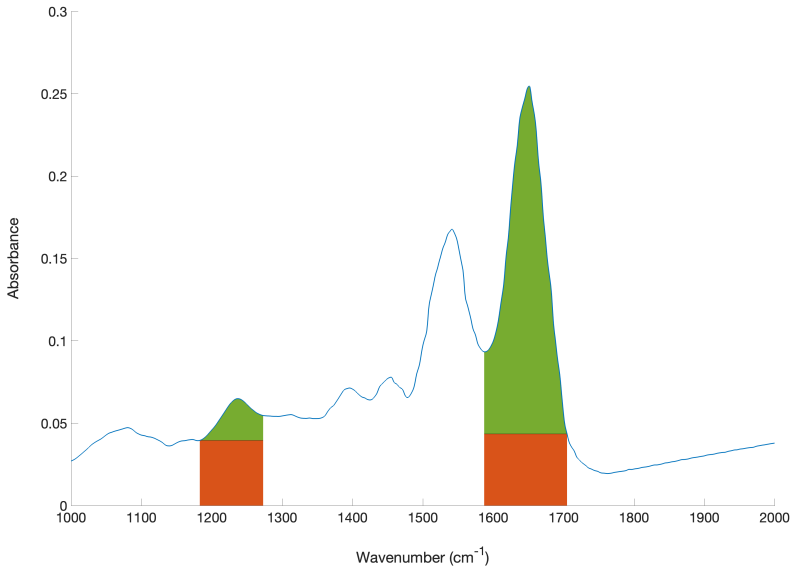



*Figure. 6.3:* Absorbance spectrum of a breast tissue sample showing areas under the amide peak (1183 $cm^{-1}$ to 1273 $cm^{-1}$) and the phosphate peak (1587 $cm^{-1}$ to 1705 $cm^{-1}$), and highlighting the areas subtracted (red) to correct for baseline fluctuations.

### 6.1.3 Peak-to-base Normalization

#### Rationale

In spectral measurements of biological samples, in particular, it is common for the same wavelengths to be absorbed by different moieties. Consequently, the resulting absorbance spectrum represents the combined influence of these moieties, rather than a single one of them. This compromises the integrity of the analysis, leading to inaccurate conclusions and misleading interpretations. Hence, it is imperative to thoroughly consider and appropriately account for these potential contributions during data analysis.

#### Implementation

The area under the curve for each absorbance peak was divided into two components: the peak and the base. The peak encompassed the region above a diagonal line connecting the two endpoints, while the base corresponded to the triangular area below this line as visually depicted in figure 6.4. The peak area, base area, and their ratio were calculated for the amide and the phosphate. By using the peak-to-base areas, effectively representing normalized absorbance, the potential influence of other moieties was corrected for.

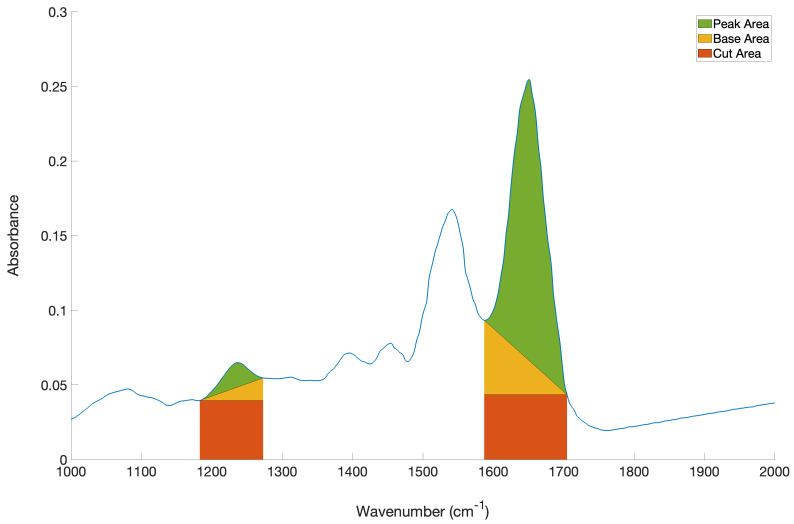



*Figure. 6.4:* Absorbance spectrum of a breast tissue sample showing areas under the amide peak (1183 $cm^{-1}$ to 1273 $cm^{-1}$) and the phosphate peak (1587 $cm^{-1}$ to 1705 $cm^{-1}$), and highlighting peak areas (green) and base areas (yellow) used to normalise absorbance peaks in addition to subtracted areas (red) used to correct for baseline fluctuations.

### 6.1.4 Area Integration

#### Rationale

Numerical integration methods state that the integral of a function, $f(x)$, from $a$ to $b$ is equal to the area under the curve within this interval and can be approximated as the summation of weighed function values. The weighing factor, $w_i$, is used to weigh the contribution of $f(x_i)$ to the sum. Depending on the integration method and application, this factor could be set as a constant for all values within the interval of integration or made to adapt to each value separately as given by

$$\int_a^b f(x)dx \approx \sum_i w_i f(x_i). \tag{6.1}$$

The interval between $a$ and $b$ can be divided into a number of steps of equal size. A smaller step size results in a more precise calculation but requires more computational resources. Conversely, a larger step size reduces computational burden but may sacrifice accuracy. The choice of step size must balance these considerations.

**Implementation**

The amide peak area, amide base area, phosphate peak area and phosphate base area were calculated using numerical integration, with a constant step size of 0.305 $cm^{-1}$. The step size was chosen to match previous versions of the DI algorithm, and was empirically validated, resulting in

$$AUC = \int_a^b A(k)dk \approx 0.305 \sum_i A(k_i). \tag{6.2}$$

### 6.1.5 Digistain Index Formulation

The DI, defined in section 6, was reformulated as

$$DI = \frac{AUC_{AP}/AUC_{AB}}{AUC_{PP}/AUC_{PB}} \tag{6.3}$$

where,

$AUC_{AP}$: Area of Amide Peak

$AUC_{AB}$: Area of Amide Base

$AUC_{PP}$: Area of Phosphate Peak

$AUC_{PB}$: Area of Phosphate Base

## 6.2 Digistain Index Validation

The validation of DI, from a computational perspective, was performed by testing the segments of the code related to area integration and DI calculation, and by checking for potential artifacts at various computational steps. These quality control tests are straightforward but essential to ensure accuracy and reliability. The specific methods and results of these validation steps are discussed in this section.

### 6.2.1 Testing Area Integration Segment

The code used numerical integration with a constant step size of 0.305 $cm^{-1}$ to calculate the amide peak, amide base, phosphate peak and phosphate base areas. To test the used methods and step size, we calculated the amide base and phosphate base areas, illustrated in figure 6.5, using the formula

$$Area = \frac{hb}{2} \tag{6.4}$$

where,

$h$: height of triangle

$b$: base of triangle

Then, we compared the results (i.e. expected areas) with the ones calculated by our code (i.e. computed areas).

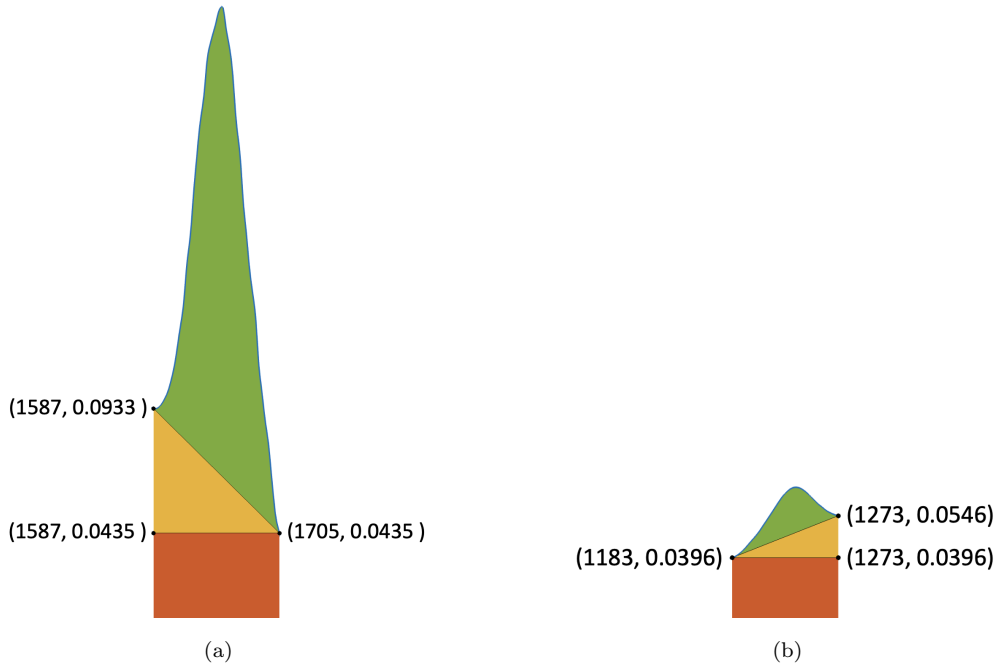



*Figure. 6.5:* Base areas used in testing the area integration segment of the code: (a) Amide peak. (b) Phosphate peak.

The results, summarised in table 6.1, demonstrated perfect agreement between the expected and computed areas, reinforcing the reliability and accuracy of the computational methods and parameters used in our algorithm.

*Table 6.1:* Expected areas and computed areas show perfect agreement for both amide base and phosphate base.

| Parameter | $\text{Area}_{expected}$ | $\text{Area}_{computed}$ | Agreement |
|---|---|---|---|
| Amide Base | 2.94 | 2.94 | Yes |
| Phosphate Base | 0.68 | 0.68 | Yes |

### 6.2.2 Testing Digistain Index Calculation Segment

The code used areas of amide peak, amide base, phosphate peak and phosphate base to calculate DI. To test the computational steps, we calculated DI by plugging the areas in equation 6.3 (i.e. expected DI), and then compared it to the DI calculated by our code (i.e. computed DI).

The results, outlined in table 6.2, demonstrated perfect agreement between the expected and computed DI, confirming the integrity of our computational approach.

*Table 6.2:* Expected DI and computed DI show perfect agreement.

| Parameter | Peak | Base | Ratio | $DI_{expected}$ | $DI_{computed}$ | Agreement |
|---|---|---|---|---|---|---|
| Amide | 10.4293 | 2.9426 | 3.3544 | 3.36 | 3.36 | Yes |
| Parameter | 0.7180 | 0.6813 | 1.0534 | | | |

### 6.2.3 Testing Overall Code Artifacts and Errors

The code read the spectrum data, applied the necessary corrections, calculated the area under the curve using numerical methods, and then calculated DI. A critical concern was to ensure that none of these steps introduced extraneous artifacts or computational errors that could undermine the spectral information and compromise the validity of any subsequent analysis. To test this, we inputted a series of artifical test spectra into the code, ranging from completely flat to step-like and partially flat, as illustrated in figure 6.6, and closely observed the outputs to discern any artifacts or deviations from the expected results.

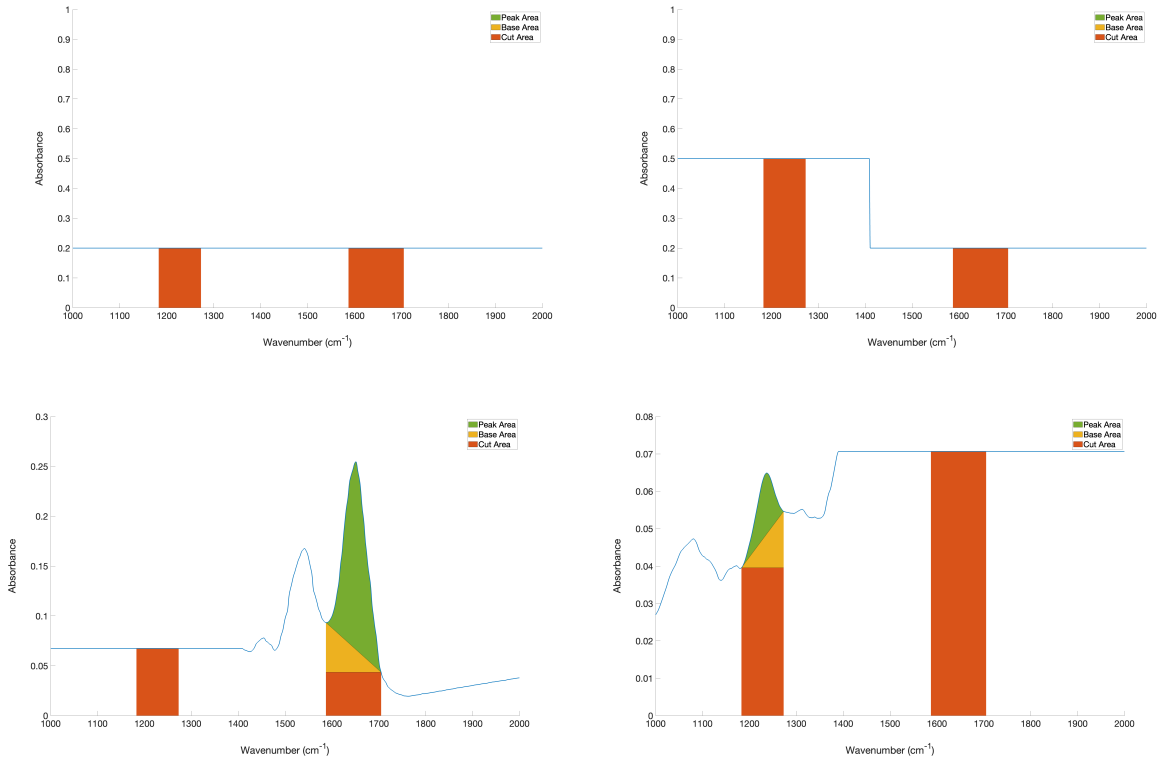



*Figure. 6.6:* Artificial absorbance spectra used in testing the overall robustness of the code.

All outputs were as expected with Not a Number (NaN) encountered at the first step where division by zero occured, leading to a DI value of NaN. This confirms that the code operated on absorbance spectra as expected and did not introduce any artifacts throughout its execution.

## 6.3 Conclusion

A computational code was developed to read FTIR measurements, correct for baseline fluctuations, account for potential contributions of interfering moieties, and calculate DI. Then, the code underwent validation through additional computational tests to ensure its accuracy and reliability.

# Chapter 7

# The Digistain Prognostic Score: Development and Validation

This chapter focuses on the second and last step of creating a prognostic model from the Digistain technique: converting DI to Digistain Prognostic Score (DPS). Initially, it outlines the study cohort. Then, it delves into the development of DPS by combining tumor grade, lymph node status, and DI. Finally, it focuses on the validation of DPS by evaluating its prognostic ability and stratification power.

## 7.1 Study Cohort

Information on samples, patients, and resulting DI values is presented in this section.

### 7.1.1 Sourcing, Preparation and Measurement of Samples

Tissue blocks of breast cancer patients were randomly selected from the records at Nottingham City Hospital where the patients received treatment between 1998 and 2006. These were used to prepare Tissue Microarray (TMA) blocks, which are blocks consisting of tissue samples from ROIs of different patients. Each TMA block contained 50 tissue cores, each with a diameter of 1 mm. A section of 4 $\mu m$ thickness was cut from each TMA block and mounted on an infrared transmitting calcium fluoride microscope slide. Then, the slides were deparaffinized and the spectra of all samples were measured on Bruker Vertex 70 Infrared Spectrometer by MSc students. It is worth noting that TMAs are commonly used in research as they reduce the time required for de-paraffinization, as well as ROI selection and localization. Additionally, the protocol for selecting ROIs

from full-section slides to construct a TMA block is well established and approved by pathologists. Therefore, using TMAs instead of full-section slides is not expected to affect the results or conclusions of this study.

### 7.1.2 Clinicopathological Characteristics of Patients

The study population consisted of 1184 patients. The ages of patients ranged from 23 to 71 years, with a mean of 54 and a median of 55. Only 15% of the patients had grade 1 cancer, while 40% and 45% had tumours graded 2 and 3, respectively. Approximately 60% were LN-negative, 30% had one to three positive lymph nodes while a smaller proportion of 10% had more than three positive lymph nodes. The sizes of tumours ranged from 0.2 cm to 7.5 cm, with a mean of 1.9 and a median of 1.7.

### 7.1.3 Calculation and Distribution of Digistain Indices

DI values were extracted from the spectra of tissue samples using the MATLAB code discussed in chapter 6. The values ranged from 1.1 to 11.0, with a mean of 2.5, a median of 2.4 and a standard deviation of 0.8. The proximity of the mean and the median, coupled with the minor skewness in the distribution as depicted in figure 7.1, makes it reasonable to regard DI as a normally distributed variable.

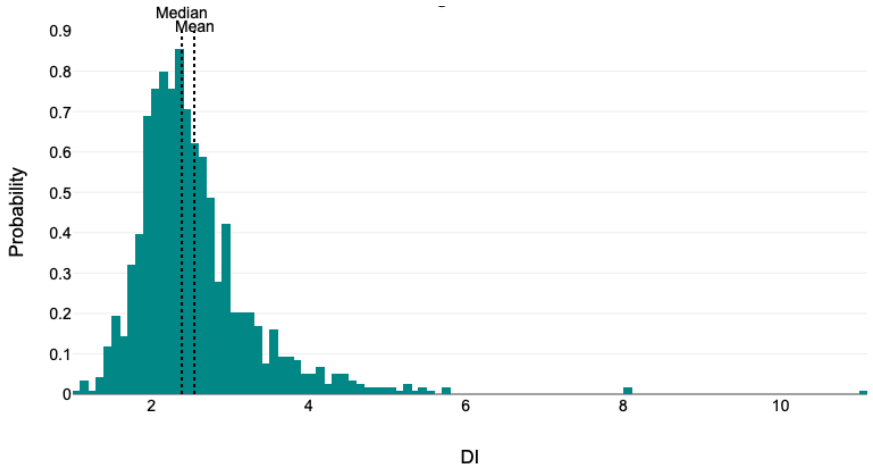



*Figure. 7.1:* Histogram of Digistain Index (DI) values shows minor skewness from normal distribution, with the mean nearly equivalent to the median.

## 7.2 Digistain Prognostic Score Development

The development of DPS involved evaluating the prognostic ability of several factors. This evaluation was conducted in two steps: first, by testing the prognostic ability of each factor univariately, and then by testing

the prognostic ability of all factors together.

The first step resulted in age scoring a hazard ratio of approximately 1 as a predictor of OS and RFS at 10 years, which suggests it does not add any significant value to our model and was therefore excluded.

In the second step, when all factors were considered together, the p-value associated with tumour size was found to be approximately 0.1 for OS and RFS at 10 years. This p-value is twice the threshold for statistical significance, indicating a weaker association than would be considered statistically significant. Hence, the size of the tumour was also excluded from the covariates list.

The exclusion of age and tumour size should not have any impact on the results due to their statistical insignificance, as confirmed by Cox regression. This leaves grade, lymph node status, and DI. Their details as individual predictors and in combination are presented in this section.

### 7.2.1 Impact of Grade, Lymph Node Status and DI on Prognosis

The ability of grade, lymph node status and DI to, univariately, predict OS and RFS at 10Y and 5Y was assessed using Cox regression, and the goodness-of-fit of the resulting models was verified using log-likelihood ratio tests. The results are summarised in tables 7.1 , 7.2 and 7.3.

*Table 7.1:* Cox model for evaluating the ability of grade to predict overall survival and recurrence-free survival at 10 and 5 years.

| | Overall Survival | | Recurrence Free Survival | |
|---|---|---|---|---|
| | 10Y | 5Y | 10Y | 5Y |
| Grade | HR: 2.114 | HR: 3.170 | HR: 1.831 | HR: 2.459 |
| | CI: 1.717 - 2.630 | CI: 2.278 - 4.556 | CI: 1.542 - 2.188 | CI: 1.932 - 3.178 |
| | P: <0.0001 | P: <0.0001 | P: <0.0001 | P: <0.0001 |
| Log-Likelihood Test | P: <0.0001 | P: <0.0001 | P: <0.0001 | P: <0.0001 |

*Table 7.2:* Cox model for evaluating the ability of lymph node status to predict overall survival and recurrence-free survival at 10 and 5 years.

| | Overall Survival | | Recurrence Free Survival | |
|---|---|---|---|---|
| | 10Y | 5Y | 10Y | 5Y |
| Lymph Node Status | HR: 2.124 | HR: 2.525 | HR: 1.934 | HR: 2.027 |
| | CI: 1.799 - 2.501 | CI: 2.022 - 3.149 | CI: 1.669 - 2.235 | CI: 1.686 - 2.429 |
| | P: <0.0001 | P: <0.0001 | P: <0.0001 | P: <0.0001 |
| Log-Likelihood Test | P: <0.0001 | P: <0.0001 | P: <0.0001 | P: <0.0001 |

*Table 7.3:* Cox model for evaluating the ability of DI to predict overall survival and recurrence-free survival at 10 and 5 years.

| | Overall Survival | | Recurrence Free Survival | |
|---|---|---|---|---|
| | 10Y | 5Y | 10Y | 5Y |
| DI | HR: 0.7977 | HR: 0.7307 | HR: 0.8364 | HR: 0.8004 |
| | CI: 0.6535 - 0.9587 | CI: 0.5451 - 0.9535 | CI: 0.7084 - 0.9756 | CI: 0.6413 - 0.9804 |
| | P: 0.0211 | P: 0.0285 | P: 0.0290 | P: 0.0404 |
| Log-Likelihood Test | P: 0.0147 | P: 0.0193 | P: 0.0219 | P: 0.0305 |

The results reaffirmed the well-established prognostic significance of grade and lymph node status, with p-values $\leq 0.0001$ for all models. The hazard ratios demonstrated that an increase in these variables corresponds to heightened risk, with a more pronounced association observed with death compared to recurrence, and at 5 years compared to 10 years.

The increase from grade 1 to grade 2 or from grade 2 to grade 3, was shown to double the risk of death within 10 years and triple it within 5 years. A similar, but less pronounced effect, was observed on the risk of recurrence.

Furthermore, the lymph node status was classified as follows: 1 for 0 positive lymph nodes, 2 for 0-3 positive lymph nodes, and 3 for more. Advancing from one status to the next was shown to increase the risk of death by 1 fold at 10 years and by 1.5 fold at 5 years. Again, a similar effect was observed on the risk of recurrence.

Most importantly, the results conclusively demonstrated that DI possesses substantial prognostic power

as a univariate predictor of clinical outcome, with p-values $\leq 0.05$ for all models. The hazard ratios demonstrated an inverse relationship between DI and risk, as the hazard ratios were all smaller than 1. This inverse relationship between DI and risk was expected because DI was defined as the ratio of phosphate moieties to amide moieties, which makes it inversely proportional to NCR and accordingly to cancer risk. For each unit increase in DI, the risk of death reduces by 20% at 10 years and almost 30% at 5 years. The association with recurrence was slightly less prominent with risk reduction of 17% and 20% respectively.

In summary, the results reinforced the roles of grade and lymph node status as key factors in breast cancer prognosis, while also paving the way for the inclusion of another factor, namely DI. The next step was to asses how well these factors perform when used together and determine if each of them maintains its predictive power and remains clinically relevant in guiding treatment decisions for breast cancer patients.

### 7.2.2 Combined Impact of Grade, Lymph Node Status and DI on Prognosis

The combined use of grade, lymph node status and DI to predict OS and RFS at 10Y and 5Y was assessed using Cox regression. The resulting hazard ratios and p-values are summarized in table 7.4, and the corresponding forest plots are shown in figure 7.2.

*Table 7.4:* Cox model for evaluating the ability of grade, lymph node status and DI to predict overall survival and recurrence-free survival at 10 and 5 years.

| | Overall Survival | | Recurrence Free Survival | |
|---|---|---|---|---|
| | 10Y | 5Y | 10Y | 5Y |
| Grade | HR: 1.924<br>CI: 1.553 - 2.406<br>P: <0.0001 | HR: 2.821<br>CI: 2.0080 - 4.0870<br>P: <0.0001 | HR: 1.675<br>CI: 1.404 - 2.011<br>P: <0.0001 | HR: 2.246<br>CI: 1.752 - 2.919<br>P: <0.0001 |
| Stage | HR: 1.888<br>CI: 1.598 - 2.227<br>P: <0.0001 | HR: 2.142<br>CI: 1.714 - 2.674<br>P: <0.0001 | HR: 1.753<br>CI: 1.512 - 2.029<br>P: <0.0001 | HR: 1.766<br>CI: 1.468 - 2.128<br>P: <0.0001 |
| DI | HR: 0.8074<br>CI: 0.6584 - 0.9775<br>P: 0.0341 | HR: 0.7470<br>CI: 0.5491 - 0.9770<br>P: 0.0413 | HR: 0.8466<br>CI: 0.7141 - 0.9933<br>P: 0.0482 | HR: 0.8019<br>CI: 0.6387 - 0.9918<br>P: 0.0495 |
| Log-Likelihood Test | P: <0.0001 | P: <0.0001 | P: <0.0001 | P: <0.0001 |

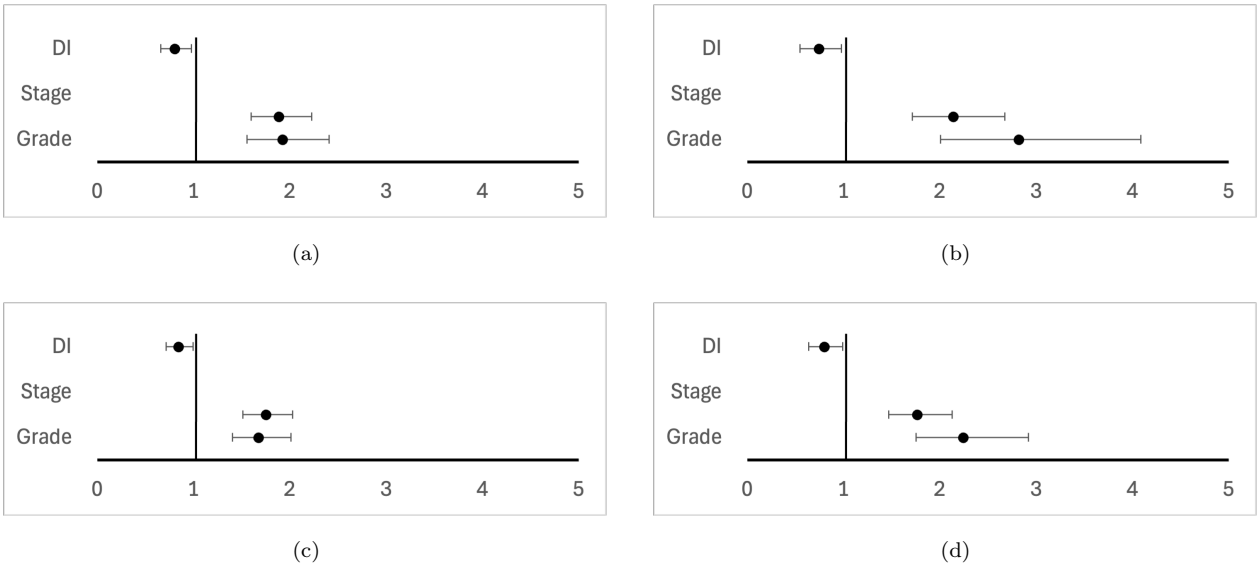



*Figure. 7.2:* Forest plots of multivariate analysis, incorporating grade, stage, and DI as covariates for: (a) overall survival at 10 years. (b) overall survival at 5 years. (c) recurrence-free survival at 10 years. (d) recurrence-free survival at 5 years. Data from the analysis highlight the independent prognostic contributions of each variable, with hazard ratios and 95% confidence intervals displayed for each factor.

The log-likelihood ratio test confirmed the global statistical significance of the resulting Cox models with grade, lymph node status and DI as predictors of OS and RFS at 10 and 5 years, as it resulted in p-values $\leq 0.0001$ for all models. Additionally, the hazard ratios and 95% confidence intervals presented in the table and illustrated in the forest plots highlighted the independent prognostic contribution of each variable, emphasizing the significance of grade, lymph node status, and DI in predicting patient outcomes at different time points. This pivotal finding is especially important for DI, as it proved its independent prognostic value and validated its use as a reliable prognostic factor. In conclusion, the results validated the combined use of grade, lymph node status, and DI to predict clinical outcomes. The next step involves effectively integrating these factors into a prognostic model.

### 7.2.3 Formulation of Digistain Prognostic Score Equation

We constructed the DPS using the parameters from the Cox model consisting of grade, lymph node status and DI as predictors of OS at 10Y. We defined the $DPS$ as the square root of the product of Partial Hazards ($PH$) as given by

$$DPS = \sqrt{\prod_p PH_p}. \tag{7.1}$$

The $PH$, for a given predictor $p$, is calculated as the $HR$ to the power of the normalized predictor value

$$PH_p = HR_p^{\frac{x_p - \mu_p}{\sigma_p}} \tag{7.2}$$

where,

$x_p$: predictor value

$\mu_p$: mean of predictor values across fitted Cox model population

$\sigma_p$: standard deviation of predictor values across fitted Cox model population

## 7.3 Digistain Prognostic Score Validation

The validation of DPS involved assessing its risk prediction ability using Cox regression, asserting its overall performance using ROC curves, assessing its stratification power using Kaplan-Meier curves and finally summarising its negative predictive values.

### 7.3.1 Calculation and Distribution of Digistain Prognostic Scores

DPS values were calculated using equations 7.1 and 7.2, with hazard ratios for DPS as a predictor of OS at 10Y from table 7.4. The values ranged from 0.1 to 3.2, with a mean and a median of 1.1, and a standard deviation of 0.6.

### 7.3.2 Cox Regression: Assessing Risk Prediction

The ability of DPS to predict OS and RFS at 10Y and 5Y was assessed using Cox regression, and the goodness-of-fit of the resulting models was verified using log-likelihood ratio tests. The results are summarised in table 7.5.

*Table 7.5:* Cox model for evaluating the ability of DPS to predict overall survival and recurrence-free survival at 10 and 5 years.

| | Overall Survival | | Recurrence Free Survival | |
|---|---|---|---|---|
| | 10Y | 5Y | 10Y | 5Y |
| DPS | HR: 2.424<br>CI: 2.077 - 2.817<br>P: <0.0001 | HR: 2.933<br>CI: 2.401 - 3.566<br>P: <0.0001 | HR: 2.183<br>CI: 1.899 - 2.498<br>P: <0.0001 | HR: 2.393<br>CI: 2.019 - 2.820<br>P: <0.0001 |

The Cox models revealed statistically significant correlations between DPS and risk, with p-values $\leq$ 0.0001 for all models. They also demonstrated strong hazard ratios. This suggests that patients with higher DPS scores are significantly more likely to experience adverse outcomes. For each unit increase in DPS, the risk of death escalates substantially, reaching 250% at 10 years and nearly 300% at 5 years. The risk of recurrence also increases, albeit by a slightly lower percentage. This evident correlation proves the predictive power of DPS, and substantiates its use to guide treatment decisions.

### 7.3.3 ROC: Assessing Overall Performance

The overall performance of DPS as a predictor of OS and RFS at 10Y and 5Y was evaluated using ROC curves and AUC values as shown in figure 7.3.

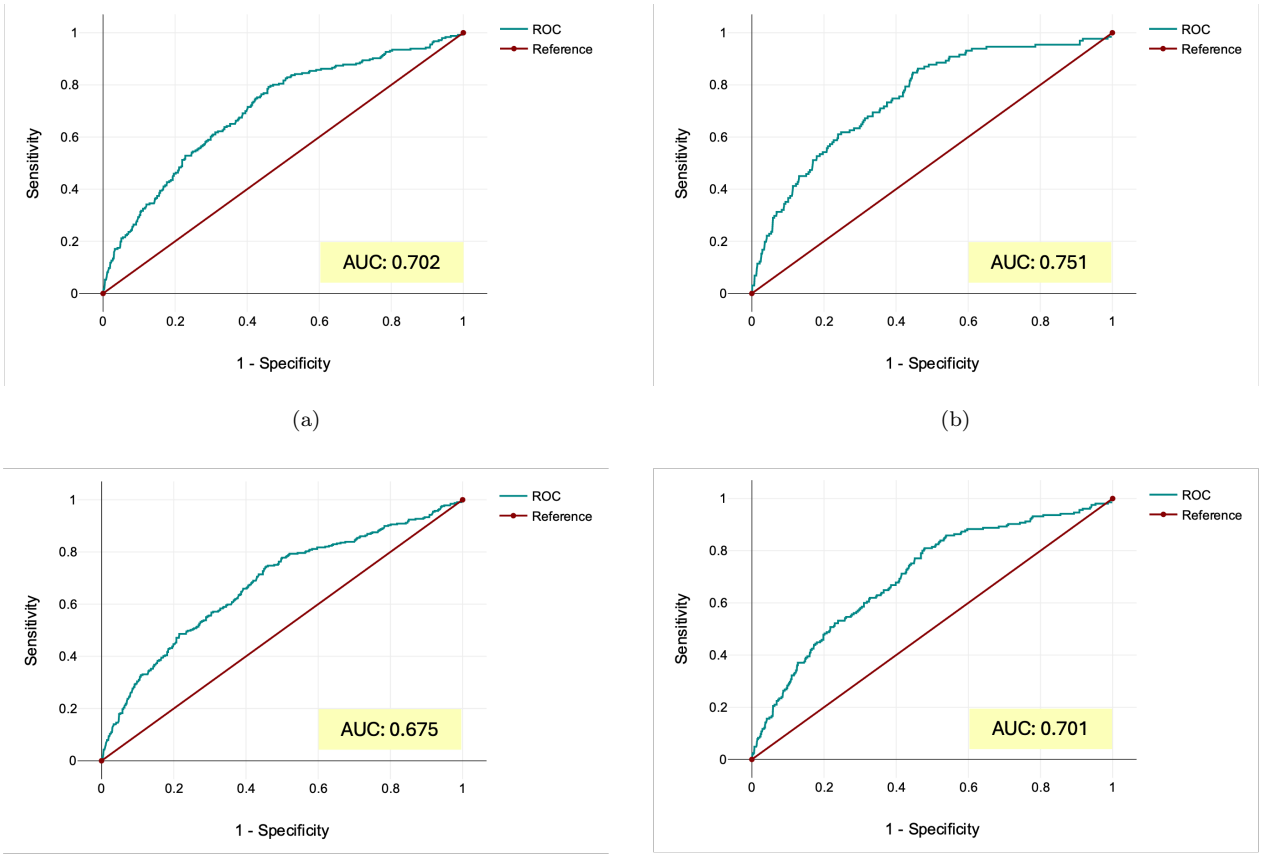


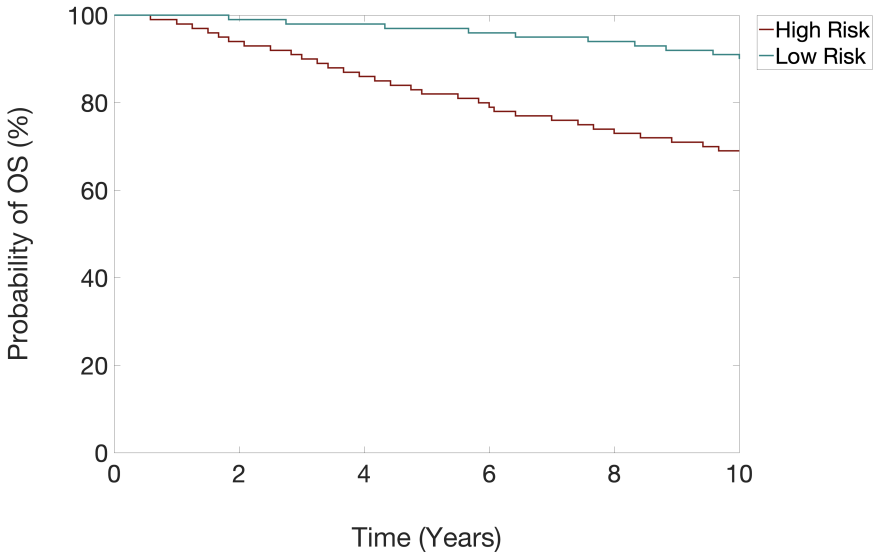

*Figure. 7.3:* ROC curves to test prognostic ability of DPS result in consistently high AUC values for: (a) overall survival at 10 years. (b) overall survival at 5 years. (c) recurrence-free survival at 10 years. (d) recurrence-free survival at 5 years.

The AUC values were approximately 0.7 for all ROC curves, with the ROC for OS at 5Y nearly reaching 0.8. These consistently high values further emphasize the predictive power of DPS, and affirm its reliability and robustness as clinical decision-making tool.

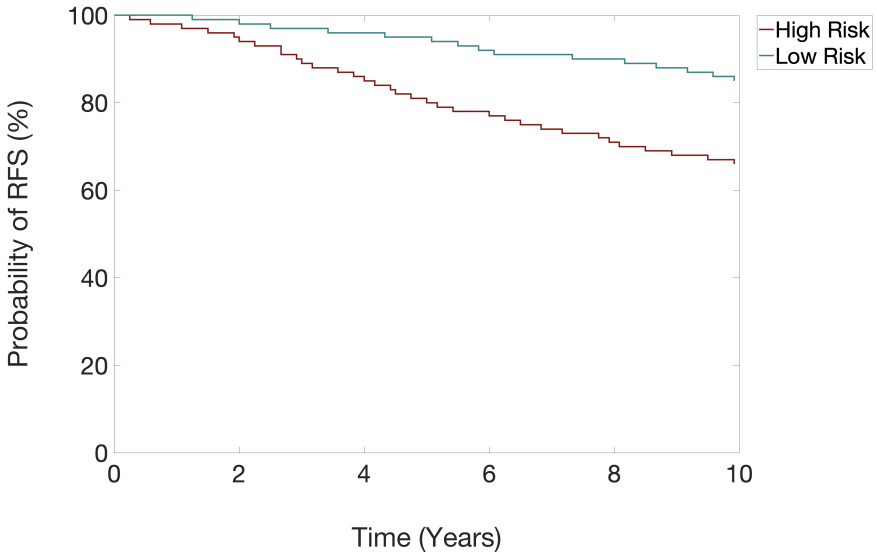
### 7.3.4 Kaplan Muier: Assessing Risk Stratification

The patients in the total population were classified into high and low risk using a DPS cut-off of 1, found by maximising the Youden Index in the ROC curves. The number of patients who fell into the high-risk category were 661. At 10 years, 70% of these patients were still alive while 63% were alive and free of recurrence. Conversely, the number of patients who fell into the low-risk category were 528. At 10 years, 91% of these patients were still alive while 85% were alive and free of recurrence. As expected, and wished for, the percentages are considerably higher for the high-risk group than for the low-risk group.

The Kaplan-Meier curves were plotted to visualize the difference in event distribution over time between the high-risk group and the low risk-group for OS and RFS, as given in 7.4.




(a)




(b)

*Figure. 7.4:* Kaplan-Meier curves to test stratification power of DPS show significant difference between high and low risk groups in: (a) overall survival (b) recurrence-free survival

Log-rank tests were calculated for 10-year overall survival and 10-year recurrence-free survival. Both tests resulted in p-values $<0.001$, indicating significant differences between the high-risk group and the low-risk group in terms of event distribution over time, and proving the stratification power of DPS.

It is important to note that the survival cutoff for categorizing high-risk patients in early breast cancer can vary, but it is a common practice to classify patients with OS of less than 90% at 5 years as high risk. For long-term outcomes, it is also common practice to deem patients with an RFS of less than 75-80% at 10 years as high risk.

The results we have align well with these established standards. Moreover, the ultimate clinical need is to identify low-risk patients in order to de-escalate chemotherapy, where overtreatment is a prevailing concern. The DPS has proven to be effective in identifying such patients, thereby enabling more tailored treatment decisions and reducing unnecessary treatments.

### 7.3.5 Negative Predictive Values

The ability of DPS to identify patients who are unlikely to experience adverse outcomes was assessed by calculating the negative predictive values for OS and RFS at 10Y and 5Y. The results are summarised in table 7.6.

*Table 7.6:* Negative Predictive Values (NPV) based on DPS high and low risk classification.

| | Overall Survival | | Recurrence Free Survival | |
|---|---|---|---|---|
| | 10Y | 5Y | 10Y | 5Y |
| NPV | 0.91 | 0.97 | 0.86 | 0.93 |

They NPV values were highest for overall survival, ranging from 0.91 at 10 years and 0.97 at 5 years, and slightly lower for recurrence-free survival, ranging from 0.86 at 10 years and 0.93 at 5 years. These values are comparable to those from leading Gene Expression Profiling (GEP) tests.[56] This proves the clinical utility and prognostic power of DPS, and makes it an attractive option for guiding treatment decisions in breast cancer patients.

## 7.4 Conclusion

A study cohort comprising over a thousand breast cancer patients was utilized for both the development and validation of DPS. Initially, the DPS was developed based on a Cox model with tumor grade, lymph node status, and DI as predictors of OS at 10Y. Subsequently, the prognostic ability of the DPS was validated using Cox regression and ROC curves. Additionally, the stratification power of the DPS was validated using

Kaplan-Meier curves and the logrank test. These findings pave the way for the clinical implementation of DPS in breast cancer management as a guide to adjuvant chemotherapy decisions.

# Part III

# Concluding Remarks and Future Work

# Chapter 8

# Concluding Remarks

Breast cancer presents as a complex disease, characterized by various sub-types and influenced by multiple prognostic markers, which is reflected in the intricacies of treatment decisions.

Adjuvant chemotherapy treatment decisions depend heavily on prognosis, often assessed through conventional prognostic scores, with the NPI serving as the gold standard. These scores rely on subjective histological factors, which has led to chemotherapy overtreatment. This underscores the urgent need for more precise prognostic tools.

In the past two decades, GEP emerged as a potential solution. However, limitations and challenges, primarily related to cost and turnaround times, hindered its widespread application.

In this thesis, our objective was to develop a prognostic score based on the Digistain technique, a mid-infrared technique that is designed to produce an index, DI, that correlates with NCR, a recognized measure of cancer. We utilised 1184 absorption spectra from a cohort of patients who did not receive adjuvant chemotherapy. The first step involved optimizing a computational code to read FTIR spectra and extract DI, while the second step focused on integrating the DI with other prognostic factors to create the DPS.

The resulting DPS showed statistically significant prognostic power and stratification capability for OS and RFS at 5Y and 10Y. The prognostic power was evidenced by the strong hazard ratios ($> 2$) and significant p-values ($\leq 0.0001$) obtained from Cox regression, in addition to the consistently high AUC values (0.7 - 0.8) from ROC curves. The stratification capability was demonstrated by the clear distinction between high- and low- risk groups represented by Kaplen-Meier curves.

The work has been published in *Breast Cancer Research and Treatment* for a sub-cohort consisting of patients with only HR-positive, HER-2 negative, primary breast cancer with up to three positive lymph nodes. This demonstrates that DPS has a great potential to address the unmet need in breast cancer prognosis and treatment.[57]

Additionally, a study by an independent body, *Health Tech Enterprise*, conducted modeling to assess the impact of using DPS. It concluded that if DPS was used in early breast cancer cases with no positive lymph nodes, instead of Oncotype DX, it would cut down cumulative cost over a period of 5 years by approximately GBP 287 million. Moreover, if DPS was used in cases with one to three positive lymph nodes instead of PREDICT, an online prognostic tool based on conventional factors, it would save 1,266 life years.[58]

# Chapter 9

# Future Work

DPS demonstrated promising clinical utility, with significant prognostic ability and stratification power. The next natural step is to establish its analytical utility. This involves assessing inter- and intra-pathologist variability, focusing on the effect of selecting the RoI, and evaluating inter- and intra-operator variability, examining the reliability of the measurement protocol. Additionally, it involves optimizing measurement sizes and measurement duration. We are already halfway there, though the results are not yet ready for publication.

Following this, we can evaluate DPS on different sub-cohorts of this total population, focusing on the effects of menopausal status and racial variations. Subsequently, we can test DPS on other sample sets, preferably full-section slides instead of TMAs.

# Bibliography


[1] Melina Arnold, Eileen Morgan, Harriet Rumgay, Allini Mafra, Deependra Singh, Mathieu Laversanne, Jerome Vignat, Julie R. Gralow, Fatima Cardoso, Sabine Siesling, and Isabelle Soerjomataram. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, 66, 2022. doi: 10.1016/j.breast.2022.08.010.

[2] Christopher Hillyar, Hirah Rizki, Omar Abbassi, Sascha Miles-Dua, Gillian Clayton, Tasha Gandamihardja, and Simon Smith. Correlation between oncotype dx, predict and the nottingham prognostic index: Implications for the management of early breast cancer. *Cureus*, 12(4):e7552, 2020. doi: 10.7759/cureus.7552.

[3] Hazem I. Assi, Ibrahim A. Alameh, Jessica Khoury, Nour Abdul Halim, Fadi El Karak, Fadi Farhat, Juliett Berro, Eman Sbaity, Maya Charafeddine, Arafat Tfayli, Ziad Salem, and Nagi El Saghir. Impact of commercialized genomic tests on adjuvant treatment decisions in early stage breast cancer patients. *Journal of Oncology*, 2020:9238084–7, 2020. doi: 10.1155/2020/9238084.

[4] E. Comen, L. Balistreri, M. Gönen, A. Dutra-Clarke, M. Fazio, J. Vijai, Z. Stadler, N. Kauff, T. Kirchhoff, C. Hudis, K. Offit, and M. Robson. Discriminatory accuracy and potential clinical utility of genomic profiling for breast cancer risk in brca-negative women. *Breast Cancer Research and Treatment*, 127(2): 479–487, 2010. doi: 10.1007/s10549-010-1215-2.

[5] P. Sinn, S. Aulmann, R. Wirtz, S. Schott, F. Marmé, Z. Varga, A. Lebeau, H. Kreipe, and A. Schneeweiss. Multigene assays for classification, prognosis, and prediction in breast cancer: a critical review on the background and clinical utility. *Geburtshilfe und Frauenheilkunde*, 73(9):932–940, 2013. doi: 10.1055/s-0033-1350831.

[6] D. Mort, M. Lansdown, N. Smith, K. Protopapa, and Mason. M. For better, for worse? a review of the care of patients who died within 30 days of receiving systemic anti-cancer therapy. Technical report, 2008.

URL https://www.ncepod.org.uk/2008report3/Downloads/SACT_report.pdf. [Accessed: 15th May 2024].

[7] Jennifer L. Ersek, Lora J. Black, Michael A. Thompson, and Edward S. Kim. Implementing precision medicine programs and clinical trials in the community-based oncology practice: Barriers and best practices. *American Society of Clinical Oncology educational book*, 38(38):188–196, May 23 2018. doi: 10.1200/EDBK_200633.

[8] Renan Gomes do Nascimento and Kaléu Mormino Otoni. Histological and molecular classification of breast cancer: what do we know? *Mastology*, 30, 2020. doi: 10.29289/25945394202020200024.

[9] Britta Weigelt, Felipe C. Geyer, and Jorge S. Reis-Filho. Histological types of breast cancer: How special are they? *Molecular Oncology*, 4(3), 2010. doi: 10.1016/j.molonc.2010.04.004.

[10] Isabelle Soerjomataram, Claudia Allemani, Adri Voogd, and Sabine Siesling. *The Global Burden of Breast Cancer in Women. In:*, pages 3–16. Breast cancer: global quality care. Oxford University Press, Oxford, 2019.

[11] Malavika Pradeep, Nivethigaa Balakrishnan, T. R. Prasanna Arvind, Piush Sharma, and K. K. Sharma. Various types and management of breast cancer: An overview. *Journal of advanced pharmaceutical technology and research*, 1(2):109–126, 2010. doi: 10.4103/japtr.japtr_158_22.

[12] Bradford Tan. Cancer cells vs. normal cells. URL https://www.cancercenter.com/cancer-cells-vs-normal-cells. [Accessed: 15th March 2024].

[13] H. Amrania, L. Drummond, R. C. Coombes, S. Shousha, L. Woodley-Barker, K. Weir, W. Hart, I. Carter, and C. C. Phillips. New ir imaging modalities for cancer detection and for intra-cell chemical mapping with a sub-diffraction mid-ir s-snom. *Faraday discussions*, 187:539–553, 2016. doi: 10.1039/c5fd00150a.

[14] H. Amrania, L. Woodley-Barker, K. Goddard, B. Rosales, S. Shousha, G. Thomas, T. McFarlane, M. Sroya, C. Wilhelm-Benartzi, K. Cocks, R. C. Coombes, and C. C. Phillips. Mid-infrared imaging in breast cancer tissue: an objective measure of grading breast cancer biopsies. *Convergent science physical oncology*, 4(2):25001, Mar 13, 2018. doi: 10.1088/2057-1739/aaabc3.

[15] Michael J. Moore, Joseph A. Sebastian, and Michael C. Kolios. Determination of cell nucleus-to-cytoplasmic ratio using imaging flow cytometry and a combined ultrasound and photoacoustic technique: a comparison study. *Journal of biomedical optics*, 24(10):106502, Oct 1 2019. doi: 10.1117/1.JBO.24.10.106502.

[16] Hannah Drescher, Sabine Weiskirchen, and Ralf Weiskirchen. Flow cytometry: A blessing and a curse. *Biomedicines*, 9(11):1613, Nov 4 2021. doi: 10.3390/biomedicines9111613.

[17] Emad A. Rakha, Jorge S. Reis-Filho, Frederick Baehner, David J. Dabbs, Thomas Decker, Vincenzo Eusebi, Stephen B. Fox, Shu Ichihara, Jocelyne Jacquemier, Sunil R. Lakhani, José Palacios, Andrea L. Richardson, Stuart J. Schnitt, Fernando C. Schmitt, Puay-Hoon Tan, Gary M. Tse, Sunil Badve, and Ian O. Ellis. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Research*, 12(4):207–207, 2010. doi: 10.1186/bcr2607.

[18] Jamie DePolo. Breast cancer grades (nottingham grade). URL `https://www.breastcancer.org/pathology-report/breast-cancer-grades#`. [Accessed: 15th March 2024].

[19] Maria Vittoria Dieci, Enrico Orvieto, Massimo Dominici, PierFranco Conte, and Valentina Guarneri. Rare breast cancer subtypes: Histological, molecular, and clinical peculiarities. *The Oncologist*, 19(8): 805–813, 2014. doi: 10.1634/theoncologist.2014-0108.

[20] Susan G. Komen. Breast cancer diagnosis. URL `https://www.komen.org/breast-cancer/diagnosis/`. [Accessed: 15th March 2024].

[21] The American Cancer Society medical and editorial content team. Lymph nodes and cancer. URL `https://www.cancer.org/cancer/diagnosis-staging/lymph-nodes-and-cancer`. [Accessed: 1st April 2024].

[22] National Cancer Institute. Cancer staging. URL `https://www.cancer.gov/about-cancer/diagnosis-staging/staging`. [Accessed: 1st April 2024].

[23] Ji Young You, Soojin Park, Eun-Gyeong Lee, and Eun Sook Lee. Detection and diagnosis of breast cancer. *A Practical Guide to Breast Cancer Treatment*, pages 1–17, 2023. doi: 10.1007/978-981-19-9044-1_1.

[24] Minh Tung Phung, Sandar Tin Tin, and J. Mark Elwood. Prognostic models for breast cancer: a systematic review. *BMC Cancer*, 19(1):230, 2019. doi: 10.1186/s12885-019-5442-6.

[25] Mary Cianfrocca and Lori J. Goldstein. Prognostic and predictive factors in early-stage breast cancer. *The Oncologist*, 9(6):606–616, 2004. doi: 10.1634/theoncologist.9-6-606.

[26] Danielle A. van der Windt, Harry Hemingway, Peter Croft, Richard D. Riley, Karel GM Moons, Thomas PA Debray, Kym IE Snell, Ewout W. Steyerberg, Douglas G. Altman, and Gary S. Collins. *Prognostic model research. In:*, pages 139–187. Prognosis Research in Healthcare. Oxford University Press, 2019.

[27] Ji Young You, Soojin Park, and Eun Sook Lee. *Brief Overview of Breast Cancer Treatment. In:*, pages 147–171. A Practical Guide to Breast Cancer Treatment. Springer Nature Singapore, Singapore, 2023.

[28] Matti Aapro, Juan Enrique Bargallo Rocha, Michele De Laurentiis, Roberto Elizalde, László Landherr, Barbro Linderholm, Terry Mamounas, Christos Markopoulos, Miguel Martin, Patrick Neven, Alexander Petrovsky, Dan Rea, Roman Rouzier, Vincent Smit, Christer Svedman, and Christoph Thomssen. The multidisciplinary application of genomics in clinical practice (magic) survey: Identification of early stage hormone receptor-positive (hr+), her2– breast cancer (bc) patients for whom multigene assays may be valuable. *Cancer research (Chicago, Ill.)*, 75(9):P6–06, 2015. doi: 10.1158/1538-7445.SABCS14-P6-08-06.

[29] Li Zhou, Maria Rueda, and Abedalrhman Alkhateeb. Classification of breast cancer nottingham prognostic index using high-dimensional embedding and residual neural network. *Cancers*, 14(4):934, 2022. doi: 10.3390/cancers14040934.

[30] Yiewfah Fong, J. Evans, D. Brook, J. Kenkre, P. Jarvis, and K. Gower-Thomas. The nottingham prognostic index: five- and ten-year data for all-cause survival within a screened population. *The Annals of The Royal College of Surgeons of England*, 97(2), 2015. doi: 10.1308/003588414x14055925060514.

[31] Sarfaraz K. Niazi. Rna therapeutics: A healthcare paradigm shift. URL `https://www.mdpi.com/2227-9059/11/5/1275`. [Accessed: 1st June 2024].

[32] Ling Xin, Yin-Hua Liu, Tracey A. Martin, and Wen G. Jiang. The era of multigene panels comes? the clinical utility of oncotype dx and mammaprint. *World Journal of Oncology*, 8(2):34–40, 2017. doi: 10.14740/wjon1019w.

[33] Peter Goodwin. In pilot study, use of genomic test spared two thirds of women with er+/her2- breast cancer from undergoing unneeded chemotherapy. *Oncology times*, 36(10):78, 2014. doi: 10.1097/01.COT.0000450363.54551.3c.

[34] Simon Gagnet, Caroline Diorio, Louise Provencher, Cynthia Mbuya-Bienge, Julie Lapointe, Claudya Morin, Julie Lemieux, and Hermann Nabi. Identifying clinicopathological factors associated with oncotype dx® 21-gene recurrence score: A real-world retrospective cohort study of breast cancer patients in quebec city, canada. *Journal of Personalized Medicine*, 11(9):858, 2021. doi: 10.3390/jpm11090858.

[35] Benjamin Harland and Mina Georgieva. *Survival Analysis. In:*, pages 93–124. Methods and Biostatistics in Oncology Understanding Clinical Research as an Applied Tool. Springer International Publishing, Switzerland, 2018.

[36] Amir Momeni, Matthew Pincus, and Jenny Libien. *Introduction to Statistical Methods in Pathology*. Springer International Publishing, Cham, 2018. URL https://library.biblioboard.com/viewer/5d193865-c380-11ea-9bfb-0ae0aa0d175d.

[37] DATAtab Team. Datatab: Online statistics calculator. URL https://datatab.net/statistics-calculator/survival-analysis. [Accessed: 1st March 2024].

[38] DATAtab Team. Log rank test, 2024. URL https://datatab.net/tutorial/log-rank-test.

[39] Samar Abd ElHafeez, Graziella D'Arrigo, Daniela Leonardis, Maria Fusaro, Giovanni Tripepi, and Stefanos Roumeliotis. Methods to analyze time-to-event data: The cox regression analysis. *Oxidative Medicine and Cellular Longevity*, 2021:1302811–6, 2021. doi: 10.1155/2021/1302811.

[40] Sthda. Cox proportional-hazards model - easy guides - wiki - sthda. URL http://www.sthda.com/english/wiki/cox-proportional-hazards-model. [Accessed: 1st March 2024].

[41] Mojtaba Hassanzad and Karimollah Hajian-Tilaki. Methods of determining optimal cut-point of diagnostic biomarkers with application of clinical data in roc analysis: an update review. *BMC Medical Research Methodology*, 24(1):84, 2024. doi: 10.1186/s12874-024-02198-2.

[42] Marwa El-Azazy. *Introductory Chapter: Infrared Spectroscopy - A Synopsis of the Fundamentals and Applications. In:*. Infrared Spectroscopy: Principles, Advances, and Applications. IntechOpen, London, 19 November 2018.

[43] Edinburgh Instruments. Beer lambert law: Transmittance absorbance. URL https://www.edinst.com/blog/the-beer-lambert-law/. [Accessed: 1st April 2024].

[44] Laurence Alexander Drummond. *Absorption spectral imaging in the mid-infrared and its application in cancer diagnosis*. PhD thesis, 2014.

[45] Yu Du, Fei Xie, Longfei Yin, Yang Yang, Houpu Yang, Guohua Wu, and Shu Wang. Breast cancer early detection by using fourier-transform infrared spectroscopy combined with different classification algorithms. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy*, 283:121715, Dec 15 2022. doi: 10.1016/j.saa.2022.121715.

[46] Cai Li Song, Martha Z. Vardaki, Robert D. Goldin, and Sergei G. Kazarian. Fourier transform infrared spectroscopic imaging of colon tissues: evaluating the significance of amide i and c–h stretching bands in diagnostic applications with machine learning. *Analytical and bioanalytical chemistry*, 411(26):6969–6981, Oct 1 2019. doi: 10.1007/s00216-019-02069-6.

[47] Hemmel Amrania, Giuseppe Antonacci, Che-Hung Chan, Laurence Drummond, William R. Otto, Nicholas A. Wright, and Chris Phillips. Digistain: a digital staining instrument for histopathology. *Optics express*, 20(7):7290–7299, 2012. doi: 10.1364/OE.20.007290.

[48] Rocky Mountain Labs. Difference between ftir and raman spectroscopy. URL `https://rockymountainlabs.com/difference-between-ftir-and-raman-spectroscopy/`.

[49] Juan Li, Xiaoting Wang, Shungeng Min, Jingjing Xia, and Jinyao Li. Raman spectroscopy combined with convolutional neural network for the sub-types classification of breast cancer and critical feature visualization. *Computer methods and programs in biomedicine*, 255:108361, Oct 2024. doi: 10.1016/j.cmpb.2024.108361.

[50] Maria Anthi Kouri, Ellas Spyratou, Maria Karnachoriti, Dimitris Kalatzis, Nikolaos Danias, Nikolaos Arkadopoulos, Ioannis Seimenis, Yannis S. Raptis, Athanassios G. Kontos, and Efstathios P. Efstathopoulos. Raman spectroscopy: A personalized decision-making tool on clinicians' hands for in situ cancer diagnosis and surgery guidance. *Cancers*, 14(5):1144, Feb 23 2022. doi: 10.3390/cancers14051144.

[51] Sebastien Jean Mambou, Petra Maresova, Ondrej Krejcar, Ali Selamat, and Kamil Kuca. Breast cancer detection using infrared thermal imaging and a deep learning model. *Sensors (Basel, Switzerland)*, 18 (9):2799, Aug 25 2018. doi: 10.3390/s18092799.

[52] Yerken Mirasbekov, Nurduman Aidossov, Aigerim Mashekova, Vasilios Zarikas, Yong Zhao, Eddie Yin Kwee Ng, and Anna Midlenko. Fully interpretable deep learning model using ir thermal images for possible breast cancer cases. *Biomimetics (Basel, Switzerland)*, 9(10):609, Oct 9 2024. doi: 10.3390/biomimetics9100609.

[53] Tom O'Haver. *Handbook of Scientific Signal Processing*. University of Maryland, 2024. doi: 0000-0001-9045-1603.

[54] Peak Spectroscopy Software. The importance of baseline correction in ftir spectroscopy. URL `https://www.peakspectroscopy.com/baselineCorrection.html`. [Accessed: 1st June 2024].

[55] Yang Xi, Yuee Li, Zhizhen Duan, and Yang Lu. A novel pre-processing algorithm based on the wavelet transform for raman spectrum. *Applied Spectroscopy*, 72(12):1752, 2018. doi: 10.1177/0003702818789695.

[56] Ben S. Wittner, Dennis C. Sgroi, Paula D. Ryan, Tako J. Bruinsma, Annuska M. Glas, Anitha Male, Sonika Dahiya, Karleen Habin, Rene Bernards, Daniel A. Haber, Laura J. Van39;t Veer, and Sridhar Ramaswamy. Analysis of the mammaprint breast cancer assay in a predominantly postmenopausal cohort. *Clinical Cancer Research*, 14(10):2988, 2011. doi: 10.1158/1078-0432.ccr-07-4723.

[57] R. Charles Coombes, Christina Angelou, Zamzam Al-Khalili, William Hart, Darius Francescatti, Nicholas Wright, Ian Ellis, Andrew Green, Emad Rakha, Sami Shousha, Hemmel Amrania, Chris C. Phillips, and Carlo Palmieri. Performance of a novel spectroscopy-based tool for adjuvant therapy decision-making in hormone receptor-positive breast cancer: a validation study. *Breast Cancer Research and Treatment*, 205(2):349–358, 2024. doi: 10.1007/s10549-023-07229-y.

[58] Health Tech Enterprise. Budget impact analysis of digistain® to guide adjuvant chemotherapy decisions in early breast cancer. Technical report, 2024.

# Appendix A: Reproduction Permissions

## Figure 2.4



## Figure 3.1

# Figure 4.2

**Zamzam** 3:38 AM

Permission to include an image from your website

To: mathias.jesussek@datatab.de

Dear Dr. Mathias Jesussek,

I am completing my PhD thesis at Imperial College London entitled: Digistain Prognostic Score Development - Infrared Spectroscopy to Guide Chemotherapy Decisions.

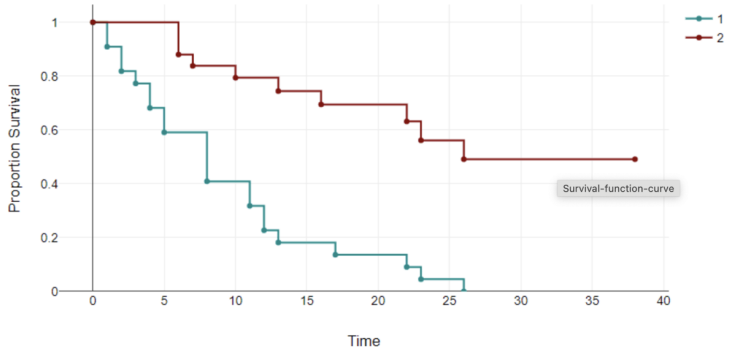
I seek your permission to reprint, in my thesis an image from (DATAtab Team. Datatab: Online statistics calculator. URL https://datatab.net/statistics- calculator/survival-analysis. [Accessed: 1st March 2024].).

I would like to include the image in my thesis which will be added to Spiral, Imperial's institutional repository http://spiral.imperial.ac.uk/ and made available to the public.

The image is attached.

Yours sincerely,

Zamzam Al-Khalili




Found in Inbox - IC Mailbox

**Mathias Jesussek** 12:22 PM

AW: Permission to include an image from your website

To: Al-Khalili, Zamzam A H



Dear Zamzam Al-Khalili,

yes sure you can use it!
But please cite DATAtab!

Thanks and Regards,
Mathias

# Figure 5.11 and Figure 5.12

Convergent Science Physical Oncology

OPEN ACCESS

PAPER

## Mid-infrared imaging in breast cancer tissue: an objective measure of grading breast cancer biopsies

H Amrania[1], L Woodley-Barker[4], K Goddard[2], B Rosales[2], S Shousha[4], G Thomas[4], T McFarlane[4], M Sroya[4], C Wilhelm-Benartzi[5], K Cocks[3], R C Coombes[2] and C C Phillips[1,6]

[1] Physics Department, Imperial College London, Prince Consort Rd., London, SW7 2BW, United Kingdom
[2] Faculty of Medicine, Department of Cancer and Surgery, Imperial College London, Du Cane Road, London W12 0NN, United Kingdom
[3] KCStats Consultancy, Leeds, United Kingdom
[4] Faculty of Medicine, Department of Surgery and Cancer, Imperial College London, Charing Cross Hospital, Fulham Palace Rd., London W6 8RF, United Kingdom
[5] Centre for Trials Research, College of Biomedical and Life Sciences, Cardiff University, 6th Floor, Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4YS, United Kingdom
[6] Author to whom any correspondence should be addressed.
Professor Chris Phillips, Experimental Solid State Group, Room 911, Physics Department, Experimental Solid State Group, Physics Department, Imperial College London, SW7 2BW, London.

**E-mail:** Chris.phillips@imperial.ac.uk

**Keywords:** bioimaging, histopathology grading, mid-infrared spectroscopic imaging, breast cancer, H + E pathology, cancer diagnosis

RECEIVED
18 December 2017

REVISED
23 January 2018

ACCEPTED FOR PUBLICATION
31 January 2018

PUBLISHED
13 March 2018



### Abstract

# Figure 6.2