



# City Research Online

## City St George's, University of London

**Citation:** Ortega-Ruiz, M. A. (2025). Analysis of immunostained histological images of breast cancer. (Unpublished Doctoral thesis, City St George's, University of London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/35418/>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

# Analysis of immunostained histological images of breast cancer

Mauricio Alberto Ortega-Ruíz

This dissertation is submitted for the degree of  
*Doctor of Philosophy*



University of London  
Department of Computer Science

January 2025



# Contents

<b>Contents</b>	<b>iii</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xiii</b>
<b>Acknowledgements</b>	<b>xv</b>
<b>Declaration</b>	<b>xvii</b>
<b>Abstract</b>	<b>xix</b>
<b>Abbreviations and Acronyms</b>	<b>xxii</b>
<b>List of publications</b>	<b>xxv</b>
<b>1 Introduction and overview</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 Problem identification . . . . .	4
1.3 Motivation . . . . .	6
1.4 Aims and objectives of this work . . . . .	6
1.5 Contributions . . . . .	7
1.6 Organization of the thesis . . . . .	7
1.7 Summary . . . . .	8
<b>2 Background</b>	<b>9</b>
2.1 Breast Cancer . . . . .	9
2.1.1 Breast tissue . . . . .	9
2.1.2 Breast cancer (BC) . . . . .	10
2.1.3 Breast Cancer Diagnosis . . . . .	11

2.1.4	Histology . . . . .	13
2.1.5	Whole Slide Imaging . . . . .	16
2.2	Digital Pathology in Breast Cancer . . . . .	19
2.2.1	Segmentation of BC images . . . . .	19
2.2.2	Digital pathology methods evolution . . . . .	19
2.2.3	Deep Learning Evolution . . . . .	22
2.3	Summary . . . . .	23
<b>3</b>	<b>Related Work</b>	<b>25</b>
<b>4</b>	<b>Materials</b>	<b>31</b>
4.1	SPIE-AAPM-NCI BreastPathQ challenge dataset . . . . .	31
4.2	Breast Cancer Semantic Segmentation (BCSS) dataset . . . . .	33
4.3	Automated Gleason Grading Challenge 2022 dataset . . . . .	33
4.4	Breast cancer ACROBAT dataset . . . . .	37
4.5	Summary . . . . .	39
<b>5</b>	<b>Tumour cellularity assessment</b>	<b>41</b>
5.1	Overview . . . . .	41
5.2	Materials . . . . .	42
5.3	Methods . . . . .	43
5.3.1	Colour separation . . . . .	44
5.3.2	Nuclei Segmentation . . . . .	45
5.3.3	Ducts segmentation . . . . .	49
5.3.4	Extraction of nuclei morphological parameters . . . . .	50
5.3.5	Extraction of regional morphological parameters . . . . .	55
5.3.6	Correlation analysis of morphological parameters to TC . . .	57
5.3.7	Training of machine learning algorithms . . . . .	57
5.4	Assessment of Tumour Cellularity . . . . .	61
5.5	Results . . . . .	61
5.6	Summary . . . . .	64
<b>6</b>	<b>Deep learning networks for multi-class segmentation</b>	<b>65</b>
6.1	Segmentation of multi-class BC histopathology images related work .	65
6.2	Materials . . . . .	66
6.3	Method . . . . .	69

6.3.1	The DL models . . . . .	69
6.3.2	Training the DL models . . . . .	72
6.4	Results . . . . .	73
6.5	Summary . . . . .	78
<b>7</b>	<b>Gleason Grading assignment for prostate cancer</b>	<b>79</b>
7.1	Motivation: Prostate Cancer and the Gleason Grading score . . . . .	79
7.2	Materials . . . . .	80
7.2.1	Extracted dataset 1 . . . . .	83
7.2.2	Extracted dataset 2 . . . . .	83
7.3	Method . . . . .	84
7.3.1	Pipeline . . . . .	85
7.3.2	Training the DL model . . . . .	86
7.4	Results . . . . .	87
7.5	Summary . . . . .	90
<b>8</b>	<b>UNet analysis and modifications for an improved architecture</b>	<b>91</b>
8.1	UNet an state-of-the-art segmentation network . . . . .	92
8.1.1	UNet motivation . . . . .	92
8.1.2	Deconvolutional Neural Networks (DeCNN) . . . . .	92
8.1.3	UNet architecture . . . . .	93
8.2	Related work . . . . .	94
8.3	Dilation, Residual and Dense (DRD) block motivation . . . . .	96
8.4	Methods . . . . .	97
8.4.1	Original UNet Model used in the experiments . . . . .	98
8.4.2	Experiments over the UNet . . . . .	100
8.4.3	Hyper-parameters selection . . . . .	105
8.4.4	Implementation Details . . . . .	106
8.5	Results . . . . .	106
8.6	Summary . . . . .	111
<b>9</b>	<b>Conclusions and Future Work</b>	<b>115</b>
9.1	Summary . . . . .	115
9.2	Main contributions . . . . .	117
9.3	Main conclusions . . . . .	117

9.4	Suggestions for future work . . . . .	118
<b>A</b>	<b>Digital images in medicine</b>	<b>121</b>
A.1	Digital images . . . . .	121
A.1.1	Binary images . . . . .	122
A.1.2	Grayscale images . . . . .	122
A.1.3	Colour images . . . . .	124
A.2	Images in the frequency spectrum . . . . .	126
<b>B</b>	<b>Wide Slide Scanners</b>	<b>129</b>
<b>C</b>	<b>Sensitivity and specificity performance of DRD-UNet model</b>	<b>131</b>
<b>D</b>	<b>Deep Learning Background</b>	<b>135</b>
D.1	Neural Network Structure . . . . .	135
D.1.1	Input Layer . . . . .	137
D.1.2	Fully connected Layer . . . . .	137
D.1.3	Convolutional Layer . . . . .	137
D.1.4	Maxpool Layer . . . . .	138
D.1.5	SoftMax Layer . . . . .	138
D.2	Network functionality: forward propagation . . . . .	138
D.3	Network performance and loss function . . . . .	138
D.3.1	Mean Absolute Error MAE . . . . .	139
D.3.2	Mean Square Error MSE . . . . .	139
D.3.3	Binary cross entropy loss BCE . . . . .	139
D.4	Fitting the model and training . . . . .	139
D.4.1	Gradient Descent . . . . .	140
D.4.2	Stochastic Gradient Descent . . . . .	140
D.4.3	Gradient descent with momentum . . . . .	141
D.4.4	Adaptive Moment estimation or ADAM . . . . .	141
D.5	Model Performance . . . . .	142
D.5.1	Binary and multi-class classification . . . . .	142
D.5.2	Accuracy . . . . .	143
D.5.3	Precision . . . . .	143
D.5.4	Sensitivity . . . . .	143
D.5.5	Specificity . . . . .	143

D.5.6	Dice coefficient . . . . .	143
D.5.7	Jaccard index . . . . .	144
D.5.8	Receiver Operating Characteristic . . . . .	144
D.5.9	F-measure . . . . .	144
D.5.10	Multi-class measurement example . . . . .	144

<b>References</b>		<b>147</b>
-------------------	--	------------



# List of Figures

1.1	Factors that cause cancer: oncogenes promote cancer and tumour suppressor genes. . . . .	3
1.2	2023 cancer statistics . . . . .	5
2.1	Comparison between an histopathology breast tissue and the structure of the breast. . . . .	10
2.2	Visual description of the evolution of ductal carcinoma . . . . .	11
2.3	Samples of H&E images of normal tissue, benign tumor, DCIS and IDC. . . . .	11
2.4	Ultrasound image example . . . . .	13
2.5	Example of an H&E staining image. . . . .	14
2.6	Example of an IHC staining image. . . . .	15
2.7	Samples from ACROBAT dataset . . . . .	16
2.8	Different magnification levels of a WSI sample image, viewed using a specialized digital pathology platform. . . . .	17
2.9	Description of the pyramidal structure of a WSI. . . . .	18
2.10	Visualization of WSI levels following specialized instructions for WSI manipulation. . . . .	18
2.11	Artificial Intelligence . . . . .	20
2.12	Biological Neuron . . . . .	21
2.13	Artificial Neuron . . . . .	22
4.1	Patch selection for tumor cellularity (TC) assessment by the pathologist from a whole slide image (WSI). . . . .	32
4.2	Four Samples of images from the BreastPathQ dataset (Peikari et al., 2017) of TC values from zero to one. . . . .	32

4.3	Examples 1 and 2 of images from the BCSSC with the annotation by the pathologist . . . . .	34
4.4	Examples 3 and 4 of images from the BCSSC with the annotation by the pathologist . . . . .	35
4.5	Scanners used for the AGGC (Guo et al., 2019) with samples of the same tissue. . . . .	36
4.6	Sample images from the AGGC dataset (Guo et al., 2019) shown at the lowest resolution . . . . .	37
4.7	Samples from ACROBAT dataset (Weitz et al., 2022) of the same biopsy with different staining methods . . . . .	38
5.1	Samples of different cellularity values. . . . .	43
5.2	Statistical distribution of the SPIE BreastPathQ dataset. . . . .	43
5.3	Pipeline of the methodology for TC assessment. . . . .	44
5.4	Three examples of colour separation by K-means . . . . .	46
5.5	Histograms before and after the enhanced image. . . . .	47
5.6	Example of nuclei segmentation and enhancement process . . . . .	49
5.7	Nuclei segmentation without the nuclei enhancement . . . . .	50
5.8	Example of clusters detection and clusters segmentation . . . . .	51
5.9	Regional analysis of a sample image . . . . .	56
5.10	Graph of extracted parameters vs. cellularity . . . . .	58
5.11	Classification with SVM . . . . .	59
5.12	KNN explanation of Euclidean distance and a new data assignment. . . . .	60
5.13	Classification with AdaBoost . . . . .	61
5.14	Visual description of the method . . . . .	62
5.15	Boxplot results for the three trained models . . . . .	63
6.1	Sample WSI image from the BCSS challenge, showing the annotated area . . . . .	68
6.2	Example of extracted $256 \times 256$ patches used for model training . . . . .	68
6.3	Depthwise and pointwise operations as described by (Howard, 2017) . . . . .	70
6.4	Description of the residual operation. . . . .	71
6.5	UNet . . . . .	71
6.6	Semantic segmentation example . . . . .	74
6.7	Pixel-wise comparison of segmentation results . . . . .	75

7.1	Sample image from the AGGC dataset . . . . .	82
7.2	Pipeline for processing WSI images from the AGGC dataset . . . . .	84
7.3	Example of background removal in a WSI; an ROI mask is obtained as a result of this pre-processing step . . . . .	85
7.4	Patch extraction during the training phase of the pipeline, performed over the region annotated by the pathologist . . . . .	86
7.5	Patch extraction during the test phase of the pipeline, performed over the entire tissue region after the background removal . . . . .	87
7.6	Example of Gleason Grade (GG) assessment on a test image . . . . .	88
7.7	Validation accuracy and validation loss curves obtained during train- ing of dataset 1. . . . .	89
7.8	Validation accuracy and validation loss curves obtained during train- ing of dataset 2. . . . .	89
8.1	UNet architecture . . . . .	93
8.2	Dilated Convolution . . . . .	96
8.3	Residual vs. Dense . . . . .	97
8.4	UNet modifications . . . . .	101
8.5	DPB3 Block . . . . .	102
8.6	ResUNet-a block . . . . .	103
8.7	Dilated Convolution in series . . . . .	104
8.8	RMS Block . . . . .	104
8.9	DRD Block . . . . .	105
8.10	Loss curves . . . . .	112
8.11	Segmentation results . . . . .	113
8.12	Segmentation results . . . . .	114
A.1	Example of a binary image of a segmented cell, illustrating the matrix elements and their corresponding pixels . . . . .	122
A.2	Example of a gray IHC image . . . . .	123
A.3	Example of an H&E-stained image, composed of blue and pink colour components . . . . .	124
A.4	Graphical representation of the colour components in an H&E-stained image . . . . .	125

A.5	Example of an IHC image, composed of blue and brown colour components . . . . .	125
A.6	Graphical representation of the colour components in an IHC-stained image . . . . .	126
D.1	3 hidden units example . . . . .	136
D.2	2 hidden layers example . . . . .	137
D.3	Illustration of performance calculation . . . . .	145

# List of Tables

1.1	Types of cancer classified by their tissue of origin . . . . .	4
2.1	Comparison of the characteristics of breast cancer screening methods.	13
4.1	Comparison between different datasets. . . . .	38
5.1	Table for the extracted parameters at nuclei, regional, clusters, and the global image. . . . .	55
5.2	Table for ICC results between hand engineering (Peikari et al., 2017), this research, and combined DL (Akbar et al., 2019). . . . .	64
6.1	BCSS dataset characteristics, the table describes the original dataset and the extracted dataset parameters . . . . .	67
6.2	Accuracy achieved with the selected parameters . . . . .	74
6.3	Segmentation results comparison . . . . .	76
6.4	Segmentation results comparison . . . . .	77
7.1	AGGC dataset characteristics, including the total annotated area marked by the pathologist . . . . .	81
7.2	Gleason Grading Performance . . . . .	90
8.1	UNet implementation in MATLAB . . . . .	98
8.2	UNet implementation in MATLAB . . . . .	99
8.3	Experiments description . . . . .	100
8.4	Jaccard similarity index results . . . . .	107
8.5	Dice coefficient Results . . . . .	108
8.6	Accuracy Results . . . . .	109
C.1	Sensitivity Index Results . . . . .	132
C.2	Specificity index Results . . . . .	133



# Acknowledgements

Firstly, I want to thank Dr. Constantino Carlos Reyes-Aldasoro for accepting me as a doctoral student, and his guidance and supervising work during the time as City student and the previous months we started this research, you always gave me great ideas to carry out. Also thanks to Dr. Edgar Román-Rangel, my second supervisor, for the support in this stage of my PhD, all your suggestions always were very valuable. I really appreciate it. Thank you, Cefa, for the opportunity to collaborate on this research.

Thanks to Dr. Emmanuel López Neri, Dr. Gerardo Dubcovsky from the National Research Division at Universidad del Valle de México UVM, and to our UVM local authorities. Also, to Dr. Humberto Ortega, my father, for your support and those enriching talks about medicine.

To my loving wife Rosana, for your support and motivation. To my daughters and son, Elizabeth, Mauricio, and Sophie, may my achievements serve as an inspiration for you.



# Declaration

I, Mauricio Alberto Ortega-Ruiz confirm that work presented in thesis is my own. Any information, idea, method, or illustration originated from other's subject work has been indicated in the Thesis. I declare that this Thesis has never been published or submitted elsewhere for obtaining an academic degree or professional qualification. This thesis is my own work and contains nothing that is the outcome of work done in collaboration with others, except as specified in text, and Acknowledgements. I grant power of discretion to the Librarian at City, University of London to allow the thesis to be copied in whole or in part without further reference. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgment.

*Mauricio Alberto Ortega-Ruiz*



# Abstract

This work analyses breast cancer histopathology images with the objective of performing segmentation of the different tissue components, with special focus on the tumour region, which is of irregular shape and of complex characteristics, in terms of texture, light intensity, and contrast. On the other hand, the images come from specialized slide scanners, that digitises the glass slide into a high resolution digital file. Manipulating a single gigabyte-scale image is a challenging and time-consuming task, requiring an efficient implementation during experimentation. The images analysed are primarily of breast cancer, although the selected approach was also used for prostate cancer images. Two approaches have been explored for this research work, a hand-craft method based on image processing algorithms for feature extraction, and Deep Learning models. Both methods reveal the evolution in digital pathology during the last decade, and both were analysed and implemented in different clinical tasks. The first task is the tumour cellularity assessment in patients under Neo-Adjuvant treatment, which is a medical treatment of breast cancer given before the main treatment. In this experiment, the tumour area of the image is estimated. Next, multi-class segmentation is analysed on breast cancer tissue, with the main goal of segmenting the different tissue components. All of the images come from the grand challenge platform, and the results presented in this work were also submitted to the corresponding challenge. This thesis provides a comprehensive workflow of the digital pathology for breast cancer analysis, covering both approaches. Future research can be done in other types of cancer as well as in newer pathology stained images.





# Abbreviations and Acronyms

ACS	American Cancer Society
AI	Artificial Intelligence
AGGC	Automatic Gleason Grading Challenge
AdaBoost	Adaptive Boosting
ANN	Artificial Neural Network
AUC	Area Under the Curve
BC	Breast Cancer
BCSS	Breast Cancer Semantic Segmentation
CAD	Computer Aided Diagnosis
CNN	Convolutional Neural Network
DCIS	Ductal Carcinoma in Situ
DCNN	Deep Convolutional Neural Network
DeCNN	Deconvolutional Neural Network
DNN	Deep Neural Network
DL	Deep Learning
CT	Computer Tomography
DPB	Dilated Pyramid Block
ER	Estrogen
FCN	Fully Convolutional Network

H&E	Haematoxiline and Eosin
HER2	Human Epidermal Growth Factor Receptor 2
ICC	Intraclass Correlation Coefficient
IDC	Invasive Ductal Carcinoma
IHC	Immunohistochemistry
HSV	Hue, Saturation, Value
ILC	Invasive Lobular Carcinoma
GAN	Generative Adversial Network
KNN	Nearest K-Network
ML	Machine Learning
MRI	Magnetic Resonant Image
MSE	Mean Square Error
NAT	Neo-Adjuvant Treatment
PGR	Progesterone
ReLU	Rectified Linear Unit
RGB	Red, Green Blue
SVM	Support Vector Machine
TC	Tumour Cellularity
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives



# List of publications

Journal publications:

- 1 Ortega-Ruiz MA, Karabağ C, Garduño VG, Reyes-Aldasoro CC. Morphological Estimation of Cellularity on Neo-Adjuvant Treated Breast Cancer Histological Images. *Journal of Imaging*. 2020; 6(10):101. <https://doi.org/10.3390/jimaging6100101>
- 2 Karabağ, C.; Ortega-Ruiz, M.A.; Reyes-Aldasoro, C.C. Impact of Training Data, Ground Truth and Shape Variability in the Deep Learning-Based Semantic Segmentation of HeLa Cells Observed with Electron Microscopy. *J. Imaging* 2023, 9, 59. <https://doi.org/10.3390/jimaging9030059>
- 3 M. A. Ortega-Ruiz, C. Karabağ, E. Roman-Rangel and C. C. Reyes-Aldasoro, "DRD-UNet, a UNet-Like Architecture for Multi-Class Breast Cancer Semantic Segmentation," in *IEEE Access*, vol. 12, pp. 40412-40424, 2024, doi: 10.1109/ACCESS.2024.3377428.
- 4 Ayala, R.E.D.; Granados, D.P.; Gutiérrez, C.A.G.; Ruiz, M.A.O.; Espinosa, N.R.; Heredia, E.C. Novel Study for the Early Identification of Injury Risks in Athletes Using Machine Learning Techniques. *Appl. Sci.* 2024, 14, 570. <https://doi.org/10.3390/app14020570>
- 5 Multiclass Semantic Segmentation of Immunostained Breast Cancer Tissue with a Deep-Learning Approach Mauricio Alberto Ortega-Ruiz, Edgar Roman-Rangel, Constantino Carlos Reyes-Aldasoro medRxiv 2022.08.17.22278889; doi: <https://doi.org/10.1101/2022.08.17.22278889>.

Scientific dissemination:

- 1 Ortega-Ruiz MA, ¿Qué es la Visión Artificial?, *Expresiones UVM* Vol.06 No 16, ISSN: 2448-7309, <https://issuu.com/expresionesuvm/docs/v6n16>



# Chapter 1

## Introduction and overview

### 1.1 Overview

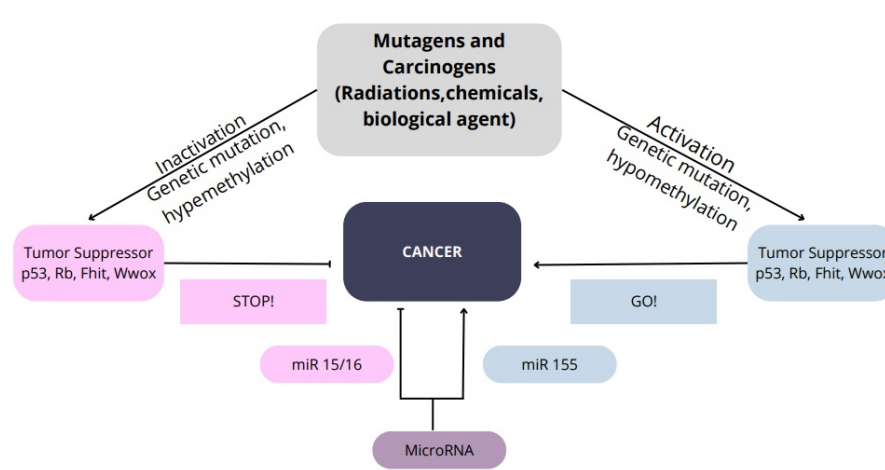
Histopathology is the study of a disease based on the observation of tissue samples under the microscope, and it plays an important role in the diagnosis of cancer, because abnormalities in tissue and cells can be identified (Dey, 2022; Veta et al., 2014; Irshad et al., 2013). The cells observed under the microscope in natural state lack of colour and contrast and are not visible, therefore a tissue section obtained by biopsy is stained to reveal cellular components. Two common staining techniques with a well-defined procedure are used in histopathology: haematoxylin and eosin (H&E), and immunohistochemistry (IHC) (Veta et al., 2014); which make the tissue visible to be analysed by the pathologist. Today, a modern histopathology scanner can digitise the image from the tissue to obtain a high-resolution digital file, that can be visualised and processed by a computer. With the aid of specialised application software, pathologists have now tools that assist them in the diagnosis, therefore, this support tool is sometimes called Computer Aided Diagnosis (CAD) (Gurcan et al., 2009), and the full process is known as digital pathology or computational pathology (Hosseini et al., 2024). This support tool is available mainly due to advances in computer processors, software tools, and the quality of scanner hardware. As a result, CAD helps relieve the workload on pathologists (Veta et al., 2014). After a positive confirmation of the disease, a quantitative measurement of the tissue reveals the severity of the cancer by the experienced pathologist. A well-defined workflow for computational pathology analysis, based on the one done by the specialist, is: colour normalisation, colour separation, selection of a region

of interest, nuclei segmentation, features extraction, and cell classification (Gurcan et al., 2009; Song et al., 2023). Colour normalisation is required due to the variations in the staining and scanning of histopathological samples, for example, the normalisation can be performed by processing the colour histograms of an image (Reinhard et al., 2001a). The selection of a region of interest (ROI) is a relevant step in computational pathology particularly when high-resolution images scanned from the full tissue are studied, due to the size of the image only the selected interest region is processed, this avoids unnecessary processing time, like doing analysis of a background area which has no information. Next, to quantify the cancer severity, the pathologist requires to identify and quantify the histological structures of the tissue (Michael et al., 2021a), which is performed by segmenting task. The features extracted from the segmented structures are based on the visual attributes (Gurcan et al., 2009) observed by clinicians for disease classification and diagnosis, some examples of these parameters are nucleus size, shape, texture, among others (Boucheron, 2008). Finally, in the classification step, special algorithms use the extracted features to determine if the segmented structure is tumour or non-tumour.

Machine learning (ML) methods have been used for classification, which are special algorithms that learn and improve with data without being explicitly programmed to do so, and are grouped into supervised, nonsupervised and reinforcement learning (Rashidi et al., 2019). Unsupervised methods uncover inherent structures or features within the data without previous labeling, for example, K-means (Shalev-Shwartz and Ben-David, 2014; Takano, 2021) is a method for grouping similar data, therefore it can be used for classifying the tissue. Reinforcement learning is a ML method for a decision-making process that is trained to learn the optimal behaviour (Rashidi et al., 2019). Supervised methods require training with labelled input-output reference data, such as deep neural networks (DNNs) (Goodfellow, 2016; Komura and Ishikawa, 2018), in addition, the inclusion of DNN into the pathology workflow process has transformed computational pathology by enhancing accuracy and efficiency (Hosseini et al., 2024). New relevant medical tasks have been studied instead of the diagnostic task (Madabhushi and Lee, 2016). For example, to discover new relationships between specific medical behaviors and morphological properties—as in the work of (Beck et al., 2011), which revealed that the stromal region has a strong relationship with cancer survival. Computational pathology has been used as a support tool for common pathological tasks, but the potential of

this tool might have an enormous impact. Also, computational pathology tends to be gradually included in daily practice for automated morphological biomarkers detection and evaluation of more complex clinical tasks (Song et al., 2023).

The word cancer has its origin in Hippocrates who named *karkinos*, the Greek word for crab, a breast tumour that resembles a crab. Cancer is a genetic disease that causes alterations in the normal growth and life cycle of cells. Abnormal cells accumulate forming tumours that can extend to other organs of the body affecting their function. This accumulation is originated by genetic mutations or errors in DNA. There are genes involved in cancer formation (oncogenes), tumour-suppressing genes, and DNA repair genes, some of which can be altered and promote tumour proliferation (Pardee and Stein, 2009). Oncogenes are activated in cancer and lead to uncontrolled cell proliferation, whereas a tumour-suppressor gene reduces the probability that a cell will turn into a tumour cell. Micro-RNA genes are non-coding genes that maintain the stability of the human genome. This complex relationship is expressed in figure 1.1, on the left side are the tumour suppressor genes, oncogenes on the right side, and stability genes on the bottom. Many causes of cancer are external factors, for example, tobacco, infectious agents, radiation, sun exposure, and even certain food diets that increase the risk of developing the disease (Pardee and Stein, 2009). There are also internal causes like inherited mutations, but these are not frequent (Aqeilan et al., 2008).



**Figure 1.1:** The complex relationship between the factors that cause cancer. Oncogenes (right side) control cell proliferation, tumour suppressors (left side) reduce it, and stability genes (bottom) support the stability. Image adapted from (Pardee and Stein, 2009).

Types of cancer can be described according to the tissue, the organ from which is originated, and also based on the degree of severity. Types based on tissue are

carcinoma, sarcoma, and lymphoma (Weigelt et al., 2010). The most common organs affected are the breast, prostate, lung & bronchus, colon & rectum, and pancreas, among others. Table 1 lists the types of cancer based on the area of origination.

Type	Area of origination
Adenocarcinoma	Glandular tissue
Blastoma	Embryonic tissue of organs
Carcinoma	Epithelial tissue
Leukemia	Tissues that form blood cells
Lymphoma	Lymphatic tissue
Myeloma	Bone marrow
Sarcoma	Connective or support tissue


**Table 1.1:** The main types of cancer based on the origin are listed in the table, carcinoma and adenocarcinoma are the common types of breast cancer, (Pardee and Stein, 2009).

Breast cancer (BC) is a serious disease, and one of the most common cancers in women, 31% of new estimated cases and 15% of estimated deaths worldwide in 2023 were of this kind of cancer, as reported in (Siegel et al., 2023), and shown in the figure 1.2.


## 1.2 Problem identification

Image processing of pathological images is important for breast cancer diagnosis (Zhang et al., 2021a). Segmentation is a crucial task within the pathology workflow. It involves partitioning an image into its objects, allowing elements such as tumours to be extracted and studied from the images. Computer-based methods for cell segmentation have been explored since the early days of digital pathology. This task is necessary for cell quantification, which measures cell proliferation, an essential parameter for grading the severity of breast cancer. The segmentation of the full tumour region and quantification of the tumour area is relevant for grading the tumour (Michael et al., 2021b). Invasive Ductal Carcinoma (IDC) and Ductal Carcinoma in Situ (DCIS), which will be described in chapter 2, have been studied by digital pathology methods. However, direct segmentation of the tumour area in histopathology images has not been fully explored (Guo et al., 2019). Segmentation of the tumour is a challenging task, due to tumour shape, intensity variation, tex-

### Estimated New Cases

		Males		Females		
Prostate	288,300	29%		Breast	297,790	31%
Lung & bronchus	117,550	12%		Lung & bronchus	120,790	13%
Colon & rectum	81,860	8%		Colon & rectum	71,160	8%
Urinary bladder	62,420	6%		Uterine corpus	66,200	7%
Melanoma of the skin	58,120	6%		Melanoma of the skin	39,490	4%
Kidney & renal pelvis	52,360	5%		Non-Hodgkin lymphoma	35,670	4%
Non-Hodgkin lymphoma	44,880	4%		Thyroid	31,180	3%
Oral cavity & pharynx	39,290	4%		Pancreas	30,920	3%
Leukemia	35,670	4%		Kidney & renal pelvis	29,440	3%
Pancreas	33,130	3%		Leukemia	23,940	3%
<b>All Sites</b>	<b>1,010,310</b>	<b>100%</b>	<b>All Sites</b>	<b>948,000</b>	<b>100%</b>	

### Estimated Deaths

		Males		Females		
Lung & bronchus	67,160	21%		Lung & bronchus	59,910	21%
Prostate	34,700	11%		Breast	43,170	15%
Colon & rectum	28,470	9%		Colon & rectum	24,080	8%
Pancreas	26,620	8%		Pancreas	23,930	8%
Liver & intrahepatic bile duct	19,000	6%		Ovary	13,270	5%
Leukemia	13,900	4%		Uterine corpus	13,030	5%
Esophagus	12,920	4%		Liver & intrahepatic bile duct	10,380	4%
Urinary bladder	12,160	4%		Leukemia	9,810	3%
Non-Hodgkin lymphoma	11,780	4%		Non-Hodgkin lymphoma	8,400	3%
Brain & other nervous system	11,020	3%		Brain & other nervous system	7,970	3%
<b>All Sites</b>	<b>322,080</b>	<b>100%</b>	<b>All Sites</b>	<b>287,740</b>	<b>100%</b>	

**Figure 1.2:** Cancer statistics for the year 2023. Breast cancer has the highest estimated new cases with 31%, and estimated death cases are also high with 15%, source of the image: (Siegel et al., 2023) under creative common license.

ture, and location (Su et al., 2015). Also, there is a lack of labeled reference images (Priego-Torres et al., 2020). The development of efficient computer-based methods would enhance digital pathology and promote efficient workflow.

### **1.3 Motivation**

Early diagnosis of breast cancer is relevant, as detecting and treating it in its early stages can prevent it from spreading to other parts of the body (Champaign and Cederbom, 2000; Sasieni et al., 2023). Smaller tumours are generally easier to remove surgically, and chemotherapy or radiation therapy could be less intensive. Breast histopathology is used to determine the grade of breast cancer, usually based on three main levels: in grade 1 (the lower grade), cancer cells look relatively similar to healthy cells and grow slowly. Grade 2 is the intermediate and cancer cells look bigger and more abnormal than grade 1 cells. Grade 3 is the highest grade and cells grow and spread quickly (Theissig et al., 1990). There are several treatment options when detected early and also, there is a higher survival rate compared to the rate in later stages (Crosby et al., 2020). When breast cancer is detected at an early stage, the 5-year relative survival rate is significantly higher. This rate refers to the percentage of patients who live at least five years after a positive diagnosis. In early-stage detection, the survival rate can be as high as 99%. However, when the cancer progresses to grade 2 or 3, this rate drops significantly. Early diagnosis gives also the opportunity for a better quality of life. In a less invasive treatment, higher survival rates positively impact patients' physical and mental well-being. Finally, early treatment reduces the chance of a future return.

### **1.4 Aims and objectives of this work**

This work is focused on the study of ML techniques to analyse, quantify, and segment the breast cancer tumor from IHC and H&E stained images. The main aim of this work is to propose a novel methodology for the segmentation of cancer histopathological images. To achieve this several objectives have to be met;

- a To study the state-of-the-art in the analysis of breast cancer images for tumour segmentation and to identify the most relevant computer pathology methods.

- b To implement a known model and compare and analyse the performance of the selected model with different histopathology images.
- c To propose an improved model that outperforms current models for the segmentation of histopathological images of breast cancer.

## 1.5 Contributions

The contributions of our work are summarized in two main areas, first, the importance of image processing and computer vision methods is emphasized, called traditional methods, as a support for newer ML and DL techniques. Traditional methods were explored for nuclei segmentation and tissue classification. A nuclei segmentation method was proposed, based on nuclei enhancement to quantify tumour cellularity. The morphological analysis of BC images, based on traditional methods, produced a journal publication (Ortega-Ruiz et al., 2020). The next contribution is the result of the study that reveals some of the important characteristics of a segmentation model called UNet (Ronneberger et al., 2015), as a result, a novel model called DRD-UNet that outperforms the original UNet is proposed. The evaluation and the results of the proposed model generated a publication (Ortega-Ruiz et al., 2024).

## 1.6 Organization of the thesis

Chapter 2 addresses the importance of image segmentation with an overview of the evolution of methods for breast cancer segmentation, starting from traditional methods to DL techniques, and it also includes a basic description of breast tissue for a better comprehension of the type of images under analysis. Chapter 3 covers the related work and describes the images used in the referenced studies. Chapter 4 includes a description of the datasets used in this work. Chapter 5 presents a full digital pathology experiment for tumour cellularity estimation in neo-adjuvant breast cancer treatment, which is a special therapy administrated before the main treatment. The experiment is mainly based on traditional computer vision methods. Chapter 6 presents a comparative analysis of three well-known Deep Learning networks for multi-class segmentation: MobilenetV2, ResNet, and UNet. A relevant conclusion derived from this chapter is the performance of UNet over the other

models, based on the accuracy. Chapter 7 describes a segmentation approach used for the assessment of the Gleason score of prostate cancer. The images came from the Gleason Grading Challenge 2022, a contest for the development of methods for grading prostate cancer. An ablation study for the UNet model is presented in Chapter 8. This study gives a better understanding of this state-of-the-art model and proposes an improvement of UNet model named DRD-UNet. In this analysis, a comparison between different relevant modifications implemented in UNet is presented. The metrics used for this comparison are Dice, Jaccard, and Accuracy. Finally, Chapter 9 presents a summary of this thesis and future work directions.

## 1.7 Summary

A general overview of digital pathology for breast cancer is presented in this chapter. The workflow for the pathology image analysis was described in this chapter, which is based on colour separation, nuclei segmentation, feature extraction, and tissue classification. This workflow is executed by a computer that is used as a support tool for the pathologist's daily work. The main motivation of this research is the relevance that breast cancer has in the total cases per year, which motivates to propose efficient methods for analysing breast cancer images.

The main goals of this research are stated as well as the main contributions and achievements during the study, with the description of the full thesis organization, this chapter gives a general comprehension of the scope of this research.

## Chapter 2

# Background

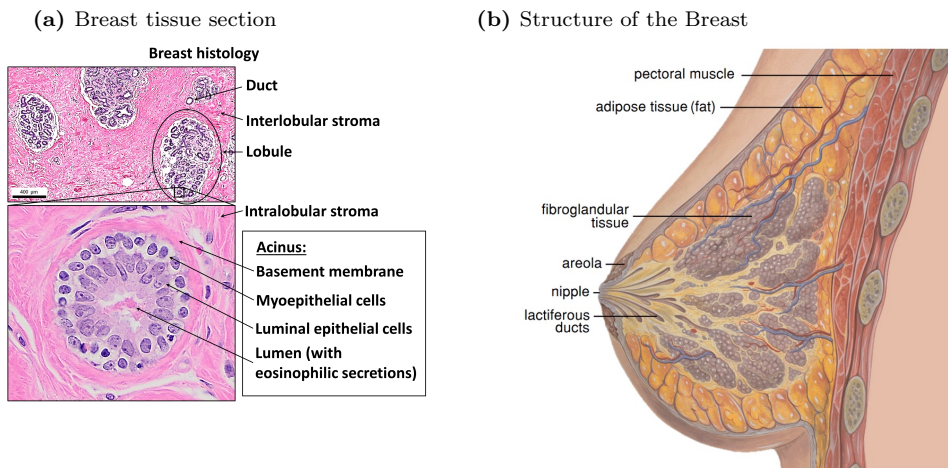
This research is focused on the segmentation of Breast Cancer (BC) histopathology images, and calculation of Tumour Cellularity (TC). A general background is presented in this chapter. It begins with a basic explanation of breast histopathology, which is essential for understanding the segmentation components of breast tissue presented in this chapter. A description of the screening methods gives support for understanding the nature of the images to be analysed. The study covers the general characteristics of a digital pathological image and its properties, obtained from a glass slide scanner, which is described in this chapter. The different imaging solutions for breast diagnosis are examined, and a special interest is in histology images. The relevance of segmentation is stated, and an overview of the segmentation methods is explored. This review starts from the traditional methods until deep learning techniques. This Chapter, helps to understand the general background on images and methods for BC analysis and segmentation.

### 2.1 Breast Cancer

#### 2.1.1 Breast tissue

Normal breast tissue is categorized into two major regions, the epithelial compartment and the stroma or supportive connective tissue (Guinebretiere et al., 2005). The epithelial is the glandular tissue responsible for milk production and is formed by the lobular units and ducts. Lobules are clusters of milk-secreting cells composed of smaller units called alveoli. In histopathology images, lobules and ducts are round structures that show a single layer of epithelial cells. In Figure 2.1, the structure

of a breast can be seen, and its corresponding histological image. Ducts are tiny tubes that drain the milk produced by the lobules towards the nipple, and in the histological image appear as branching structures with a double epithelial layer. The stroma is the supportive connective tissue that surrounds and embeds the epithelial compartment, is composed of various elements including fibroblasts, which are cells that produce collagen fibers capable of providing structural support to the breast tissue. Adipose tissue is formed by fat cells that contribute to the size and shape of the breast. Blood vessels and lymphatics are the responsible to deliver nutrients and remove waste products (Pandya and Moore, 2011). The normal function of breast cells and structures are affected when cancer arises.



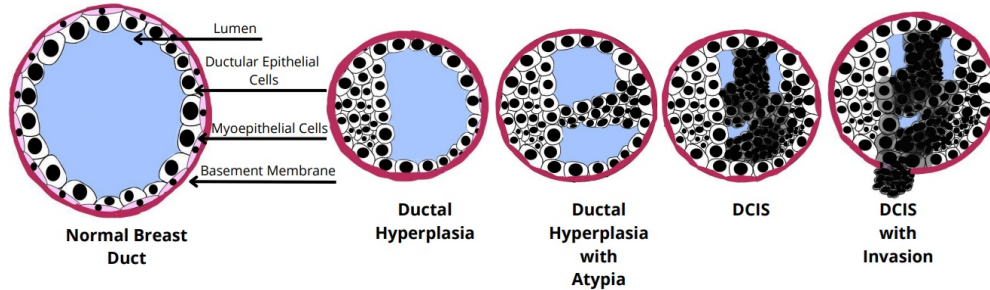
**Figure 2.1:** Description of breast tissue, figure (a) is the histopathology description of breast healthy tissue, image from commons wikimedia (de Bel et al., 2022), and (b) is the structure of the breast anatomy, the glandular tissue contains the milk producing lobules and the ducts. Image retrieved from (Zolfagharnasab et al., 2018), image available under Creative Commons License.

### 2.1.2 Breast cancer (BC)

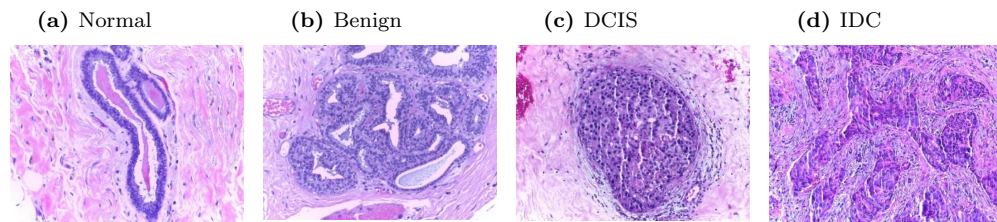
The type of BC can be identified through breast histopathology, Invasive Ductal Carcinoma (IDC) is the most common type, and Invasive Lobular Carcinoma (ILC) is the second most common invasive BC (Thomas et al., 2019). IDC starts in the milk ducts, and invades surrounding tissues; ILC starts in the milk lobules, and can spread to other tissues. The origin and evolution of IDC are shown in Figure 2.2, with a normal breast duct formed with a single layer of epithelial cells on the left, after an alteration, cells start to proliferate into a ductal hyperplasia, and a lesion is produced. After a greater proliferation, ductal carcinoma in situ (DCIS) is

formed from tumour cells confined in the duct. When proliferation extends to the surrounding tissue, there is IDC (Wang et al., 2024).

The main goal of this work is to develop efficient methods for segmenting the different regions of cancerous breast histopathology tissue.



**Figure 2.2:** A graphical description of the origin and evolution of ductal carcinoma, showing cancer evolution from left to right, image adapted from (Coleman, 2019).



**Figure 2.3:** Samples of H&E normal breast tissue in (a), a benign tumor in (b), DCIS in (c), and IDC in (d). The images are from ICIAR 2018 grand challenge (Campilho et al., 2018).

### 2.1.3 Breast Cancer Diagnosis

Cancer diagnosis involves various clinical tests and the observation of specific abnormalities, along with the collection of patient symptoms and clinical history (Grobstein, 2005a).

#### Self-examination

Breast cancer detection starts with regular breast self-examination, which is a simple method for detecting abnormalities, and it is done by a visual observation of both breasts and a gentle examination with the fingers to detect any irregularity or difference between breasts. The purpose of this self-test is to identify any abnormality but also to promote the self-awareness of BC in the patient (Iacob et al., 2024).

## **Mammography, ultrasound, and MRI tests**

The most common image modalities used for BC diagnosis are: mammography, ultrasound, Magnetic Resonant Imaging (MRI), Computer Tomography (CT), and Histopathology imaging (Jafari et al., 2018).

A regular clinical laboratory test or screening is recommended. Mammography is considered the gold standard for BC and is used for tumour detection (Sood et al., 2019). In this technique, a low X-ray dose is used to generate images of breast tissue. Mammograms show the internal structure of the tissue and radiologists can identify masses, and distortions associated with breast cancer. Mammography is usually used to detect cancer in an early stage when it is too small to be detected by self-touch and its relevance is that it can increase the survival rate (Champaign and Cederbom, 2000). The American Cancer Society (ACS) recommends mammography screening once a year after 45 years (Oeffinger et al., 2015), it is the best option for diagnosis, but cancer is not detected in the case of a dense breast, in which ultrasound can be used. Also, mammography is a cost-effective screening tool.

Ultrasound images are generated from high-frequency sound waves (which is why they are called ultrasound), used to describe different tissues and organs of the body (Aldrich, 2007). The waves travel inside the body, and those waves that are reflected, present a delay (or echo) used to create an image based on the type of tissue (Bartrum Jr, 1986). Ultrasound complements this screening process (Tadesse et al., 2023); it is used for the analysis of breast masses and the assessment of their density. Ultrasound is a complementary tool to mammography and is also recommended for young women under 40. It is important to mention that mammography is a radiation technique, and must not be applied to pregnant women, for whom ultrasound is preferred.

Magnetic Resonance Imaging (MRI) is a technique with high resolution and high sensitivity response when detecting IDC and DCIS (Harms, 1996; Mann et al., 2019). MRI is a time-consuming and expensive method that is used for high-risk patients and in preoperative cases. It is used when mammography is not conclusive, although is a more expensive technique. In the cases when abnormal mammography is observed, ultrasound, or MRI are recommended for a malignant tumour confirmation (Gao et al., 2020).

A comparison of screening methods for breast cancer diagnosis is presented in



**Figure 2.4:** An ultrasound image with a breast lesion, as presented by (Iacob et al., 2024)

table 2.1

	Advantages	Disadvantages
Mammography	Most common method	Not effective in dense breast
Ultrasound	detection in dense breasts	not detects misscalcifications
MRI	high sensitivity	expensive method

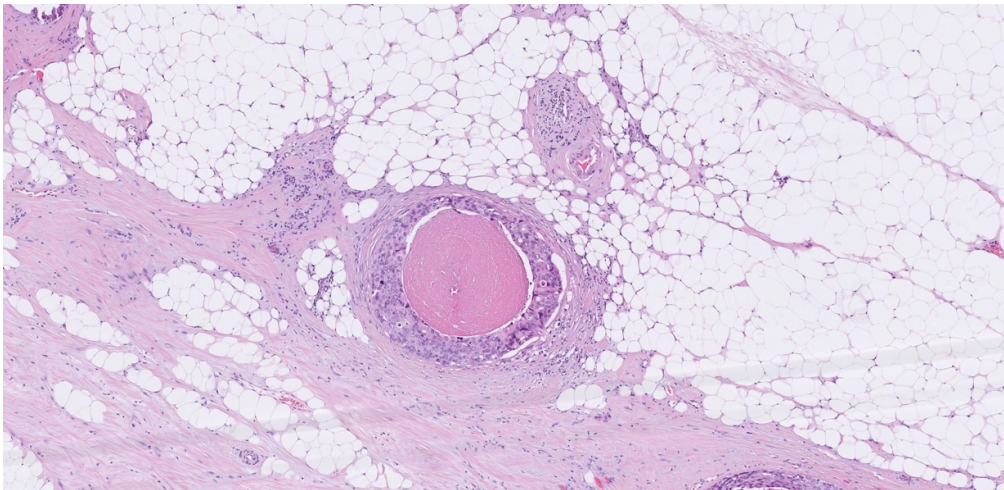
**Table 2.1:** Table description of the advantages and disadvantages between screening methods in breast cancer detection.

### 2.1.4 Histology

Histology is the study of tissues and their structure, when a disease is present, histopathology images from the affected tissue are analysed. The image corresponds to a tissue sample that is extracted and observed under the microscope. This method is used in many diagnosis procedures, for example in infections and inflammatory cases (Kradin et al., 2017). After a biopsy analysis, the pathologist presents a detailed report indicating if cancerous cells are present and the grading (Dey, 2022). A tissue slice is extracted and stained to highlight different tissue structures. A dye is applied to reveal the morphological features of the tissue observed under light microscopes. The two most common laboratory staining techniques used by pathologists to generate visual images are: Haematoxyline and Eosin (H&E) and Immunohistochemistry (IHC).

### Biopsy test: Haematoxilyne and Eosin (H&E)

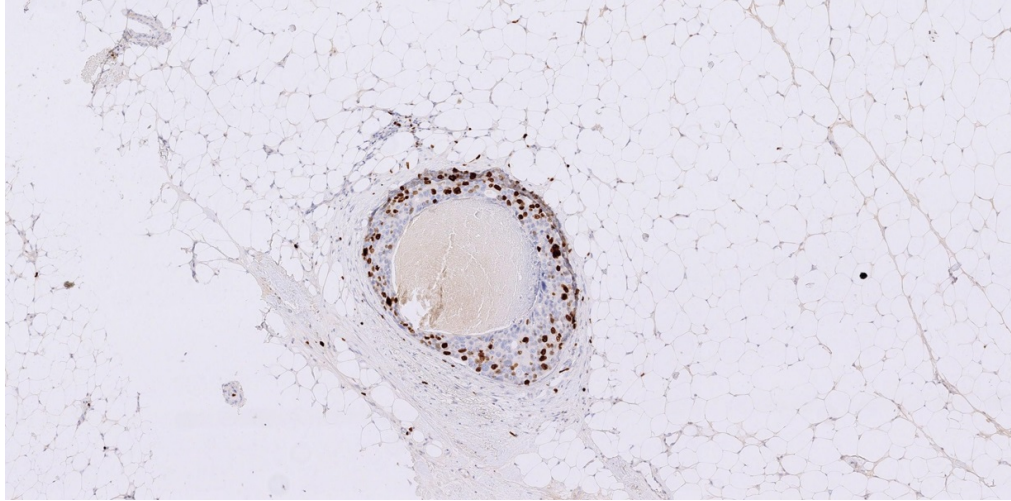
Haematoxilyne binds to acidic (basophilic) structures like the ADN in the nucleus of the cell, and it stains this in a blue colour. Eosin binds to basic (acidophilic) structures, and stains cytoplasm or stroma with a pink colour (Alturkistani et al., 2015). H&E staining is used to identify the morphological features of the tissue. The process for staining is the following, first the extracted tissue is preserved in formalin to prevent degradation. Next, thin tissue slices are cut (between  $3 - 5\mu m$ ) and embedded in paraffin wax. Paraffin is removed and the staining starts. Haematoxylin is applied which is responsible for staining the nuclei in a blue colour, and after a time, a rinse removes the staining. Finally, eosin is applied, to stain cytoplasm and extracellular components in pink. This technique reveals morphological features like the tumour region, which can be well identified. This technique is widely used because of its simplicity and reliability (Fischer et al., 2008). An example of an H&E histopathology image is presented in Figure 2.5.



**Figure 2.5:** Example of an H&E stained image from the ACROBAT dataset (Weitz et al., 2022). Pink colour corresponds to stroma and blue colour is for the cell nuclei.

### Immunohistochemistry

H&E staining confirms whether the tumor is malignant, while immunohistochemical tests provide additional information for classification and treatment guidance. Immunohistochemistry uses antibodies to detect specific proteins, such that the distribution of the proteins in the tissue section is revealed (Hussaini et al., 2022). In the image in Figure 2.6, we can see a sample image from the ACROBAT dataset (Weitz et al., 2022) for proliferation marker Ki-67 in (b).



**Figure 2.6:** Example of an IHC stained for Ki67 protein, image from the ACROBAT dataset (Weitz et al., 2022). It is interesting to observe that the image corresponds to the same tissue section as the previous example presented in Figure 2.5 from the same patient, a slice from biopsy is used for H&E staining and the next slice is used for IHC staining.

Examples of IHC biomarkers are estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2 (Maleki et al., 2013), and proliferation index Ki-67 (Bertrand et al., 2022). These are specific biomarkers that provide information on tumour behavior, and a guide for adequate therapy. In this type of test, diaminobenzidine (DAB) is the most common color substrate used, and it produces a brown stain (Prichard, 2014).

### **Estrogen(ER) and Progesterone (PGR) receptor**

Estrogen receptor (ER) is a biomarker that indicates that tumour growth is stimulated by estrogen hormone in the same way as the progesterone receptor (PGR) biomarker test. In these cases, the tumor is treated by hormone blocking therapies.

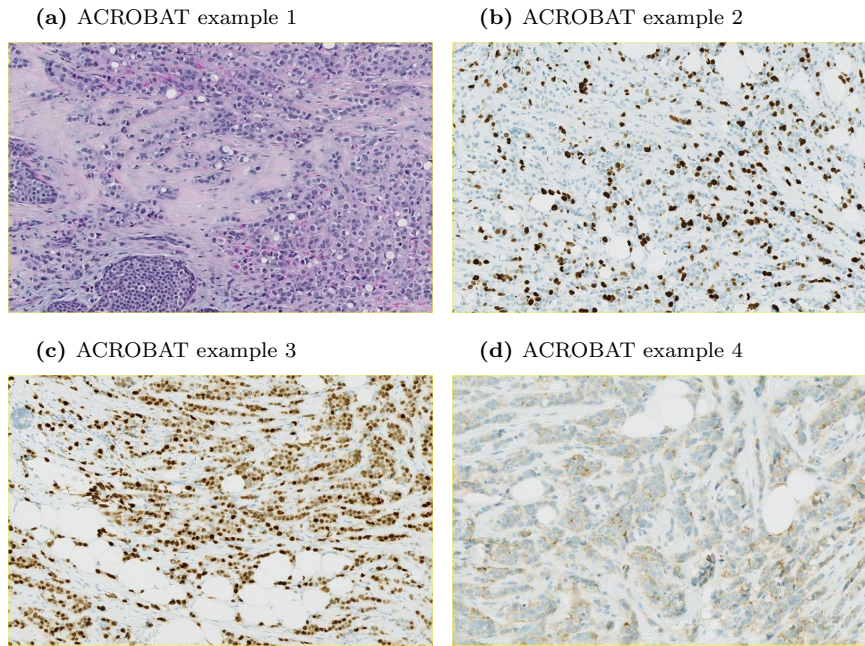
### **Proliferation marker Ki-67**

The proliferation marker or Ki-67 indicates the proliferation rate, or how fast the tumour grows. A higher index indicates a more aggressive tumour, which is usually treated with chemotherapy (Bertrand et al., 2022).

### **Human Epidermal Growth Factor Receptor 2**

Human Epidermal Growth Factor Receptor 2 or HER2 is also a immunohistochemistry (IHC) test that measures the amount of HER2 protein on the surface of the cancer cells, this over-expression of protein HER2 indicates cancer cell growth and

cell division (Vandenberghe et al., 2017). After the visual analysis by the pathologist HER2 is classified as HER2-negative or HER2-positive (in this case, when tumour is over-expressing HER2 protein). The different IHC biomarkers described can be seen in Figure 4.7, which is an image from ACROBAT dataset (Weitz et al., 2022). This dataset contains digitized samples from the different biomarkers.

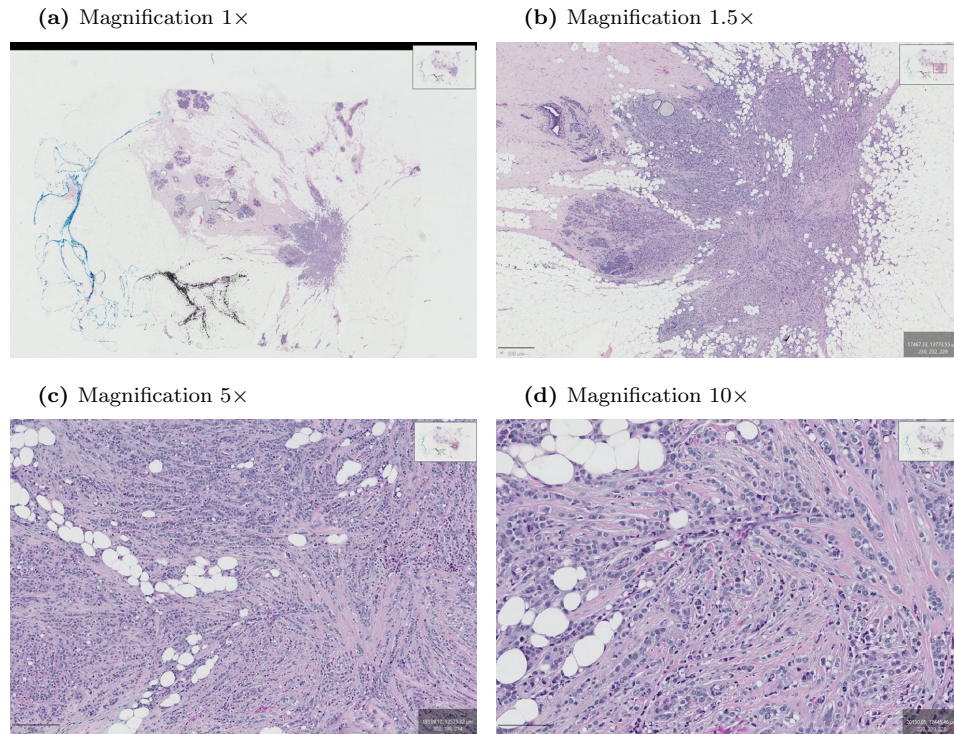


**Figure 2.7:** The ACROBAT dataset (Weitz et al., 2022) contains images from different laboratory tests: a)H&E image, b)IHC staining for Ki-67, c) IHC for progesterone and d)HER2 response. All the different images correspond to the same patient.

### 2.1.5 Whole Slide Imaging

Whole Slide Imaging, or WSI is an image obtained by a specialized scanner that digitizes a glass slide into a high-resolution digital image (Veta et al., 2014). The scanner is a specialized device with a mechanism for digitizing several glass slide samples, it takes a time to scan from 30 seconds to several minutes. The scanner is a trinocular microscope with electronic control of illumination intensity, a mechanical system, objectives, and coarse and fine focusing, and is equipped with a high-resolution camera. The scanner is then connected to a computer with specialized software for visualizing the image and sometimes with some image-processing algorithms already integrated. WSI scanners are now widely used by the pathologists (Humphries et al., 2024), however, this technology demands high computer resources in terms of processor and memory. A high-resolution scanned image is

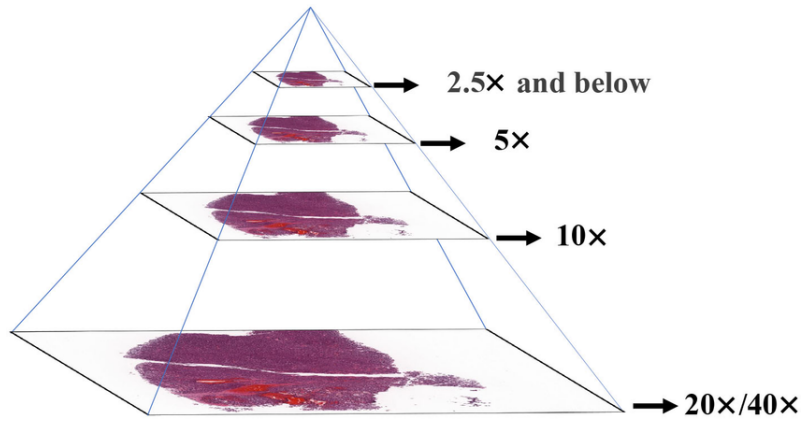
up to  $40,000 \times 60,000$  pixels (Amgad et al., 2019). The software visualization of the WSI also includes tools for adding annotations by the pathologist (Bankhead et al., 2017). Images of this type were used in this research. Images of BC from the ACROBAT dataset (Weitz et al., 2022) at different magnifications are shown in Figure 2.8.



**Figure 2.8:** Image of an H&E WSI sample taken from the ACROBAT dataset (Weitz et al., 2022), the image is visualized and magnified by Q-Path software (Bankhead et al., 2017), (a) is the WSI image, and images (b-d) are at magnifications of 1.5 $\times$ , 5 $\times$ , and 10 $\times$ .

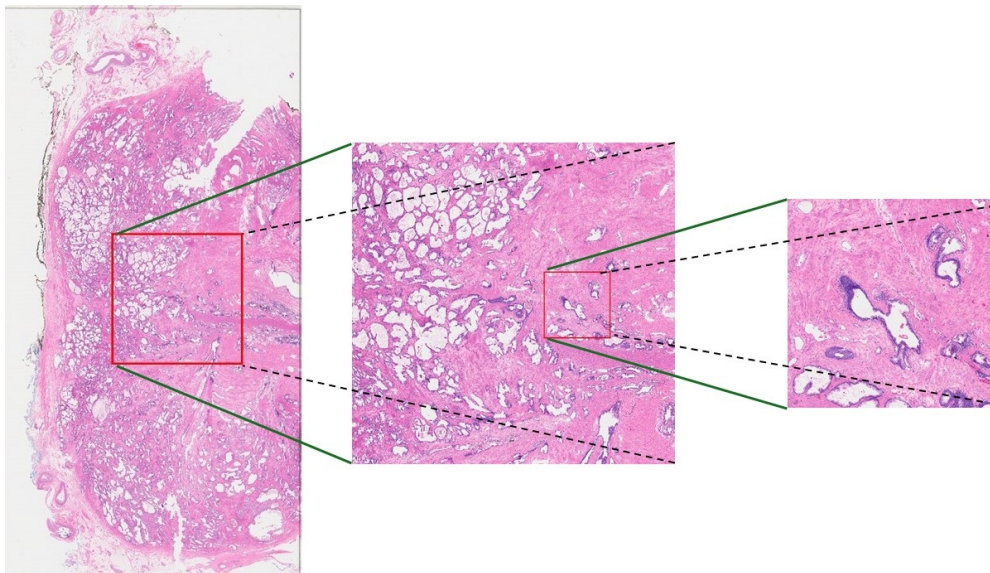
WSI are created at different resolutions. The image is organized in a pyramidal structure. In the coarse resolution level, users can identify large regions, and the entire glass slide is seen as a thumbnail view, this is at the top of the pyramid. At the top magnifications are at 5 $\times$  or lower. At the bottom of the pyramid is the finest resolution-level, usually at magnifications of 20 $\times$  or 40 $\times$ , where the finest details of the image are observed. Figure 2.9 shows this representation in a pyramidal view. The pyramid is generated through a downsampling process, at each level the image is downsampled by 2, which generates an efficient storage and retrieval process. The user can access the low resolution and after selecting an ROI the high resolution of the selected ROI is loaded efficiently without loading the entire full-resolution image.

Several software platforms are available to manipulate WSI, Q-path (Bankhead



**Figure 2.9:** Pyramidal structure of a WSI, the top of the pyramid is a coarse view and the bottom is the finest resolution, image is from (Rasoolijaberi et al., 2022), under creative Commons License CC BY-NC-ND 4.0.

et al., 2017), is an open-source software for digital pathology that offers annotation tools, cell detection, and classification; OpenSlide (Goode et al., 2013) is a library available for custom development; SlideViewer is a proprietary software for visualization and annotation; MATLAB has also specialized functions for opening and manipulating WSI for academic and research purposes (MATLAB, 2023). In this work MATLAB was used for manipulation the WSI, Figure 2.10 shows a WSI from the AGGC dataset, of size,  $(87120 \times 47040)$ .



**Figure 2.10:** WSI sample image from the AGGC Challenge dataset. This image was opened and processed by special MATLAB (MATLAB, 2023) functions for WSI manipulation.

## 2.2 Digital Pathology in Breast Cancer

Manipulating a digital image, represented as a matrix, alters the original pixel values. This process can enhance the image, detect object edges, or encode the entire image. Computer vision is a subfield of Artificial Intelligence (AI). Originally, computer vision methods relied solely on digital image processing techniques, often referred to as traditional handcrafted methods. However, modern AI-based techniques, especially those using neural networks, have become more popular due to their superior performance compared to traditional approaches.

### 2.2.1 Segmentation of BC images

An accurate detection of the tumour boundaries is a crucial task for the diagnosis process, and is required by the different image modalities (mammography, ultrasound, MRI or histopathology slices) (Michael et al., 2021b). Image segmentation consists of assigning a label or class to each image element. Semantic segmentation means that every pixel has a class or label assigned (Michael et al., 2021b; Irshad et al., 2013). Several segmentation approaches have been proposed over the years. At the early years of digital pathology, segmentation was centered on the nuclei segmentation. In addition, other tasks have attracted attention, like the differentiation of various types of tumours, IDC from DCIS, tumour delineation, and other structures present in histopathological images (Irshad et al., 2013), for instance, segmentation of epithelial tissue, stroma, lymphocytes, blood vessels, necrotic regions, and other histological features (Amgad et al., 2019). Therefore, this is called multi-class segmentation, and is critical for comprehensive histopathological analysis, enabling pathologists to assess tissue composition, tumour micro-environment, and disease progression (Song et al., 2023; Michael et al., 2021b). As mentioned, tumour size quantification, shape, and spatial distribution are useful in cancer diagnosis, staging, and treatment planning (Song et al., 2023). The Present research is centered on multi-class segmentation for BC tissue.

### 2.2.2 Digital pathology methods evolution

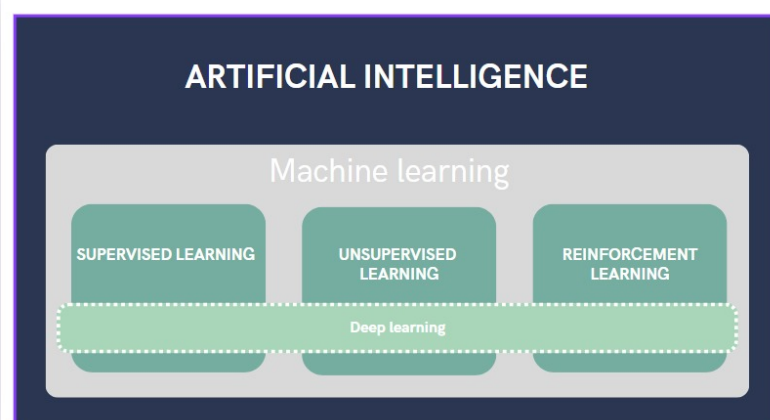
#### Traditional Methods

The first stage of digital pathology was in the 80's-90's with the analysis of images obtained from the microscope by the first digital cameras. Image processing opera-

tions were used for the analysis of these images, some examples are: thresholding, clustering, texture-based methods, and morphological operations. Wide Slide Image (WSI) scanners were first introduced in 1999 and since then, they have been widely adopted by the pathologist community. A WSI scanner is a device capable of digitizing an entire tissue slide, and a high-resolution image can be displayed on a computer screen. With the use of specialized software, the image is visualized as it is seen on the microscope.

## Machine Learning

Important progress in computer pathology was achieved with the implementation of Artificial Intelligence (AI) methods. Machine Learning (ML), is a type of AI algorithm that improves and learns based on the features and insights of the data. Therefore, ML is a subset of AI and can be understood as a software system that simulates intelligent behavior. The subsets of AI are seen in Figure 2.11.



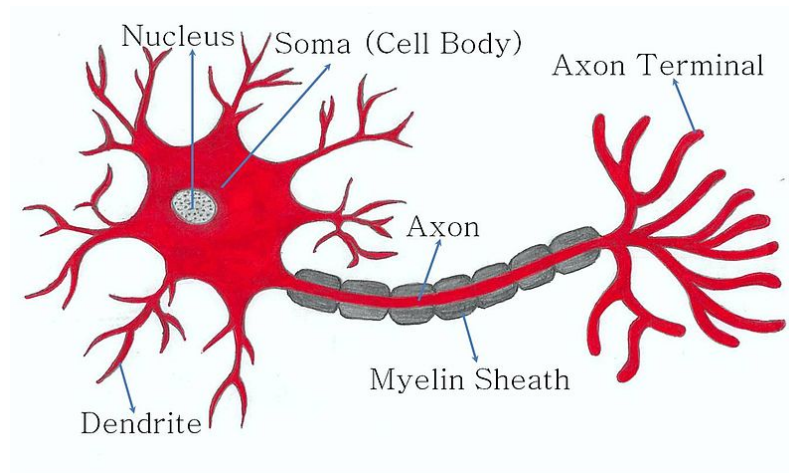
**Figure 2.11:** Machine learning is a subset of artificial intelligence and refers to those computer methods with the property of learning from the data itself, Deep-Learning is a subset of Machine Learning. Image adapted from (Prince, 2023).

In supervised learning models, the output is known and the model defines a mapping from the input data to the output. A relevant supervised method for the classification task is the Support Vector Machine (SVM), which is based on finding an optimal hyperplane that separates the different classes. It is a well-accepted method within the digital pathology workflow. Unsupervised learning is a model constructed from the input without any known of the output, this means no supervision. The model must understand the structure of the data and uncovers the insights of data without labeling it. K-means is a classification method used

to group data into clusters of similar characteristics, it can be used for segmenting, counting, and basic tissue classification tasks. Reinforcement learning is an ML method for a decision-making process that is trained to learn the optimal behavior (Rashidi et al., 2019). ML has been extensively used in digital pathology. Some of the relevant methods are Decision Tree (Karalis, 2020), Linear Discriminant (Li and Wang, 2014), Gaussian Naive Bayes (Bansal et al., 2022), Linear SVM (Cortes, 1995), Fine KNN, SVM Kernel, Boosted Trees (Elith et al., 2008), and Logistic Regression (Cox, 1958). These ML methods have been jointly used with computer vision methods to classify cancerous tissue. Computer vision techniques extract the image features, and next, ML performs the classification task. Examples of features are: intensity, colour, texture, and shape morphology. Because of the way features are extracted, they are also called hand-craft methods.

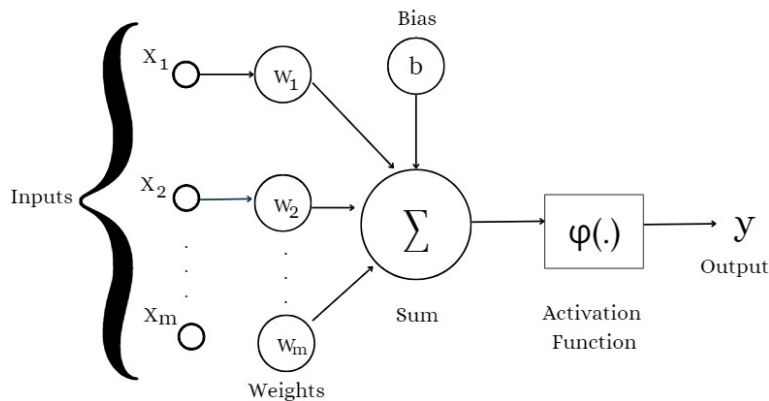
## Deep Learning

Deep learning (DL) is a supervised method based on neural networks with multiple layers that outperforms traditional methods. DL requires large amounts of data for training. An artificial neural network or ANN is a machine learning algorithm inspired by the human brain. Our brain is a collection of millions of neurons as the one represented in Figure 2.12. Each neuron receives electrical and chemical signals and transmits its output through a dendrite to another neuron at a junction named synapse.



**Figure 2.12:** A biological neuron is formed by a cell body connected to another cell by a dendrite, the joint connection is the axon, image under creative common licence, source: commons wikimedia.

An ANN is a computational model inspired by the behavior of the biological



**Figure 2.13:** An artificial neuron emulates the biological neuron,  $x_i$  are the inputs,  $w_i$  are weights and  $\varphi(\cdot)$  is the activation layer. Image adapted from (Prince, 2023).

neuron. Figure 2.13 shows an example of an artificial neuron. A collection of multiple artificial neurons can be connected and each neuron processes the input to obtain an output to the next one. A model based on multiple neural network layers is called Deep Neural Networks. ANN started being studied many years ago, however, it has attracted great interest recently due to its potential in real problem-solving. This is due to modern computers capable of training multiple neural networks.

### 2.2.3 Deep Learning Evolution

The first artificial Neural Network was introduced by McCulloch and Pitts (Palm, 1986) in 1943 by a model that processed a binary input to generate an output. Rosenblatt (Rosenblatt, 1958) developed the perception, an electronic device with a certain threshold that linearly combines inputs to generate a no/yes decision outcome. Minsky and Papert (Minsky and Papert, 1969) demonstrated the limitations of a two-layer perceptron, which cannot handle the OR exclusive problem, however extra hidden layers with non-linear activation functions can handle more complex input/output systems. The common activation functions used in a neural model were sigmoid and tanh functions. The Rectified Linear Unit, or ReLU function, was later used by Fukushima who trained a multilayer network for hand-written characters recognition (Fukushima, 1975). An important achievement in the ANN evolution was the backpropagation algorithm presented by Paul Werbos, in his PhD thesis. Werbos formalized the backpropagation algorithm to compute the gradients in a neural network. Backpropagation was rediscovered and refined by Rumelhart, (Rumelhart et al., 1986) along with Geoffrey Hinton, who demonstrated its practical

application in training a multilayer neural network. This period is considered the second generation of NN. However, there were several limitations in training a network by that time, only shallow networks were well trained and not deeper networks. For that reason, NN were not well accepted, mainly by the Machine Learning community, and were temporarily forgotten (Kurenkov, 2020). Hinton (Hinton et al., 2006) used for the first time the term *Deep-Learning* to the idea of a neural network with many layers, and presented a clever way to train a deep neural network, this contribution represented a re-born of neural networks. The Convolutional Neural Network (CNN) is based on the use of digital filters to learn features from data (Lecun et al., 1998), it has been successfully used for the image analysis problem. During the 2012 ImageNet Large Scale Visual Recognition (ILSVRC) Challenge, Krizhevsky (Krizhevsky et al., 2012) introduced the network called AlexNet which demonstrated outstanding performance over the rest of its competitors. This event is considered a turning point in the evolution of neural networks and artificial intelligence. The factors that allowed this event, were larger training datasets, improved processing power by use of computer graphics units or GPU's for training, the use of the ReLU activation function, and stochastic gradient descent algorithm. In summary, what modern deep-learning is nowadays. Since 2012, more than a decade ago, the impact of Deep Neural Networks is being seen in the solution of real life problems like image classification, segmentation, regression.

In 2025, Goeffry Hinton was awarded with the Nobel Prize in Physics 2024, as stated by the Nobel Committee *“for foundational discoveries and inventions that enable machine learning with artificial neural networks”*.

## 2.3 Summary

This chapter presents a general overview of computational pathology for breast cancer (BC). For a better understanding of the problem, a description of the breast anatomy and BC disease, including the most relevant diagnosis tests, are also presented. Tumour segmentation of BC images is a pertinent task in cancer diagnosis and is addressed in this research work. A general overview of BC image segmentation is presented in this chapter. It described digital pathology on BC image analysis. Nuclei segmentation was first addressed, also, tumour segmentation, and more recently multi-class segmentation for all the different elements visible in the

tissue. Finally, the evolution of segmentation methods is discussed, from traditional image processing techniques to modern deep learning approaches.

## Chapter 3

# Related Work

In this section, the related work on BC image analysis and segmentation is presented. Several works have been published on breast cancer image analysis, which have evolved from hand-engineering feature extraction to a deep learning approach; the works presented in this chapter also show this evolution. Although much work has been published on the analysis of mammograms, ultrasound, and MRI, as summarized by (Michael et al., 2021a), the overview presented in this chapter is oriented to the most relevant studies published on immunostained and H&E tissue images.

Early works performed on immunostained BC images were done on the mitotic count of cells. The International Conference of Pattern Recognition (ICPR) (Heyden et al., 2014) introduced the Mitosis and Atypia contest for mitosis score and atypia grading from 2012 to 2014. Mitosis detection is also an important indicator of BC aggressiveness (Heyden et al., 2014), although this is not a segmenting task, several contributions were made within this challenge for the BC computational pathology. It is interesting to note that in the 2012 contest, several works were still based on hand-craft feature extraction and classification algorithms (Ludovic et al., 2013).

The work presented by (Dong et al., 2014) aims to classify tissue into DCIS and UDH (Usual Ductal Hyperplasia). It is based on the extraction of a high-dimensional vector of features, mainly from the nuclei, and it performs a tissue sample classification task based on machine learning methods. This method is not fully automated because the samples need to be manually selected by a specialist containing a tumour area on the image itself.

The detection of IDC was addressed by (Cruz-Roa et al., 2014), which was one of the early works based on a deep CNN for the automated detection of IDC. The

model was built with a CNN, based on a two-layer network, and a logistic regression classifier to predict if an image patch belongs to an IDC tissue or not. It is interesting to mention that the delineation performed by pathologists was not very precise, which the author states is the main cause of the misclassification of tissue cases.

A more complex network was proposed by (Chen et al., 2016), which combined two CNNs in cascade, which is an example of a deeper architecture by increasing the size of a single convolutional layer. This approach outperformed the accuracy of other methods by a large margin within the mytosis score and atypia grading challenge.

(Janowczyk and Madabhushi, 2016) presented a comprehensive tutorial with seven DL tasks, some of which are for BC: IDC classification, nuclei segmentation, epithelium, stroma, and tubule segmentation. The authors explored segmentation, classification, and localization tasks, and employed Alexnet network (Krizhevsky et al., 2012), which was specifically trained for the selected tasks.

Several works were published for the detection of metastases in H&E images under the Camelyon 16 challenge, organized by the IEEE Internet Symposium on Biomedical Imaging (ISBI) (Bejnordi et al., 2017). Two tasks were addressed: tumour localization and classification. In the classification task, some of the works also performed the segmentation of metastases regions. The most relevant works were based on DNN implementation. For example, the work by (Wang et al., 2016), which was the winner of the contest, evaluated the performance of VGG16, GoogLeNet, and AlexNet, with the highest accuracy obtained by the GoogLeNet model.

The work by (Mehta et al., 2018) for the detection of benign, atypia, DCIS, and IDC (Elmore et al., 2015) of H&E images was based on a modification of UNet (Ronneberger et al., 2015) for joint segmentation and classification tasks. A special dataset was prepared for this research.

The determination of the different BC categories was explored over the BREAKHIS challenge (Pereira, 2023). It is formed of H&E images acquired in magnifications of  $40\times$ ,  $100\times$ ,  $200\times$ , and  $400\times$  images for the classification of benign or malignant tissue and as well as to determine different categories: adenosis (A), fibroadenoma (F), phyllodes (PT), tubular adenoma (TA), ductal Carcinoma (DC), lobular Carcinoma (LC), mucinous carcinoma (MC) and papillary carcinoma (PC). The main goal is a multi-category classification task. Several results were published based on this dataset, which also showed the transition from traditional models (based on

handcraft feature extraction) to DL models like DCNN (Spanhol et al., 2016; Yari et al., 2020; Benhammou et al., 2020).

The task of localizing the tumour area in a WSI image was included in the automated workflow within the Grand Challenge on Breast Cancer Histology images (BACH) organized with the 15th International Conference on Image Analysis and Recognition (ICIAR 2018). It comprises 400 training and 100 WSI test images that contain annotations by pathologists for benign, in Situ Carcinoma, and invasive Carcinoma (Aresta et al., 2019a). The data is divided into two parts: part A contains images for a classification task into the 4 categories, while part B is used for a rough segmentation of a scaled version of the WSI into the same 4 classes categorized in a mask numerical image. Most of the relevant works were done by a pre-training of models like ImageNet (Russakovsky et al., 2015), VGG (Balasubramanian et al., 2024), ResNet, and Densenet (Aresta et al., 2019b). These networks were fine-tuned, which means that the classification layer of the original pre-trained network, usually a fully connected layer, is replaced by a new reclassification layer and trained for the medical images of interest. For the localization task over the WSI image, some models used UNet (Ronneberger et al., 2015).

The dataset from the BreastPathQ challenge organized jointly by the International Society for Optics and Photonics (SPIE), the American Association of Physicists in Medicine (AAPM), the U.S. National Cancer Institute (NCI), and the US Food and Drug Administration (FDA) provides a set of images to determine the tumour cellularity (TC) of breast cancer images (Peikari et al., 2017). This task does not request a tumour region delimitation but rather the tumour area estimation or TC as an important factor for neo-adjuvant (NAT) treatment (Akbar et al., 2019; Peikari et al., 2017). Methods based on feature extractions as well as deep learning were used for this task (Akbar et al., 2019; Ortega-Ruiz et al., 2020).

Most of the works presented so far have used pre-trained, well-known deep learning models. For example, the work by (Celik et al., 2020) used Resnet50, and Densenet 161 for IDC detection. Few studies have proposed new models or significant modifications over the pre-trained models. The work by (Bolhasani et al., 2020) presented a network named FusioNet, which is an autoencoder network based on residual convolution, and it was calibrated for IDC detection.

By the other hand, most of the works presented perform a classification task on the tissue image sample. A direct segmentation of a tumour region of BC images was

not addressed until 2019 by (Guo et al., 2019), whose work has the aim to delimitate the lesion area of the image. Semantic segmentation implies the class assignment at a pixel level; previously described works were trained for a classification task and not for segmentation. This difference is also explained by (Guo et al., 2019). In his work, segmentation is performed in two steps: first, a classification network by an Inception V3 (Szegedy et al., 2016), and next, a DCNN was used for a refined segmentation. A heatmap likelihood image at  $\frac{1}{8}$  of the original WSI image is obtained in 11.5 minutes. Due to the lack of image sets for the segmentation task, (Priego-Torres et al., 2020) created a dataset for direct tumour segmentation. A web-based platform was implemented for pathologists to visualize and make annotations on WSI images, and a DCNN encoder-decoder with Atrous convolution, i.e., convolution with dilated kernel architecture, was implemented. The goal was to determine the type of tumour: ductal carcinoma in situ, invasive ductal carcinoma, carcinoma tubular in situ, or invasive tubular carcinoma (Priego-Torres et al., 2020).

Finally, there are few works reported on multi-class segmentation of H& E images. For example, a non BC task was the head and neck tumour segmentation (HEKTOR) 2022 Challenge (Andrzejczyk et al., 2021) which is the segmentation of head and neck (H & N) tumours and lymph nodes. The work done by (Liu et al., 2022) proposed a 3D UNet network with an attention module, and (Yuan, 2021) presented an encoder-decoder network similar to UNet for performing the segmentation of primary tumours and metastatic lymph nodes. Another multi-class segmentation example for the delineation of the different breast tissue elements was presented at the Breast Cancer Semantic Segmentation Challenge (BCSS) by (Amgad et al., 2019). Within this challenge, the reference data consists of 151 WSI images with the delineation of 22 different regions. Due to the lack of labeled data, this dataset is a valuable support for BC segmentation research. The authors also present results from a segmentation model based on a VGG-16 network, which are used as a baseline reference. This data was used in the present work and will be explained in detail in Chapter 4.

In summary, this review highlights that there is still significant work to be done to achieve accurate tumor segmentation directly from immunostained and H&E-stained cancer images, particularly in the context of breast cancer (Guo et al., 2019). While numerous studies have addressed tumor area segmentation in mammogram images, Guo et al.'s work is among the earliest efforts in this direction, and progress has

been limited mainly due to the lack of annotated reference images. Furthermore, segmentation of additional breast tissue components remains an underexplored area (Priego-Torres et al., 2020).

The BCSS dataset introduced by (Amgad et al., 2019) offers a valuable resource for training and evaluating models for this particular task, enabling segmentation of various breast tissue structures, something previously not fully addressed. Additionally, many existing works reviewed in the literature rely on pre-trained, well-established models. There is a clear need to explore different architectures and modifications to improve performance (Spanhol et al., 2016).

Therefore, this research addresses both of these challenges: improving segmentation of multiple breast tissue components and exploring novel or adapted model architectures. Moreover, the methods reviewed in this section reflect the broader evolution of the field—from traditional image processing approaches to modern machine learning (ML) and deep learning (DL) methodologies.



## Chapter 4

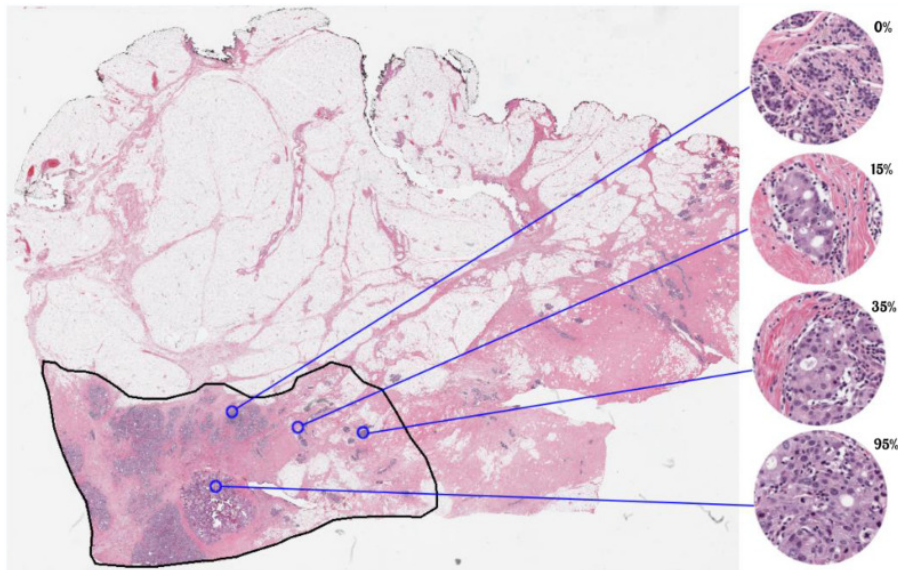
# Materials

This chapter presents a description of the images and datasets used in this research. Most of the images were sourced from selected challenges focused on breast cancer, along with one dataset related to prostate cancer grading. The datasets include: BreastPathQ, a collection of histopathology images for cellularity estimation; Breast Cancer Semantic Segmentation dataset (BCSS), which provides images along with corresponding labels for the segmentation of various tissue components; Automated Gleason Grading Challenge (AGGC), used for assessing prostate cancer severity; and ACROBAT, which includes multiple breast imaging modalities. Each of these datasets is explained in detail in this chapter.

### 4.1 SPIE-AAPM-NCI BreastPathQ challenge dataset

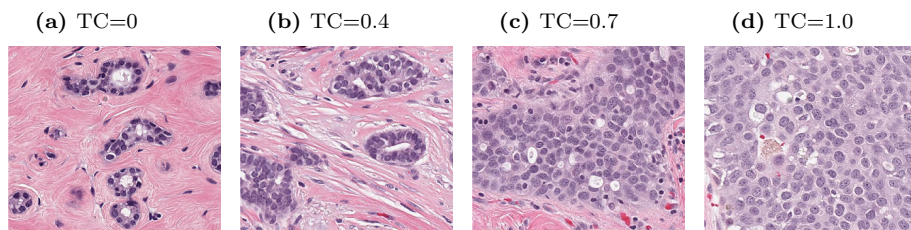
The SPIE-AAPM-NCI BreastPathQ challenge dataset was presented in 2019 by the International Society for Optics and Photonics (SPIE), along with the American Association of Physicists in Medicine (AAPM), and the National Cancer Institute (NCI) in the Grand Challenge forum (Peikari et al., 2017). It is a set of 96 breast cancer hematoxylin and eosin (H&E) stained pathological slides from 64 patients with residual invasive Breast Cancer under Neoadjuvant Treatment. Each slide was scanned by magnification of 20X using an an Aperio AT Turbo 1757 scanner (Leica Biosystems Inc.). The dataset was collected at the Sunnybrook Health Sciences Center, Toronto, Canada. The dataset is divided into the training, validation, and test datasets. From each slide, patches of size  $512 \times 512$  pixels from a region of interest (ROI) were selected and extracted as uncompressed TIFF images. For each patch, the Tumor Cellularity (TC) rate was assigned by the pathologist. Seeden

Viewer platform was used to visualize each slide by selecting a specific point in the image and a square of  $512 \times 512$  pixels was extracted at each specific point as seen in Figure 4.1.



**Figure 4.1:** After visualizing the full WSI, patch sections were extracted by the pathologist for selected points of interest. After a point is selected a  $512 \times 512$  image centered at that point is extracted. Image taken from (Peikari et al., 2017), used with permission of the authors.

A total amount of 2394, 185, and 1119 patches for training, validation, and testing were obtained and provided for the challenge. Some representative samples are seen in Figure 4.2. We can observe images from low to high TC. The TC values assigned to each path have values from 0.1, 0.2,  $\dots$ , 1.0.



**Figure 4.2:** Four images with cellularity values of 0, 0.4, 0.7, and 1, where the prevalence of cancerous cells can be observed.

The challenge was held in November-December 2018, in which participants submitted algorithms and results for the test set. The submission consists in a .csv file with predictions for the TC assignment for each patch image. The Ground Truth was provided for the training and validation sets only, and not for the test set. After December 2018 the challenge was still open for submissions in an ongoing phase. There were 317 participants in the initial phase. The submissions were measured by

the Intra-Class Correlation coefficient (ICC)

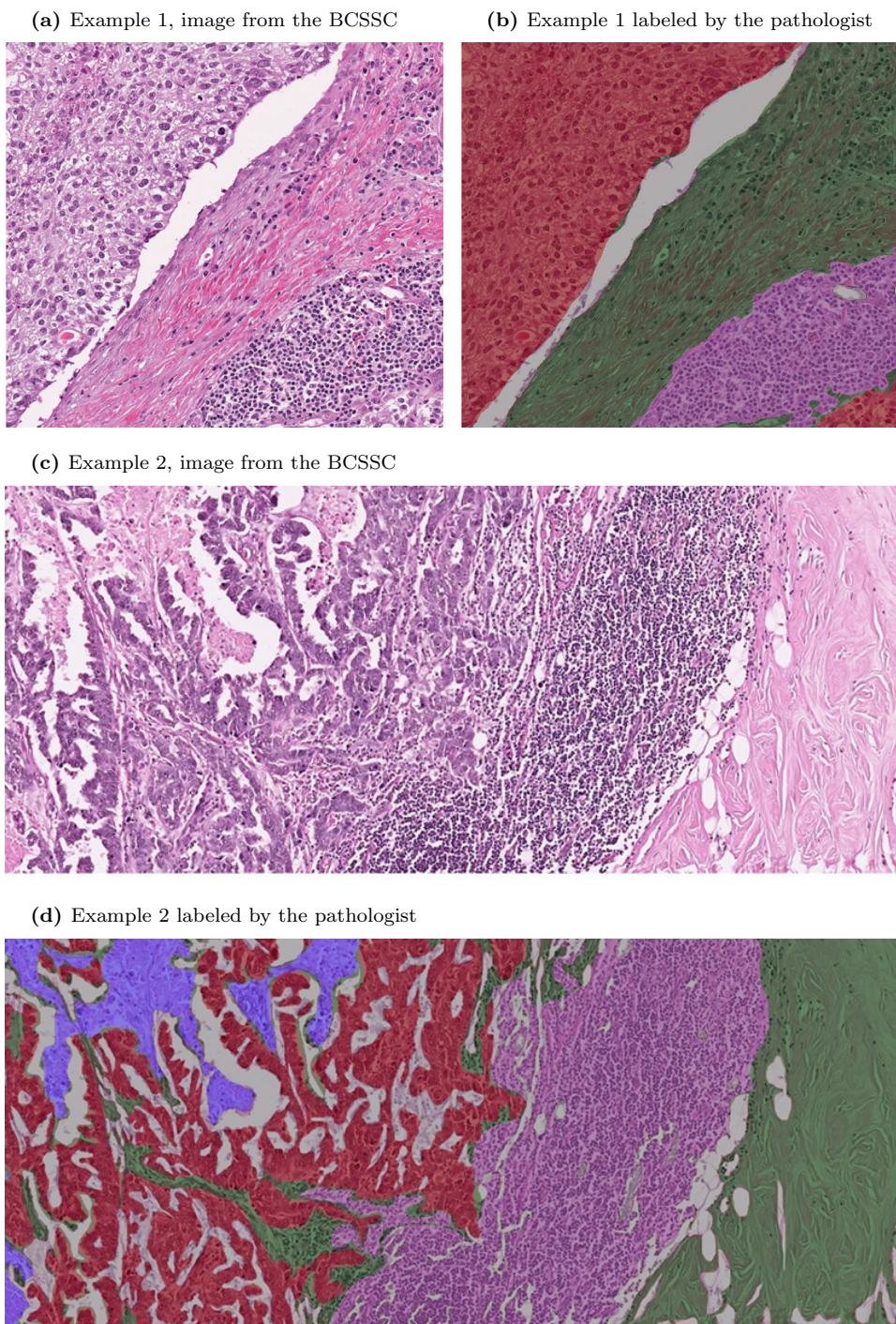
## 4.2 Breast Cancer Semantic Segmentation (BCSS) dataset

The main goal of this dataset is to provide valuable material for training algorithms for H&E breast cancer image segmentation of different tissue regions (Amgad et al., 2019). The images within this dataset are 151 WSI from the Cancer Genome Atlas. The images are tissue samples stained with H&E of different sizes and resolutions. Each image has a representative ROI selected by a doctor and used to do the annotations. Images were manually annotated by pathologists, pathology residents, and medical students, in a crowd-sourcing process as described by (Amgad et al., 2019), and more than 20,000 segmentation annotations are provided. Structured crowd-sourcing refers to a process of assigning tasks based on the experience of participants. Each participant was asked to annotate 5 or 6 different ROI's and the more challenging slides were transferred to the more experienced participants. The different regions were carefully delineated by the participants. A total of 22 different classes were annotated: tumour, stroma, inflammatory, necrosis, glandular secretions, blood, fat, plasma cells, immune infiltrate, mucoid material, normal acinus or duct, lymphatics, nerve, skin, blood vessels, and undetermined. The labeled image is a mask image whether each pixel contains an integer value that corresponds to one of the 22 labeled classes. Samples of this dataset are shown in Figures 4.3 and 4.4, the original WSI image can be observed, and an overlaid image of the original image with the annotation mask using different colors.

The full dataset is presented in the Breast Cancer Semantic Segmentation (BCSS) Challenge

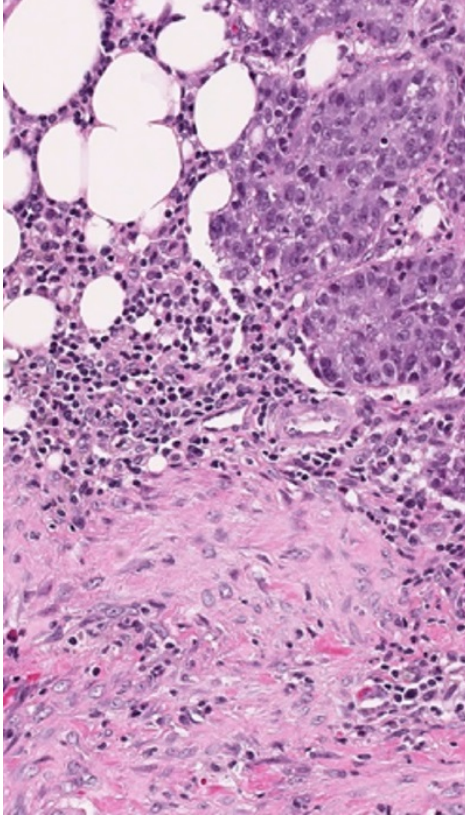
## 4.3 Automated Gleason Grading Challenge 2022 dataset

This dataset of H&E Whole Slide images consists of prostatectomy and biopsy specimens with annotations performed by experienced pathologists. Tissue specimens were collected from the Department of Pathology, at the National University of Singapore Hospital (NUHS) (Huo et al., 2024). The samples were prepared following the standard operating procedures of a CAP-accredited histopathology laboratory (Guo et al., 2019; Dey, 2022). The slides consist of  $4\mu\text{m}$  sections stained with haematoxylin

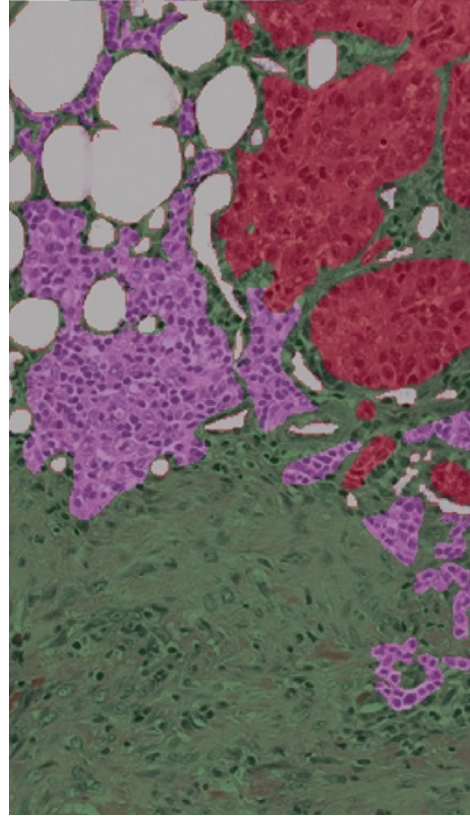


**Figure 4.3:** H&E images taken from the BCSS challenge dataset, (a) corresponds to the file TCGA-E2-A1AZ-DX1, and (b) is the overlaid image for the region annotated by specialists. The tissue regions selected in this work are tumour (red), stroma (green), inflammatory (purple), necrosis (blue), and other (gray). Sample TCGA-AO-A129-DX1 (c) from BCSS dataset and the corresponding annotation done by the pathologist is in (d). Notice that some regions are easier to distinguish, whereas stroma and necrosis are irregular regions and harder to distinguish

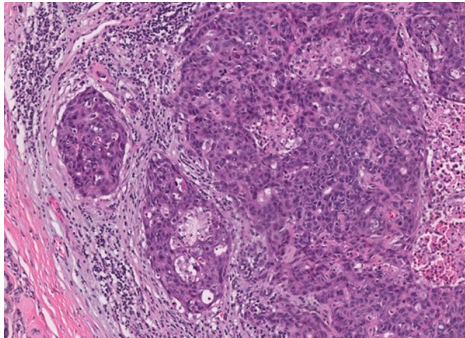
(a) Example 3, image from the BCSSC



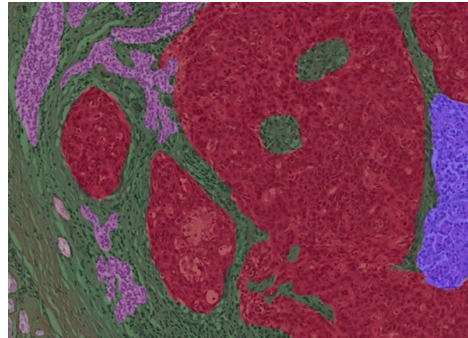
(b) Example 3 labeled by the pathologist



(c) Example 4, image from the BCSSC

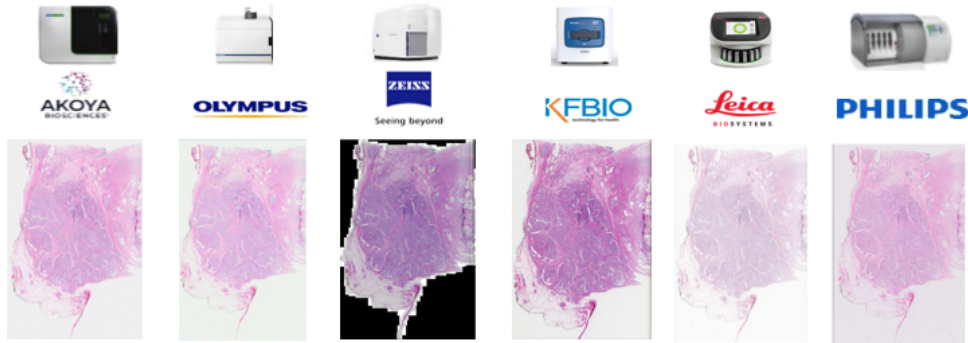


(d) Example 4 labeled by the pathologist



**Figure 4.4:** Figures (a) and (b) are the corresponding H&E and annotated images for file TCGA-E2-A1AZ-DX1, (c) and (d) are the images for the TCGA-S3-AA15-DX1 file from the BCSS dataset.

and eosin from tissue blocks of radical prostatectomy and biopsy specimens. There were 187 prostatectomy specimens corresponding to a tissue area of  $112,400\text{mm}^2$  and 156 biopsy specimens (tissue area  $7,723\text{mm}^2$ ) from 214 patients.

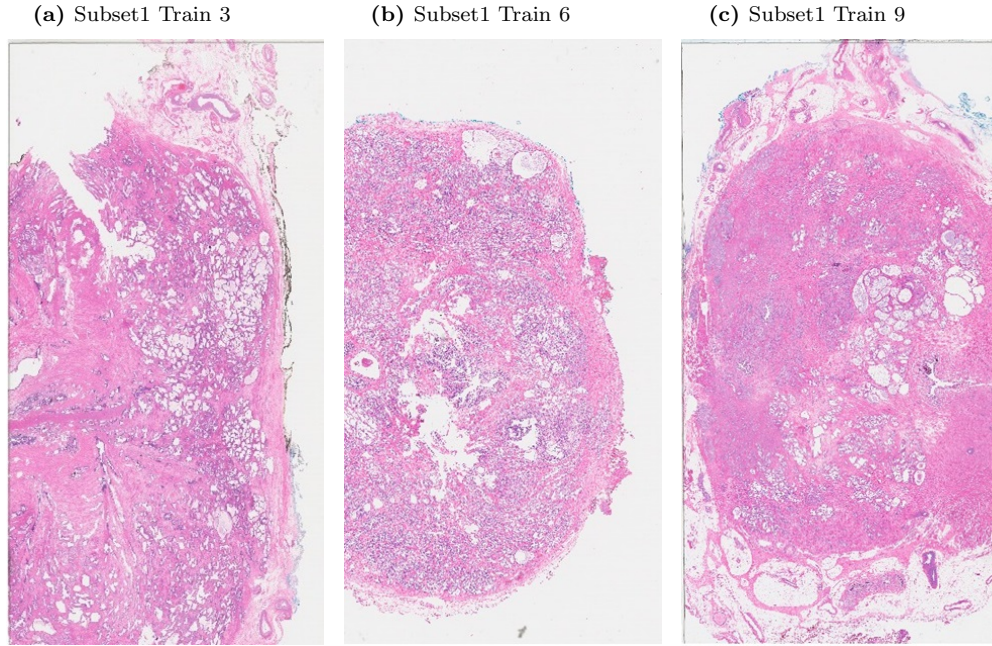


**Figure 4.5:** Scanners used within AGGC (Guo et al., 2019) and their corresponding scanned image is below. It can be seen differences in colour and light intensity, as well as the background in *Zeiss* scanner. Image retrieved from (Guo et al., 2019).

The images come from different types of scanners to analyse variations caused by image digitization. Initially, the images were acquired using an Akoya Biosciences scanner with bright-field imaging at  $0.5\mu\text{m} \times 0.5\mu\text{m}$  per pixel resolution. To analyse the effect of the scanning, 38 prostatectomy specimens were scanned at the same resolution by 5 other scanners (Olympus, KFBio, Zeiss, Leica, Philips). The dataset is divided into three subsets; subset 1 consists of 105 training and 45 test cases, subset 2 consists of 37 training and 16 test cases and, finally, subset 3 is composed by 144 training cases and 67 test cases, all of them scanned by multiple scanners: Akoya Biosciences, Olympus, Zeiss, Leica, KFBio and Philips. Figure 4.5 shows the scanners used for the generation of this dataset and Figure 4.6 shows three slide samples of the biopsy.

The annotations for GG are selected regions in the image with the categories: *Gleason Pattern 3*, *Gleason Pattern 4*, *Gleason Pattern 5*, *Normal*, and *Stroma* performed by the pathologists under a special platform called !AHistonotes (Huo et al., 2024). The annotations are binary masks corresponding to each class. It is important to remark that in some cases there are pixels assigned to more than one class, for example, there might be pixels with annotation for Gleason pattern 3 and Gleason pattern 4, in the experiment presented in section 5.3 from this research a selection criteria is explained for this cases.

The main objective of this challenge is to develop algorithms for Gleason grading by using the test images. Participants submitted the test images as PNG images



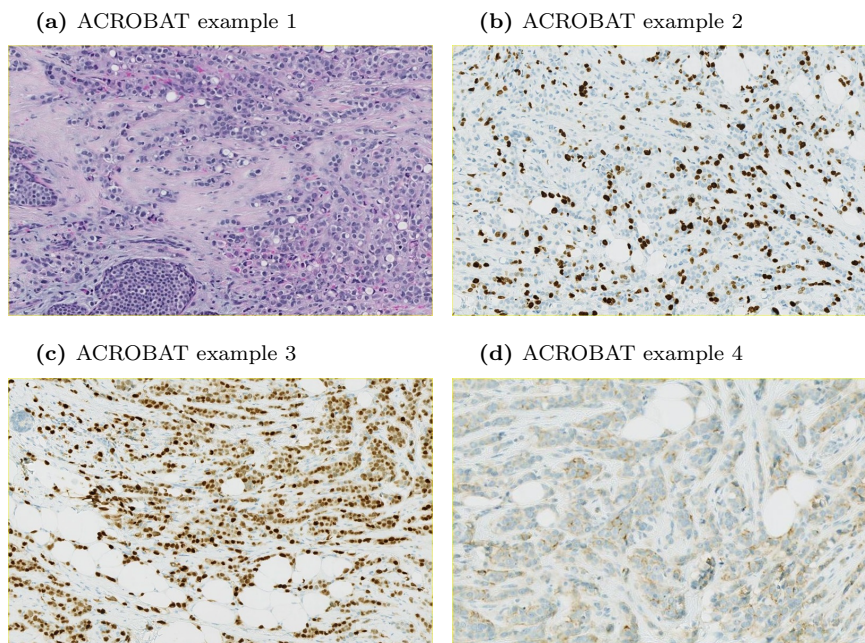
**Figure 4.6:** Three sample images from AGGC dataset. The visualization is done at the lowest resolution. The images correspond to the files a) Subset1 Train 3, b) Subset1 Train 6, and c) Subset1 Train 9.

with pixels assigned to numbers corresponding to each of the selected categories.

#### 4.4 Breast cancer ACROBAT dataset

The ACROBAT dataset consists of 42124 whole slide images (WSIs) from 1,153 patients. There are image sections stained with H&E staining and IHC (Weitz et al., 2022). The main objective of the dataset is to develop algorithms for image corregistration, which is the task of performing an alignment between H&E and IHC images taken from the same biopsy (Weitz et al., 2022). Registration is particularly challenging due to the gigapixel size of the images. The dataset also contains image tests for Ki-67, estrogen, progesterone, and HER2 response for each biopsy.

Although the main purpose of this data set is to develop WSI registration methods, the dataset has an enormous potential for research in the context of computational pathology. Figure 4.7 shows an example of all the images contained in the dataset, it is interesting to remark that all of the images displayed are for the same patient biopsy sample.



**Figure 4.7:** The ACROBAT dataset (Weitz et al., 2022) contains images from different staining, the H&E image is in (a), an IHC staining for Ki-67 is shown in (b), an IHC for progesterone (PGR) in (c) and the HER2 response is in (d). All the four images are from the same patient.

Feature	BreastPathQ	BCSS	AGGC22	ACROBAT
Type	Breast	Breast	Prostate	Breast
Modality	H&E	H&E	H&E and IHC	H&E
	stained WSIs	stained WSIs	stained WSIs	stained WSIs
Task	Tumour cellularity	Tissue segmentation	Classification	Grading (classification)
Resolution	40×	40×	40×	40×
No. of WSIs	151	257	800	435
License	research use	open (CC BY 4.0)	open (CC BY 4.0)	open (CC BY 4.0)

**Table 4.1:** Comparative table of main characteristics of the datasets used in this thesis research.

## 4.5 Summary

This chapter provides a general description of the materials used in this research, which include pathology images from breast cancer cases and biopsy tissue samples from prostate cancer. A comparison of the datasets is summarized in Table 4.1. As it can be seen from the table, most of the images employed in this research are whole-slide images (WSIs) acquired using high-resolution scanners. As a result, they are often large in size—on the order of gigabytes—and can be challenging to process and manipulate. It is important to note that the BreastPathQ and BCSS datasets were used in their entirety, while approximately 50% of the images from the AGGC dataset were utilized. In contrast, only a small number of images from the ACROBAT dataset were used, primarily for validating certain programming routines. Details regarding the slide scanners used in the AGGC dataset can be found in Appendix B.



## Chapter 5

# Tumour cellularity assessment

This chapter presents a comprehensive digital pathology example for Tumour Cellularity (TC) estimation, utilizing traditional computer vision and image processing methods. The approach includes Otsu’s binarization (Otsu, 1979), K-means clustering for colour separation (Ghosh and Liu, 2009), watershed segmentation (Meyer, 1994), and morphological analysis for feature extraction. Machine learning techniques are then applied to assess TC. This work follows the traditional digital pathology pipeline, as outlined by (Dong et al., 2014): colour separation, segmentation, feature extraction, and classification. The results were compared against a baseline established by (Peikari et al., 2017), and the outcomes of this research were published in a scientific journal (Ortega-Ruiz et al., 2020).

### 5.1 Overview

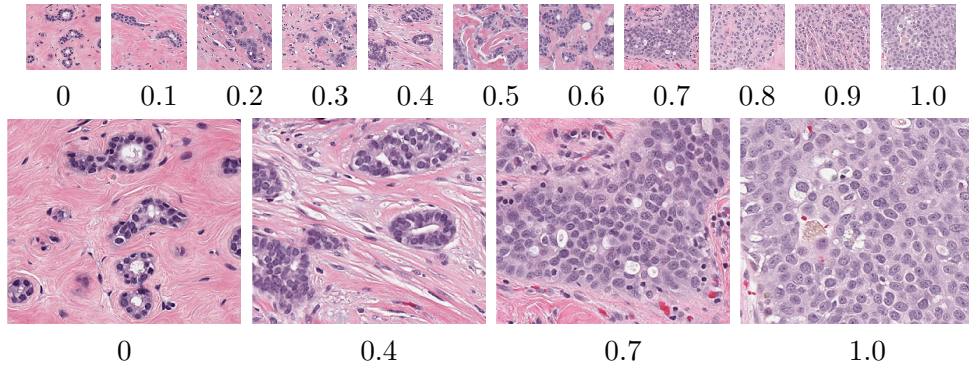
TC is a relevant parameter of breast cancer for patients under Neo-Adjuvant Treatment (NAT). NAT is a preliminary treatment administrated to the patient before the main treatment (Grobstein, 2005a), the response to NAT gives information on therapy efficacy; this response is measured by the residual cancer burden index composed of two metrics: TC and assessment of lymph nodes (Symmans et al., 2007). TC is given by the percentage of the tumour area over the total area of the image, thus, the tumour needs to be measured. Although this experiment is not a direct tumour segmentation task, our approach estimates an approximation of the tumour region in an intermediate step to determine the TC. This work presents a novel nuclei segmentation technique, as well as computer vision routines for the analysis of regional parameters, like features from breast duct. Therefore, the main

contributions of this experiment are: (1) an efficient nuclei segmentation algorithm based on the enhancement of nuclear structures, and (2) a method for breast duct detection grounded in set theory. This work relies on traditional image processing techniques, underscoring their continued relevance. The cellularity assessment results were validated using the Intraclass Correlation Coefficient (ICC) and compared with the results from (Peikari et al., 2017). While our method demonstrates improved performance over that baseline, it does not outperform results obtained using deep learning approaches; however, it highlights the potential of traditional techniques when properly optimized.

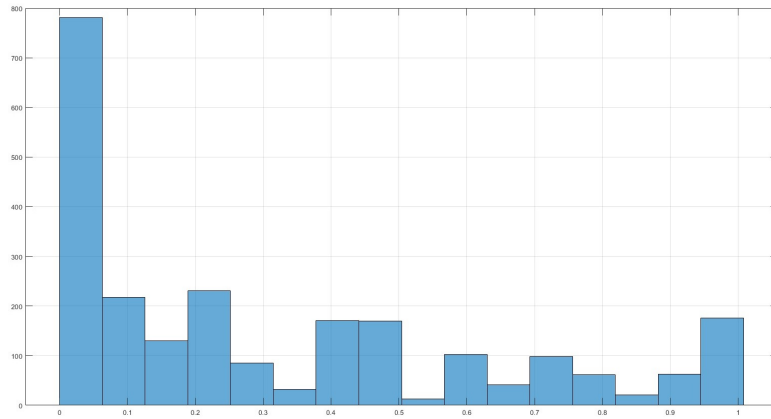
The morphological parameters from histopathology images are extracted, and the correlation between the features and the residual TC is examined. Linear regression and statistical methods were applied, and twenty-two morphological features from the nucleus, epithelial region, and the complete image were extracted. Then, Machine Learning (ML) algorithms were trained with the selected key parameters for a cell classification. The methodology was validated with the score assigned by pathologists through the ICC, given in the work by (Peikari et al., 2017). The selection of key morphological parameters improved other results reported based on ML, and were very close to DL techniques.

## 5.2 Materials

The data used for this experiment is the SPIE BreastPathQ Challenge Dataset described in chapter 4. It consists of a total of 2,395 images divided into a training and testing that were extracted from 15 whole slide images (WSI) of breast tissue stained with Haematoxylin and Eosin (H&E) released in the 2019 SPIE Breast Challenge. In the bottom of Figure 5.1, we can observe images from low to high TC, and the TC value assigned to each case. In this dataset, every image from the 2,395 has assigned a TC value by a specialized pathologist given by a real number in the interval  $TC \rightarrow (0, 1)$ . Only the training dataset is known to the participants. The statistical distribution of the training dataset can be seen in Figure 5.2.



**Figure 5.1:** The sample images show cellularity values from zero to one and magnified versions of cases with values 0, 0.4, 0.7 and 1.



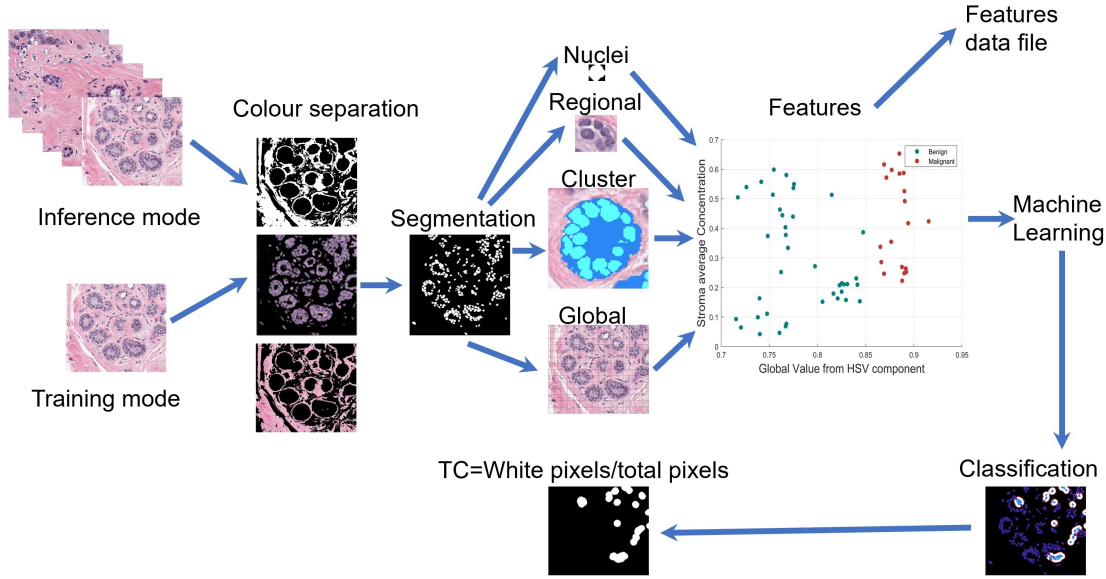
**Figure 5.2:** The statistical distribution of the 2394 images from the SPIE BreastPathQ dataset in a histogram reveals the total amount of patches in the 16 selected intervals.

### 5.3 Methods

The methodology proposed in this experiment is based on the traditional computer vision workflow described in the overview chapter of this thesis. We divided this methodology into three basic steps, a pre-processing for the colour separation and nuclei segmentation. Next, feature extraction is done by manual handcraft methods using image processing and computer vision techniques. Finally, the classification to estimate TC.

This approach based on feature extraction parameters is similar to the one described by (Peikari et al., 2017). Also, the methodology is similar to (Dong et al., 2014), but we also include parameters from the neighborhood of the nuclei and the whole image patch. It also has similarities to the work presented by (Fondón et al., 2018), in which the region around the nuclei is analysed; however, we included the morphology parameters from the ducts region and the whole breast patch image.

The full process of this method is presented in figure 5.3. The program runs for



**Figure 5.3:** The computer program runs under two operational modes, one for training the machine learning models and another for cellularity assignment. The extracted features are obtained from the nuclei, regional, the ducts cluster regions, and the full image.

training the machine learning model and runs for inference to process the testing data. In the training mode, parameters from each image are extracted and saved to train the ML classifier. In the inference mode, the full set of testing images is processed to estimate TC. The main computer vision routines are explained in the next sections.

### 5.3.1 Colour separation

Image is converted to HSV colour space and separated in colour components by K-means clustering (Shalev-Shwartz and Ben-David, 2014; Takano, 2021). Clustering algorithms such as K-means can also be used for image segmentation by grouping similar pixels based on their feature vectors; then, the separation of colours is obtained. This is a non-supervised method based on grouping the elements closest to a reference parameter through a minimization function, usually the Euclidean distance. Let the image  $I = \{x_1, x_2, \dots, x_{M \times N}\} = \mathbf{x}$ , be represented in vector form  $\mathbf{x}$  where  $x_i$  is a 3-space colour,  $K$  be the number of clusters (for the H&E histopathology image case are three colour components pink, blue and background) and  $\mathbf{C}_k = \{C_1, C_2, \dots, C_j\}$  be the set of pixels assigned to the  $k_{th}$  cluster, where  $k = 1, 2, \dots, K$ . The centroid  $\mu_k$  of cluster  $k_{th}$  is given by the formula,

$$\mu_k = \frac{1}{|\mathbf{C}_k|} \sum_{x_i \in \mathbf{C}_k} x_i \quad (5.1)$$

where  $|\mathbf{C}_k|$  is the size of  $\mathbf{C}_k$ , i.e., the total number of pixels in cluster  $\mathbf{C}_k$ .

This method attempts to find the given  $K$  clusters through an iterative process by an objective function that minimizes the sum of the square distances between each pixel and its corresponding centroid.

$$\min \sum_k \sum_{x \in \mathbf{C}_k} \text{dist}(x, \mu_k) \quad (5.2)$$

where  $\mathbf{C}_k$  denotes the  $k_{th}$  cluster  $k$ , and  $\mu_k$  is the centroid of  $\mathbf{C}_k$ .

This method was used to extract the three main colour component: blue for nuclei, pink for stroma and the remaining is the background. Let,  $I_s$ ,  $I_n$ , and  $I_{ba}$  be the binary mask images for each corresponding cluster, in which  $I_s$  is for stroma (pink colour),  $I_n$  is for the nuclei (blue colour) and  $I_{ba}$  for the remaining background of the image. These binary images can be considered subsets from the original image and are mutually exclusive, which means:

$$I_s \cap I_n \cap I_{ba} = \emptyset \quad (5.3)$$

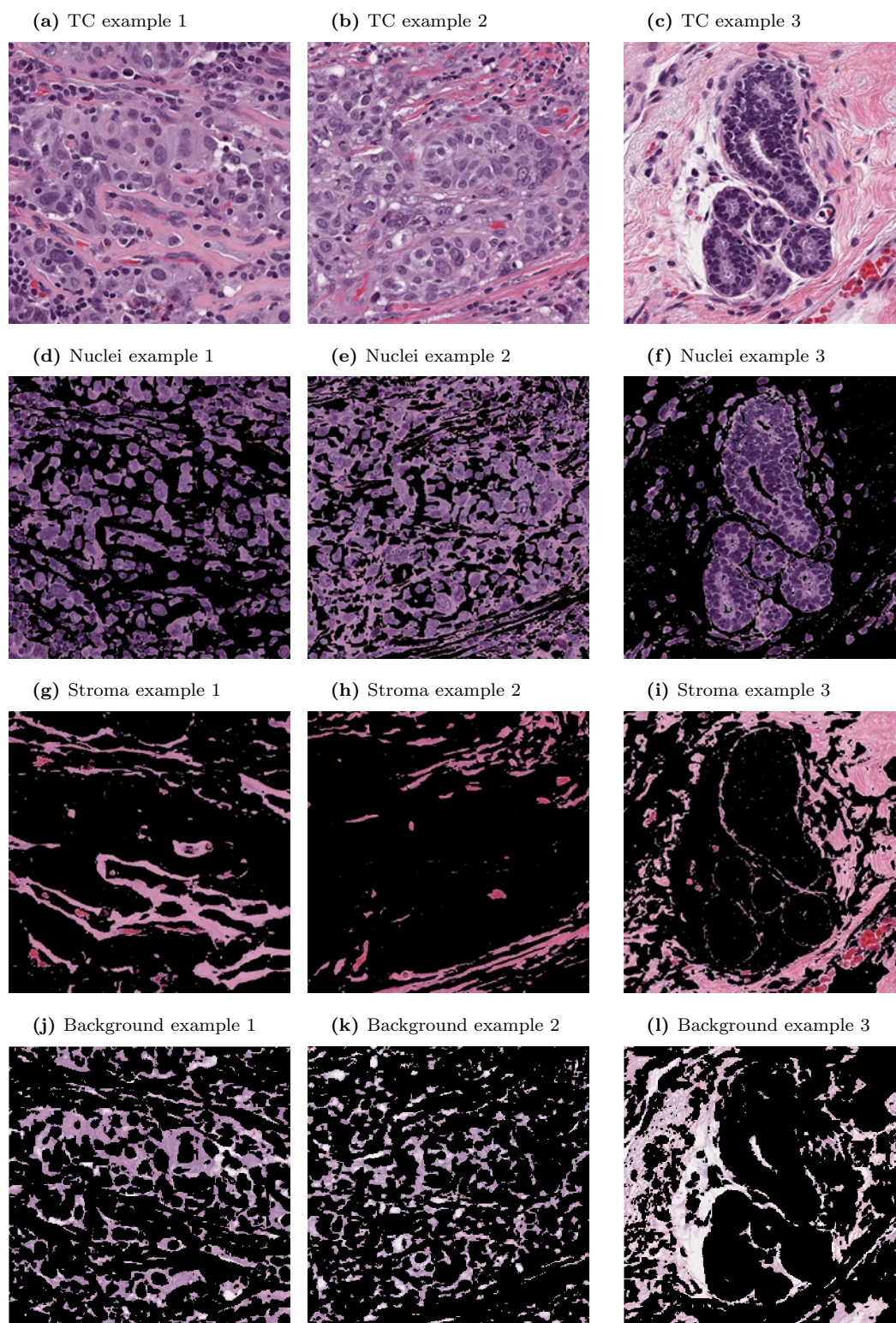
and

$$I_s \cup I_n \cup I_{ba} = 1 \quad (5.4)$$

Samples of the extracted images are shown in Figure 5.4.

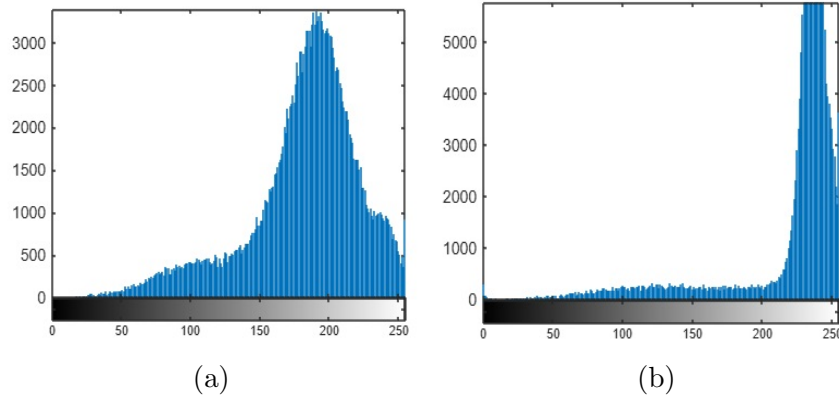
### 5.3.2 Nuclei Segmentation

A nuclei segmentation method based on a novel nuclei enhancement algorithm is applied after the implementation of K-means colour separation, and before the Otsu's binarization step. The nuclei enhancement function increases the contrast of the nuclei in the H&E grayscale original image. The motivation of the nuclei enhancement method is due to the idea of using the nuclei colour information obtained after the colour separation step to improve the contrast between nuclei and background, which avoids the segmentation of artifacts and elements that are not part of the nuclei region, as in most of the cases are of different colours. The enhancement pro-



**Figure 5.4:** The images represent the colour separation by K-means algorithm, the first row are three H&E patches and second row are the coloured images for the nuclei region (blue cluster), the third row are the colour images for the stroma region (pink cluster). The fourth row are images for the background region.

cedure weakens the background and makes the nuclei clearly visible by the use of the gamma function. The result of the proposed enhancement algorithm modifies the histogram making an image with higher contrast as can be seen in the histograms shown in Figure 5.5.



**Figure 5.5:** Histograms before and after the enhancement algorithm. In figure a) the histogram shows a soft transition between background and foreground (nuclei), which means there is no clear boundary between background and foreground; and in b), the transition is sharp, which indicates the background can be better removed.

The enhanced image is calculated by:

$$I_{en} = K_1 * [I_o * I_m] + K_2 I_o + \gamma(I_m, 0.1) \quad (5.5)$$

where  $I_o$  is the RGB image converted to graylevel,  $I_m$  is the gray image after a median filter of size  $3 \times 3$  was applied, and  $\gamma(I_m, 0.1)$  is the gamma correction  $I_m$  of parameter 0.1. This value was selected low to start from a lighter background.  $K_1$  and  $K_2$  are constants that control the enhancement of nuclei and weakens background intensity respectively, and were experimentally adjusted. First, both constants were fixed to 0.5 and as  $K_1$  is increased and  $K_2$  is reduced the nuclei region is enhanced. The enhanced image  $I_{en}$  is binarized by Otsu's algorithm (Otsu, 1979) combined with watershed (Meyer, 1994) separation of overlapping nuclei.

## Thresholding

A straightforward segmentation solution is thresholding. It offers high performance when the images have enough contrast between objects of interest and the background. It lacks robustness when texture changes and cannot separate overlapping objects (Irshad et al., 2013).

Let  $I_g$  be a grayscale image as defined in equation (2.3). It can be converted into a binary image  $I_{BW}$  by selecting a threshold value  $T$  and classifying the  $i_{th}$  pixel

in the image ( $i \in I_g$ ) as foreground or background based on whether its intensity is above or below the threshold. This can be expressed,

$$I_{BW}(i) = \begin{cases} 1, & \text{if } I(i) > T, \\ 0, & \text{otherwise.} \end{cases} \quad (5.6)$$

In case  $T$  changes during the conversion to binary, it is a variable threshold, which means the value of  $T$  depends on the neighborhood properties of the pixel  $i$ . This is called adaptive or dynamic threshold. An optimal adaptive method for the threshold estimation was presented by Nobuyuki Otsu (Otsu, 1979), which is based on a histogram distribution analysis and it maximizes the inter-class variance and minimizes the intra-class variance. It is the well-known *Otsu's Method*.

### Watershed algorithm

In region-based methods, pixels are grouped into regions based on certain criteria such as color, texture, or intensity similarity. One popular region-based method is the watershed algorithm (Meyer, 1994), which treats the image as a topographic surface and simulates flooding to segment regions. First, the gradient of the image,  $\mathbf{LS}(i)$ , indicates the slope connecting pixel  $i$  to its neighbors,

$$\mathbf{LS}(i) = \max_{j \in N(i)} \left( \frac{I(i) - I(j)}{d(i, j)} \right) \quad (5.7)$$

where  $d(i, j)$  represents the Euclidean distance between pixels  $i$  and  $j$ , and  $N(i)$  is the set of neighbors of the pixel  $i$ .  $\mathbf{LS}$  is a measure of the local intensity variation or gradient variation between pixel  $i$  and its neighbors. Pixels with low  $\mathbf{LS}$  values (this means low-intensity differences with their neighbors) are considered markers for starting the flooding process in the watershed algorithm. To perform the flooding, a cost function is used to analyse the path from pixel  $i$  to pixel  $j$ , and is defined as

$$\text{cost}(i, j) = \begin{cases} \mathbf{LS}(i)d(i, j) & \text{if } I(i) > I(j) \\ \mathbf{LS}(j)d(i, j) & \text{if } I(i) < I(j) \\ \frac{1}{2}(\mathbf{LS}(i) + \mathbf{LS}(j)) & \text{if } I(i) = I(j) \end{cases} \quad (5.8)$$

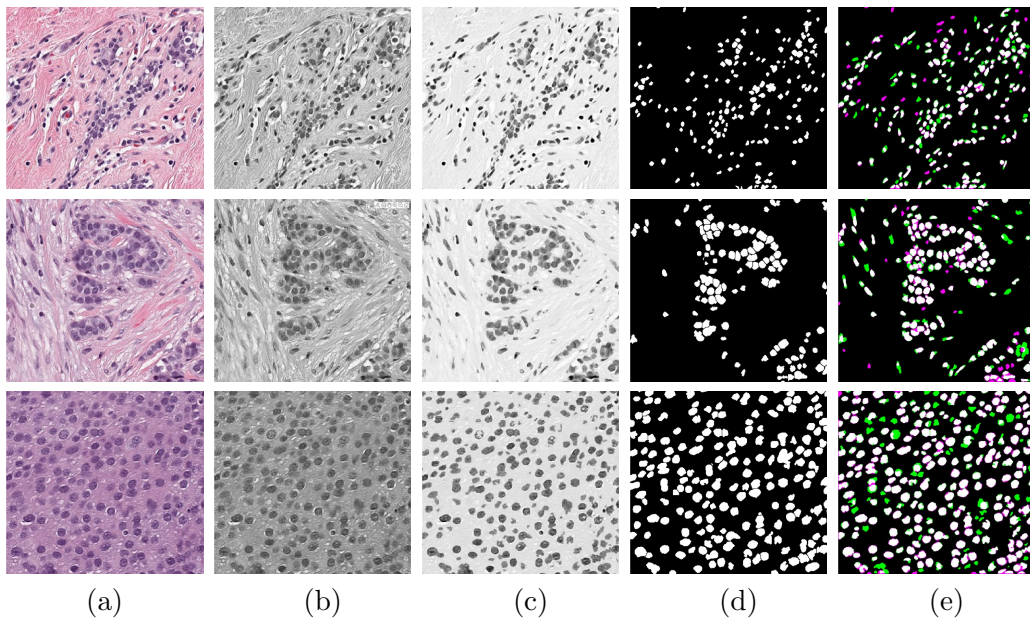
The topographical distance between pixels is measured based on intensity values,

which is the topographic distance, and it is calculated by,

$$\min_{(i_0, \dots, i_t) \in \Pi} \sum_{k=0}^{t-1} d(i_k, i_{k+1}) \text{cost}(i_k, i_{k+1}) \quad (5.9)$$

and  $\Pi$  represents all paths from  $i$  to  $j$ . The topographic distance is used to determine the flow of *water* between neighboring pixels.

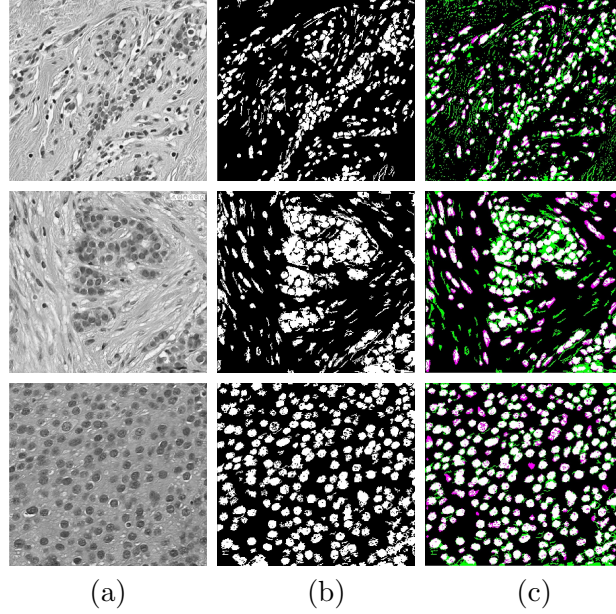
The watershed transform is suitable for overlapping object separation and is commonly used for nuclei separation. This segmentation procedure is seen in Figure 5.6 for three TC cases. The enhanced grayscale image is in column c) and watershed result, which is applied after Otsu's binarisation, is presented in column d). A validation analysis of this procedure was done by means of Jaccard Index (Jaccard, 1901) shown in (e). The same images segmented without the enhancement method, and without the watershed separation can be seen in Figure 5.7



**Figure 5.6:** The first row corresponds to a TC value of zero, the second row is an example of TC= 0.5 and the third row is TC= 1, cases of low, medium and high cellularity values were chosen. Images in column (a) are the original image, (b) grayscale images, (c) enhanced images, notice that the nuclei region is darker whilst the background becomes lighter, (d) Binary image obtained with an adaptive threshold and after applying the watershed algorithm and (e) validation of nuclei segmentation by Jaccard index. The reference image was segmented manually by using QuPath Platform, in column e) green colour are false positives, and pink colour are false negatives. It can be seen for high TC that false negative cases are reduced.

### 5.3.3 Ducts segmentation

As described, ducts are tube-like structures responsible to carry milk to the nipple. In DCIS cancerous cells grow within the ducts and do not invade surrounding tis-



**Figure 5.7:** Otsu segmentation without any pre-processing or post processing. Images are the same of previous example and column (a) are the original grayscale image, (b) are the binarised images and c) are the validation of nuclei segmentation by Jaccard index. The reference ground truth image was segmented manually by using QuPath Platform, green colour are the false positive and pink colour are the false negatives. Some artifacts, which are elements that are not positive cells, were segmented and presented in green, which corresponds to false positives. When the enhancement method is applied, artifacts are weakened and then, they are not segmented.

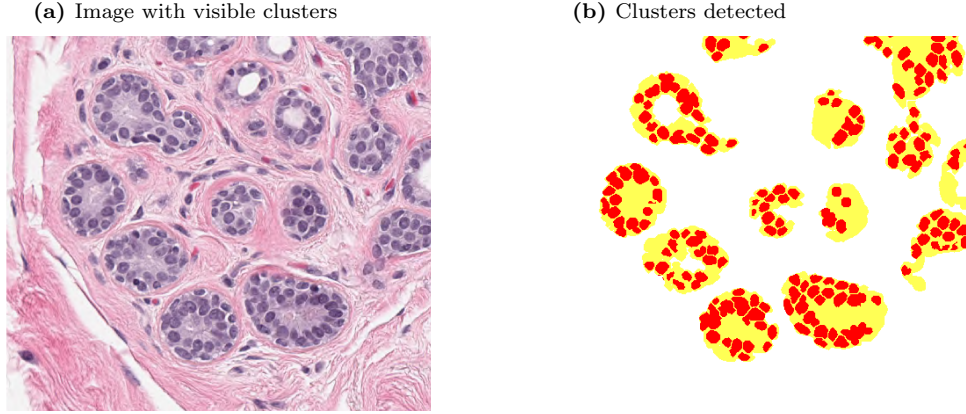
sues. In this step of the method, ducts segmentation is performed, the goal is to determine features from this particular structure. A new colour separation by the K-means algorithm of size  $K=2$  was performed, and a new background image, say  $I_{ba}$  is obtained. Ducts are detected based on set theory: let  $I$  be the original color image,  $I_D$  is the ducts image,  $I_D \subseteq I$ . Then  $I_D$  is obtained by subtracting the new background mask from original image  $I$ ,

$$I_D = I - (I_{ba} + I_s) \cdot * I \quad (5.10)$$

when the mask  $I_{ba}$  is multiplied by  $I$ , it gives the remaining region outside the ducts. Although it is a straightforward method, it obtains the ducts efficiently as seen in figure 5.8.

### 5.3.4 Extraction of nuclei morphological parameters

The binary image of the nuclei cells is then analysed by morphological methods to obtain a set of features from the nuclei region: area, perimeter, roundness, eccentricity, centroid X, centroid Y, major axis, minor axis, and orientation angle. Also,



**Figure 5.8:** (a) The original patch with visible clusters. (b) Clusters detected by the method, the image was modified to colors red and yellow to emphasize the result.

a sub-region inside the segmented cell body was obtained to compute the texture and mean HSV values of the nuclei cell body.

### Morphological Methods

Morphology operations are used to extract the main features of an object in a binary image. Set theory is used to explain this type of processing (Gonzalez, 2009). In the morphology operations, a predefined shape called structure is compared with the original image. Let  $U_f = I^{-1}(\{1\})$  be the foreground pixels of image  $I$ . The structure  $S$ , where  $S \in \mathbb{Z} \times \mathbb{Z}$ , is a small array of pixels, which is compared with all the corresponding pixels of an image neighborhood. Some of the common structures are square, diamond, disk, cross, and line (Gonzalez, 2009). The structure is positioned at all possible regions of the image, this is a morphological operation between the image and the structure. Two basic morphological operations are erosion and dilation.

### Dilation

It is an operation associated with the growth of the object, that is, it enlarges the object by the shape of the structure. Let  $U_f$  be the object and  $S$  be the structure, the operation  $x + s$  represents the translation of the structure along the foreground object. The dilation operation is defined as:

$$U_f \oplus S = \{x + s | x \in U_f \wedge s \in S\} \quad (5.11)$$

when a pixel from the structure overlaps with the foreground object the structure is added to the image.

### **Erosion**

The erosion is the opposite of dilation, which means it reduces the object. This operation is defined as follows:

$$U_f \ominus S = \{x | \forall s \in S, x + s \in U_f\} \quad (5.12)$$

when any part of the structure fits the background, then the central pixel of the structure is removed.

### **Opening**

The opening is defined as an erosion followed by dilation and is defined as:

$$U_f \circ S = (U_f \ominus S) \oplus S \quad (5.13)$$

The erosion operation removes those pixels in the image that are smaller than the reference structure then dilation returns the image object to its original size and the final image is smoothed out.

### **Closing**

It defined as a dilation followed by erosion

$$U_f \bullet S = (U_f \oplus S) \ominus S \quad (5.14)$$

through this operation, the holes smaller than the reference structure are filled.

Opening and closing have distribution properties

$$(U_f \circ S) \circ S = U_f \circ S \quad (5.15)$$

$$(U_f \bullet S) \bullet S = U_f \bullet S \quad (5.16)$$

### Top-Hat Transformation

This operation is defined by subtracting the original image by its opening. The top-hat transformation of image  $I$  defined as  $T_w(I)$  is:

$$T_w(I) = U_f - (U_f \circ S) \quad (5.17)$$

This structure is important to enhance details and small elements within the image.

### Bottom-Hat Transformation

Bottom Hat transformation is obtained first by closing the image and then subtracting the original image from the result of the closing. It is defined as  $T_b(I)$ :

$$T_b(I) = (U_f \circ S) - U_f \quad (5.18)$$

The object properties can be calculated by morphology functions of binary objects: size, perimeter, roundness, centroid, elongation, and orientation of all elements. These operations are defined next.

### Area

The area is the total number of pixels that form a binary object. First, the object  $O$  of the binary image  $I$  is formed by the  $K$  connected pixels, the area is calculated by

$$Area(O) = \sum_{x \in O} 1 \quad (5.19)$$

### Perimeter

The perimeter is defined from the boundary of the object  $B$ , which is the set of pixels of the object which are adjacent to a background pixel. The perimeter can be determined by the length of the boundary,

$$Perimeter(O) = \sum_{x \in O} length(x) \quad (5.20)$$

where the contribution of the boundary is based on its connectivity, pixels in diagonal have a norm of 1 pixel size =  $\sqrt{2}$ , and its length is

$$length(x) = \begin{cases} 1 & \text{horizontal and vertical} \\ \sqrt{2} & \text{diagonal} \end{cases} \quad (5.21)$$

### Circularity

Circularity indicates how close to a perfect circle an image object is. It is measured by the following equation,

$$C(O) = \frac{4\pi Area(O)}{Perimeter(O)^2} \quad (5.22)$$

### Centroid

The centroid coordinates  $(\bar{m}, \bar{n})$  is the point that concentrates the object mass and can be determined by

$$\bar{m} = \frac{1}{Area(O)} \sum_{(m,n) \in O} m \quad (5.23)$$

, where  $m$  and  $n$  are iterated through all the object coordinates.

$$\bar{n} = \frac{1}{Area(O)} \sum_{(m,n) \in O} n \quad (5.24)$$

where  $m, n$  are coordinates of pixels within the object.

### Elongation

Elongation is the relation between the longitude and width of the object and is obtained by the relation between the major axis length (which is the longest dimension of the object), and minor axis length (which is the shortest dimension perpendicular to the major axis), and is evaluated by the equation

$$Elongation(O) = \frac{\text{Major Axis length}}{\text{Minor Axis Length}} = \frac{b}{a} \quad (5.25)$$

### Eccentricity

Eccentricity describes how much an object's shape deviates from a circle, and it can be calculated as

$$\text{Eccentricity}(O) = \sqrt{1 - \frac{b}{a}} \quad (5.26)$$

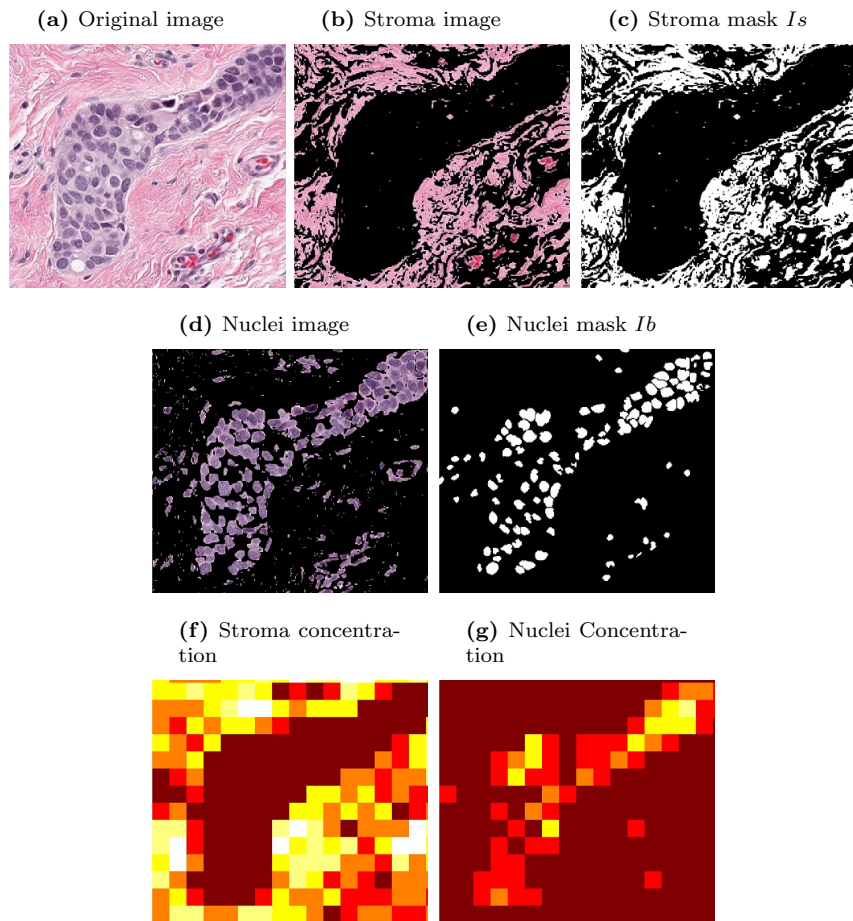
where  $a$  and  $b$  are the semi-major and semi-minor axes of the ellipse that matches the second order central moment.

Nuclei parameters			
Area	Eccentricity	Roundness	Centroid x,y
Perimeter	Orientation	Major Axis	Minor Axis
Contrast 1	Contrast 2	Homogeneity 1	Homogeneity 2
Mean H	Mean V	Mean S	
Regional parameters			
Stroma Ip	Background Iba	Mean Ib	Epithelial Tissue
Mean H	Mean S	Mean V	
Histogram H1	Histogram H2	Histogram H3	Histogram H4
Histogram S1	Histogram S2	Histogram S3	Histogram S4
Histogram V1	Histogram V2	Histogram V3	Histogram V4
Ducts parameters			
Area	Roundness	Total cells	Distance to centroid
Global image mean values			
Stroma Ip	Background Iba	Nuclei Ib	

**Table 5.1:** Summary of extracted features. Nuclei parameters measure the nuclei shape. Homogeneity and contrast are texture parameters. Features at a regional level are computed at 4 different window sizes. Features from the full image represent mean values.

### 5.3.5 Extraction of regional morphological parameters

The original image was divided in sections of different window sizes to extract also parameters of a regional area surrounding the segmented nuclei. Table 5.1 indicates the parameters extracted from this new sub-image indicated as regional parameters. Figure 5.9 shows examples of mean values of stroma and nuclei for a window size of  $30 \times 30$  pixels, which can be seen in each small square in images f) and g). The mean intensity is presented in a colour scale in which brown color indicates the lowest and white is the highest mean value. Different window sizes were selected:  $30 \times 30$ ,  $60 \times 60$ ,  $90 \times 90$ , and  $120 \times 120$  pixels. The correlation analysis (as will be explained in the next section) indicated that the best window size was the smallest. At each



**Figure 5.9:** Regional Analysis of the image, figure (a) is the original image  $I$ . Figure (b) is the stroma region image obtained from color separation by  $I * I_s$ . The figure (c) is the mask image  $I_s$ . Figure (d) is the nuclei image obtained from color separation and calculated by  $I * I_n$ . Figure (e) is the nuclei mask  $I_n$  image. Regional image mean values of the stroma region is in (f), in which brown color indicates the lowest and white is the highest region concentration. Figure (g) is the regional concentration of  $I_n$ . This example was processed at  $30 \times 30$  pixels window.

window area, the parameters estimated are the four bins histogram from the HSV image and also the mean values of  $R$  of  $I_s$ ,  $I_{ba}$ , and  $I_n$  masks as described in the colour separation section. The mean colour value mask is a real value computed by the formula

$$R = T_W/T_P \quad (5.27)$$

where  $T_W$  is the total number of white pixels, and  $T_P$  is the total pixels inside the selected window neighborhood. Another regional analysis is the estimation of the parameters from ducts and lobular units, already described in previous section (see Figure 5.8). The following morphological parameters can be extracted from the cluster regions: total cluster area, roundness of cluster, number of cells inside the cluster, and distance from cells centroid to cluster centroid. A summary of the full set of extracted parameters is presented in table 5.1.

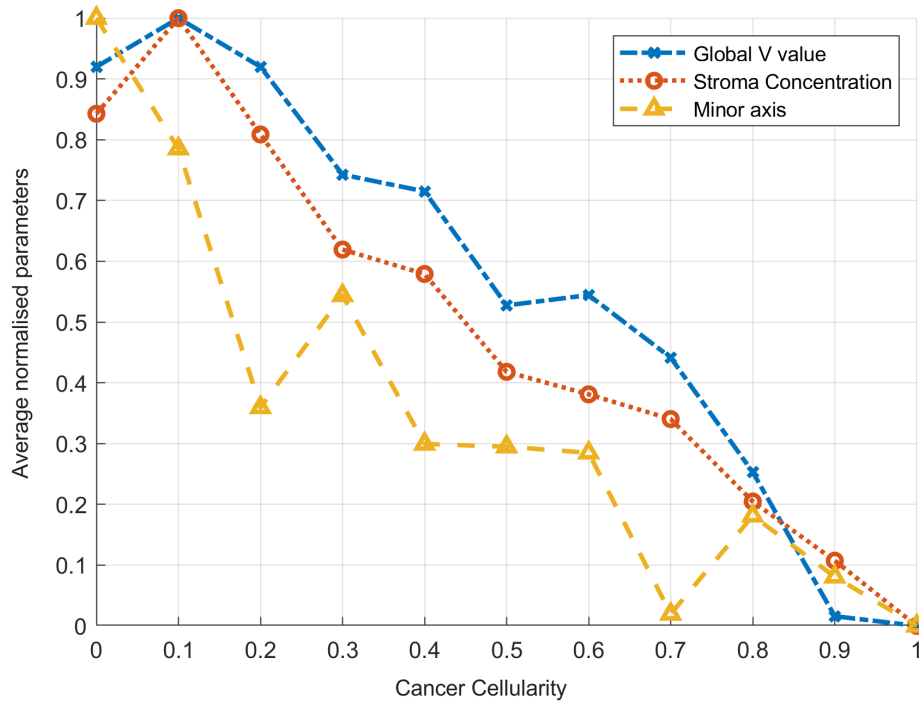
### 5.3.6 Correlation analysis of morphological parameters to TC

From the full dataset of training images, those corresponding to cellularity values equal to (0,0.1,0.2...1.0) were selected and processed to determine all the morphological parameters at the three regions of interest, and their mean, standard deviation, maximum, and minimum values were computed. Over 150 features were obtained, and linear regression and lasso analysis were determined. Parameters of coefficient above 0.80 from linear regression and in concordance with lasso-selected parameters, after redundant removal, yield 22 parameters that have the strongest correlation with TC. The plot in Figure 5.10 illustrates three parameters with the strongest correlation with TC.

The 22 selected parameters are described next. Features related to segmented nuclei are eccentricity, roundness, major axis, minor axis and perimeter. Parameters computed at neighborhood region are nuclei density concentration, Hue, Value, and Saturation histograms of the regional HSV colour map image. Finally, parameters computed from the full image are HSV average values from HSV components, nuclei concentration, background concentration, and stroma average concentration.

### 5.3.7 Training of machine learning algorithms

Machine learning algorithms were trained with the 22 vector size parameters described before. To generate the ground truth training data, a subset of random

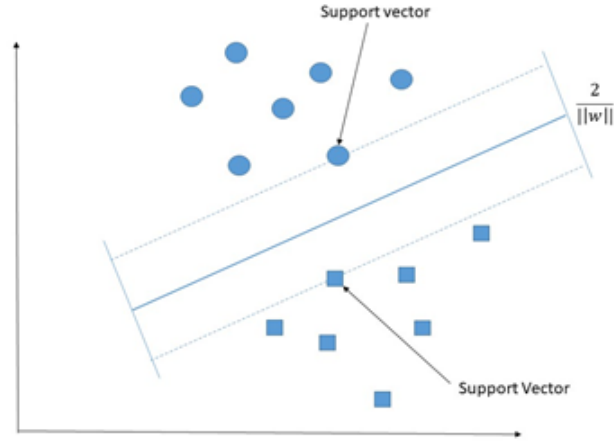


**Figure 5.10:** Graphical display of three selected morphological parameters against the cellularity: Global stroma region (o), minor axis ( $\Delta$ ) and *Value* (from HSV) concentration in region (x). Cellularity was manually ranked by a pathologist for each patch, and the corresponding parameters extracted from each patch also. Plots represent average parameter value normalized with maximum value to be compared in the same graph. The graph indicates an inverse relationship between respective parameter and cellularity.

images with TC zero was used to extract features for the class non-tumour, and a subset of images from low cellularity to high cellularity was used to extract features for the tumour class. From these sample images, 4533 nuclei cell images were obtained, and their morphological parameters were used as ground truth for training a prediction model for nuclei classification between malignant and normal cells. Three algorithms were tested: Support Vector Machines (SVM) (Boser et al., 1992), Nearest K-Network (KNN) (Fix, 1985), and Adaptive Boosting (AdaBoost) (Freund and Schapire, 1997).

### Support Vector Machines SVM

Support Vector Machine is an algorithm used for the classification of multiple variables into a defined number of classes (Boser et al., 1992). Let's start with a basic linear classification of two classes, for example, squares and circles. A linear classifier separates objects by a dividing line called a hyperplane. SVM is a method that separates both classes, maximizing the margin  $M$ , which is defined as the distance between the hyperplane and the closest data point. This is explained in figure 5.13.



**Figure 5.11:** The image describes a basic linear SVM classification between two categories (circles and squares) and Support vectors with the distance to be maximized on the hyper-planes.

Let  $\mathbf{x}$  be a vector of dimension  $n$  of data that belongs to class A or B. There are  $p$  examples of the data given by  $(x_1, y_1), (x_2, y_2), \dots, (x_p, y_p)$  in which  $y_k = 1$  if  $x_k \in A$  or  $y_k = -1$  if  $x_k \in B$ . This algorithm is based on a decision function defined as

$$D(\mathbf{x}) = w\varphi(\mathbf{x}) + b \quad (5.28)$$

where  $w$  are adjustable parameters and  $\varphi$  is a predefined function. Starting with a dot product, two hyper-planes that define the offset on each side of the main hyper-plane can be created. No objects can be located within the region that spans between the main hyper-plane, and edge hyper-planes. On one side, the equation is

$$w \cdot x - b = 1 \quad (5.29)$$

and on the other side is

$$w \cdot x - b = -1 \quad (5.30)$$

The objects that lie on the edge hyper-plane are the support vectors. Having a bigger gap between the categories there would be an increase in confidence in the predictors, and maximizing the distance would be minimizing the value  $w$ . The distance  $d$  is:

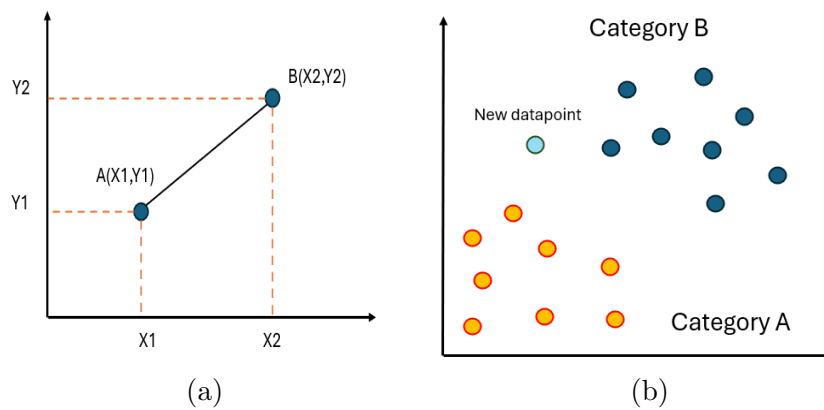
$$d = \frac{D(x)}{\|w\|} \quad (5.31)$$

Where  $\| \cdot \|$  is the Euclidean norm. The maximization is done by the Lagrange multipliers. Also, the method can be extended into a multidimensional space to separate categories as much as possible. Non-linear classification can be applied.

### Nearest Neighbor pattern classification KNN

KNN was selected because it is a well-known algorithm within clustering and data classification, which was used in our case to classify outcomes between normal and malignant classes. KNN, as described by (T.M. Cover) is a classification method that predicts the class of a data point based on the values of its neighbors by using a distance metric, in which the Euclidean distance is generally used, given by

$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (5.32)$$

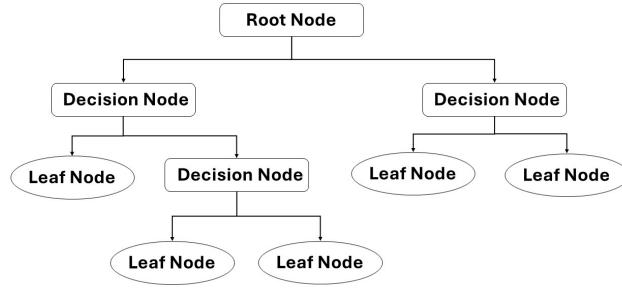


**Figure 5.12:** The Euclidean distance a) is used to measure how close one point is to another data point. New data is classified by the calculation of the K number of neighbors, and the class is selected by majority voting.

### Adaptive Boosting AdaBoost

AdaBoost is a decision-tree learning algorithm that starts from observations of a certain item. In a simple decision tree, a classifier has a structure similar to a tree, with its root at the top and a certain condition splits the tree into branches, until a leaf is reached, which is a node that do not split, as can be seen in Figure 5.13.

There is a special tree classifier named bagged-tree, which has a time  $T$ , and a number of  $n$  training samples with replacement; thus,  $T$  training sets of size  $n$  are obtained. The decision tree is trained with these sets, combined with a majority voting rule. Adaptive boosting, also called AdaBoost, is a bagged tree classifier with



**Figure 5.13:** A tree classifier has a tree structure, based on one root node, multiple internal nodes, and multiple leaf nodes. The leaf nodes correspond to the decision outcomes. Image adapted from (Zhou, 2021).

a weight added to it. This weighted sample is the input to the next tree for training, and the process is repeated. In this process, the final trees are focused on classifying hard samples.

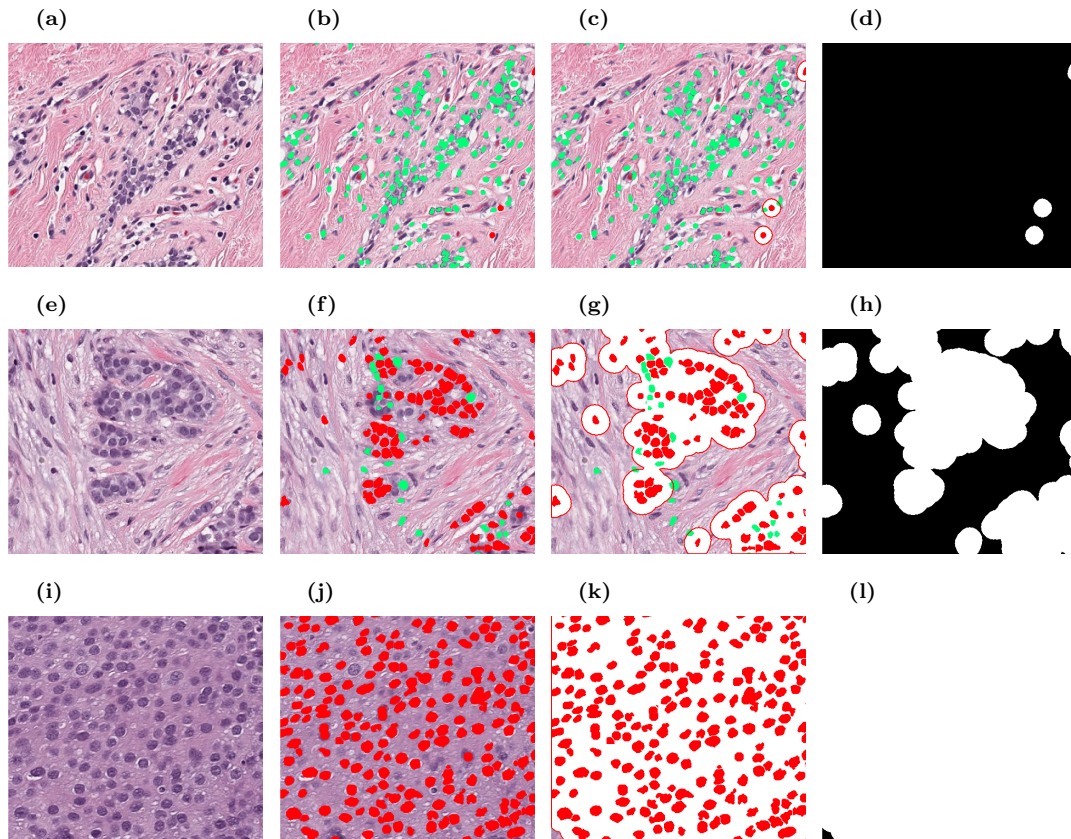
The accuracy of the training process for every selected method with the training data showed values up to 0.99 due to a high correlation between the selected features used for training. True Positive Rate (TPR) achieved are 0.97 for SVM, 0.95 for AdaBoost, and 0.97 for KNN.

## 5.4 Assessment of Tumour Cellularity

The methodology to estimate TC is illustrated in Figure 5.14 with three representative images with increasing TC from top to bottom. The nuclei are segmented, and their corresponding parameters are extracted. Then, the prediction function classifies every cell as either normal (green) or malignant (red) as illustrated in the second column in Figure 5.14. Next, an estimation of full cell cytoplasm is done by morphological dilation drawn with the white circles around malignant cells, see column 3 in Figure 5.14. The cellularity detected region is shown in column 4, the white area is a rough representation of tumour segmentation. TC is computed as the ratio of the area covered by cellularity (white in the figure) over the total area (white and black).

## 5.5 Results

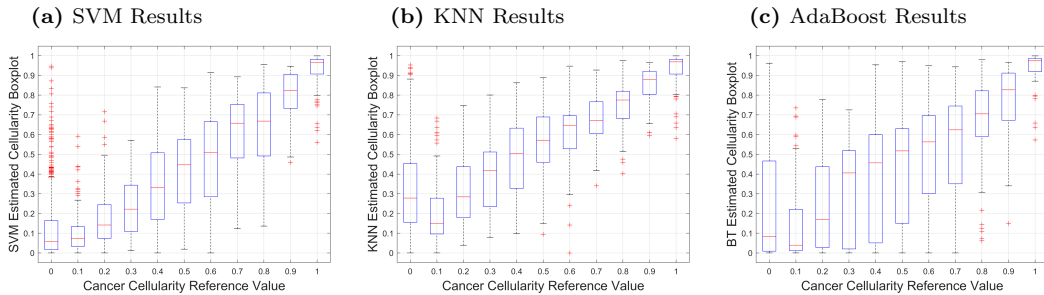
The experiment presented in this chapter performs the estimation of TC of BC images under NAT therapy. However, for this task, the nuclei segmentation is first obtained. A validation analysis of the nuclei segmentation was done through the



**Figure 5.14:** Three TC cases are presented in each row: 0, 0.5 and 1. Column 1 are the original images, the image is segmented and key parameters are computed, then a classification predictor estimates either malignant or normal cells, shown in red and green respectively in column 2. A dilation of malignant nuclei estimates full cytoplasm of every detected cell in column 3, and TC region is shown in white in column 4. The cellularity metrics calculated for the three examples by the proposed methodology are:  $TC = 0.0113$ ,  $TC = 0.5181$  and  $TC = 0.9936$ .

Jaccard Index (Jaccard, 1901). As there is no ground truth for the segmented cells, the reference nuclei masks were obtained by a manual segmentation of a selected subset of images using the QuPath platform. Results indicate an average Jaccard index of 0.73, after comparing cases of low, medium and high TC. Although this result might be improved, (Peikari et al., 2017) suggests that cell segmentation accuracy does not have a significant effect in TC assessment.

Three machine learning prediction functions were trained to classify cells into tumour and non-tumour classes. TC was computed by measuring the total tumour cell area. The method was tested with the set of 2579 images already classified by a pathologist with a TC value, without using the selected subset of training images. A dispersion plot for each model is presented in Figure 5.14. The dispersion plot indicates how close the method approximates the pathologist classification assignment, larger boxplots indicate poorer concordance. Results have a better approximation at higher cellularity values ( $TC > 0.70$ ) and perform well with the KNN algorithm. Also, around the middle region ( $0.4 < TC < 0.6$ ) has a good approximation with AdaBoost. At low cellularity values ( $TC < 0.3$ ) three methods present deviation, with its higher at cellularity zero, which corresponds to images with only benign nuclei cells. According to Mean Square Error (MSE), SVM performs better overall the cellularity region. This result can be validated by a visual inspection of boxplots of Figure 5.15, in the three cases there is a positive correlation between the actual cellularity (horizontal) and the estimated cellularity (vertical). However, SVM shows less dispersion, especially in the lower values of cellularity as compared with the other two techniques.



**Figure 5.15:** Results of implementation on Training Data shown as boxplots. a) SVM, b) KNN and c) AdaBoost. Every boxplot corresponds to the statistical result of all the same TC value results. Large boxplot indicate large deviation. A visual inspection indicates the lowest deviation corresponds to SVM method in (a).

Intra-correlation coefficient was computed over the test set after submission to

the challenge, and indicates a 0.78 ICC. This result outperformed those obtained by the method of (Peikari et al., 2017). Results based on DL techniques like a combined hand engineering Deep Neural Network reported by (Akbar et al., 2019) are slightly above the proposed methodology. These results were compared over the 1119 testing images used during the submission to the SPIE Breast Challenge, with an unknown TC value. Our participation in the challenge resulted in an average ICC=0.76, which resulted in the middle score ranking of participants.

	Hand Engineering (Peikari)	Key Parameters (This research)	Combined Deep Network (Akbar)
ICC	0.75	0.78	0.79
[L,U]	[0.71, 0.79]	[0.75, 0.80]	[0.76, 0.81]

**Table 5.2:** The Intraclass Correlation Coefficient (ICC) of the proposed methodology against a Hand Engineering Methodology (Peikari et al., 2017), and a combined DL and hand engineering methodology (Akbar et al., 2019). Lower and upper bounds are shown in square brackets. Notice how close are the results of the proposed methodology against the combined DL approach.

## 5.6 Summary

This chapter presents an automated computational pathology experiment that processes H&E-stained histopathology digital images by extracting key morphological parameters at the cellular, regional, and global levels. After a statistical analysis of the extracted features, 22 key morphological parameters were found to be strongly correlated with tumour cellularity. Notably, the most significant parameter was stroma concentration, aligning with the findings of (Beck et al., 2011), which suggests that stromal histology is closely associated with breast cancer prognosis.

To identify tumour regions, three different machine learning algorithms for cell classification were evaluated, with the best performance achieved using Support Vector Machine (SVM) algorithm. The relevance of this work lies in the careful selection of morphological features used to train the model, which yielded results comparable to those achieved by deep learning methods.

While modern deep learning techniques dominate the field, this work highlights that traditional, handcrafted methods—when properly designed and applied—can still deliver competitive results. These approaches should not be disregarded, as they continue to offer valuable insights and remain relevant in certain contexts.

## Chapter 6

# Deep learning networks for multi-class segmentation

This Chapter explores the use of Deep-Learning (DL) methods to perform multi-class segmentation of BC H&E images. Three different DL models were compared to perform multi-class segmentation: MobileNet, ResNet, and UNet. The Breast Cancer Semantic Segmentation Challenge (BCSS) dataset was used in this experiment, which was prepared by a collaborative effort to develop machine learning models for automated tissue segmentation (Amgad et al., 2019). To train the selected models, the hyper-parameters were adjusted, and a performance comparison in terms of Dice, Accuracy, and Mean AUC is presented. The results reveal that UNet and ResNet18 performed better; and the tumour class had the highest accuracy and dice coefficient. These results motivated us to select the UNet model for a prostate cancer Gleason grading system, and also, to perform a study for an improved UNet network.

### 6.1 Segmentation of multi-class BC histopathology images related work

Segmentation of the different breast tissue components was studied earlier using computer vision methods; a study based on spectral analysis is presented by (Khan et al., 2013). In this work, segmentation is based on color analysis and separation combined with graph-based spectral segmentation for the determination of histological structures in H&E images, which is an example of a traditional computer vision

method approach. In the work by (Su et al., 2015), region segmentation of breast cancer is analyzed by a deep learning neural model, by means of a 7-layer DCNN trained to perform segmentation of different tissue regions. Other models have been used for BC analysis, like AlexNet, GoogleNet, ResNet, and VGG16. In the work by (Ehteshami Bejnordi et al.), BC is analyzed based on deep learning methods for the detection of lymph node metastases. Tumour segmentation has been explored in the work by (Priego-Torres et al., 2020); a DL model for segmentation of DCIS, IDC, LCIS, and ILC was proposed based on an architecture formed by an encoder and decoder network. The authors also prepared a breast cancer image dataset for the analysis. (Khalil et al., 2022a) presented a segmentation of metastatic foci model based on different neural models like UNet, SegNet, and a FCN. Similarly, the work by (Benmabrouk et al., 2022), the tissue sections are segmented by a UNet model, which was trained by a dataset prepared by a computer vision method to prepare the tumour region masks by colour detection and colour separation. Also, this fact indicates the lack of annotated data for the segmentation task. In more recent works, the combination of different models have been explored (Demir, 2021). In the work by (Liu et al., 2023), segmentation of invasive carcinoma regions is obtained by a combination of multiscale morphological features with a deep learning model. From this review, deep learning methods have better performance over traditional methods (Khan et al., 2013). Early DL approaches were based on convolutional neural networks. Next, some known models like AlexNet, VGG, ResNet, and UNet were used for accurate segmentation of BC tissue. More recent approaches are the combination of different models, as well as traditional computer vision and image processing techniques.

## 6.2 Materials

The BCSS dataset (Amgad et al., 2019), previously described in Chapter 4, was used in this task. It consists of 150 whole-slide images (WSIs) along with corresponding annotations for 22 different tissue classes. To train the deep learning model, image patches of size  $256 \times 256$  were extracted. Following data augmentation—including rotations and horizontal/vertical flips—over 10,000 patches were generated.

Fifteen WSIs, not used during training, were randomly selected and reserved for testing. As a point of comparison, in the work by (Su et al., 2015), a dataset of 92

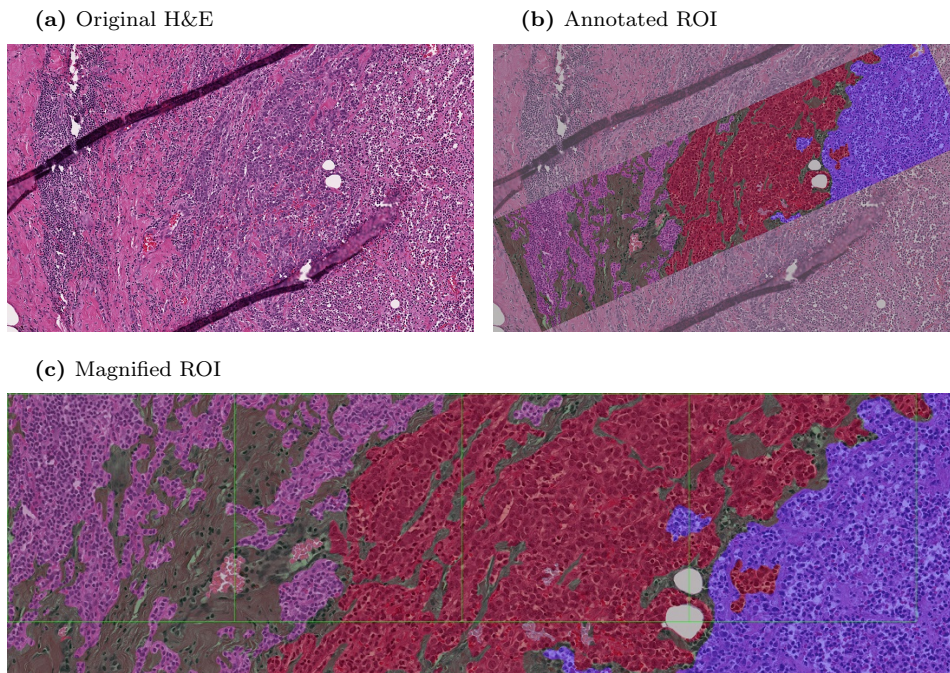
WSIs was used, with 75 images for training and 17 (approximately 18%) for testing. In our case, 10% of the total dataset, i.e., 15 WSIs, were used for testing. During testing, each WSI was divided into  $256 \times 256$  patches; after inference, the full-size WSI and its corresponding segmentation mask were reassembled.

Table 6.1 summarizes the original dataset and the extracted patches used for training. Figure 6.1 shows an example of a pathologist-labelled image used in this experiment. For simplicity, only five tissue classes were selected for segmentation: tumor (red), inflammatory (purple), necrosis (blue), stroma (green), and other (gray). Figure 6.1 displays four sample patches extracted from the same image. Additionally, patches with a 126-pixel overlapping stride were included in the dataset (not shown).

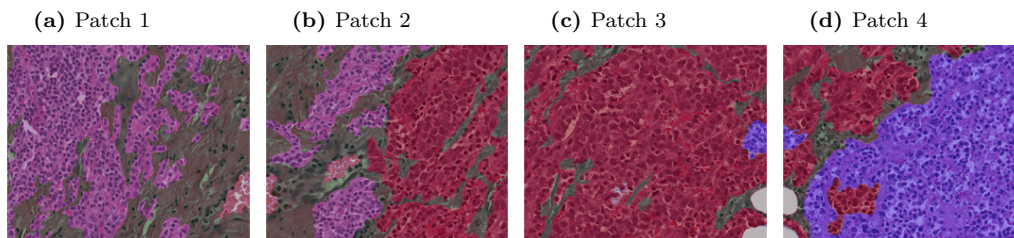
The next section describes the training process of the models.

<b>WSI from BCSS</b>	
Description	Parameter
Total WSI	150
Magnification	40 $\times$ , 20 $\times$
Maximum image size	6813 $\times$ 7360 pixels
Minimum image size	1036 $\times$ 1222 pixels
Mean ROI	1.18 mm <sup>2</sup>
Total Classes	22
<b>Extracted Patches</b>	
Description	Parameter
Patch size	256 $\times$ 256 pixels
Image resize	1/4
Data Augmentation	+90,+180, flip up/down, flip left/right
Total patches	12,930
Training set	10,736 (patches)
Validation set	2,194 (patches)
Test set	15 WSI images
Total Classes	5

**Table 6.1:** Characteristics of the dataset used in this work, the above table indicates original WSI parameters provided by Breast Cancer Semantic Segmentation (BCSS) challenge and the lower one describes patches extracted for training the different networks.



**Figure 6.1:** Figure (a) is one H&E image example taken from BCSS challenge dataset. Figure (b) is a rectangular region of interest (ROI) annotated by the specialists. The tissue regions selected in this work are tumour (red), stroma (green), inflammatory (purple), necrosis (blue), and other (gray). Magnification of the region of interest with lines that indicate patches that were extracted is presented in (c). Notice that tumor region is easier to distinguish, whereas stroma and necrosis are irregular regions and harder to distinguish.



**Figure 6.2:** Four representative images extracted from previous image ???. After applying data augmentation patches are used to train the DNN model. The image shows separated patches, but also patches with an overlap gap of 126 pixels were extracted.

## 6.3 Method

Three different architectures were evaluated for the segmentation task: MobileNet V2 (Howard, 2017), Resnet (He et al., 2015), and UNet (Ronneberger et al., 2015). As described in the related work section, these three models are examples of some of the relevant models used for BC tissue analysis. The models selected for this research were chosen based on the fact that we trained them with local computer resources, for that reason UNet which can be successfully trained with a limited number of images (Ronneberger et al., 2015), and MobileNet, that is an efficient model created for mobile applications (Howard, 2017). Also, due to the quality and its precision (Aljuaid et al., 2022), as well as for the efficient training process, ResNet was selected. However, these models were created for a classification task, and they needed to be modified for the segmentation task. First, the final classification layer (fully connected) is removed and is replaced by a convolutional layer. Next, to recover the original image resolution, a decoder upsampled the image until the original resolution was restored. Then, each pixel has a class prediction to perform semantic segmentation.

We present a brief description of the models, and the results of the training process are also described.

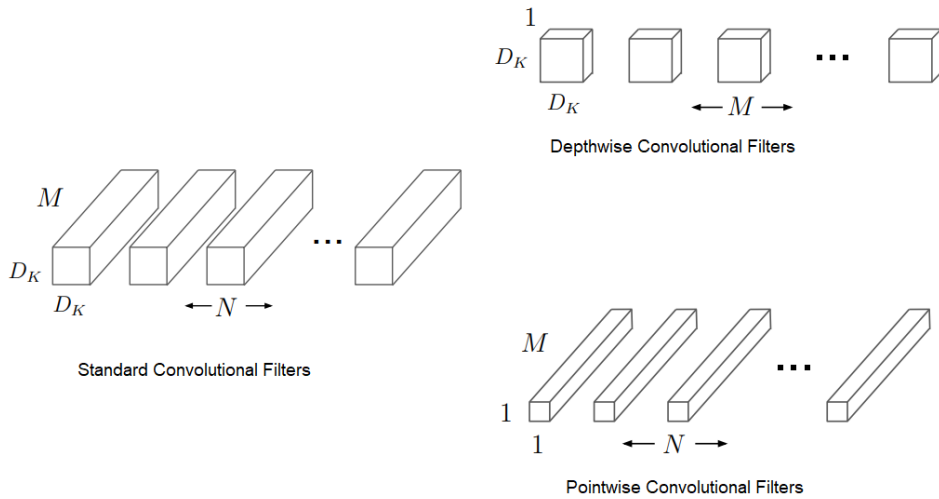
### 6.3.1 The DL models

#### MobileNet V2

MobileNet V2 is a network designed to reduce the computation of the convolution operation so it can run properly on a mobile device. It was developed to separate the normal convolution into a two-step operation, the depth-wise convolution and a point-wise convolution. This is described in Figure 6.3,

In a normal convolution with kernel  $k$  of size 3 with  $c = 3$  channels from a feature map of input  $M \times N$ , the convolution filters are applied  $c$  times. In a depthwise convolution, a single filter is applied to each input channel as seen in the upper part of Figure 6.3. For the pointwise convolution a  $1 \times 1$  convolution is applied that combines the outputs of the depthwise convolution; at the end, the total amount of operations can be reduced between 8 to 9 times. Applications of this network can be seen in face detection implemented on mobile cameras.

(a) Depth and Point Wise convolutions



**Figure 6.3:** The Convolution is reordered in two separate operations, depthwise and pointwise that combined results in the same  $3 \times 3$  kernel convolution, image adapted from (Howard, 2017)

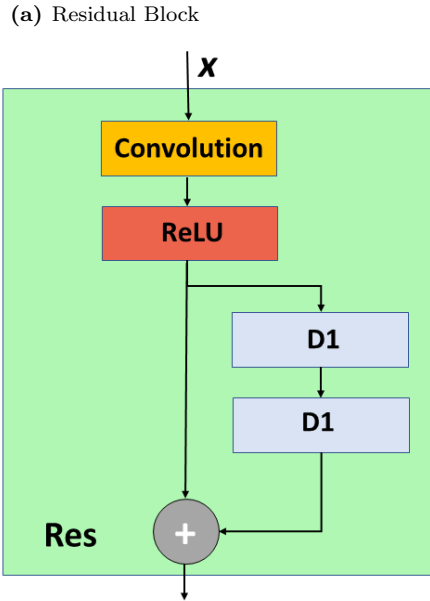
## Resnet

Deep networks have demonstrated better performance over shallow networks on complex tasks, however, as the network becomes deeper it is difficult to train (Bengio et al., 2013). The main problem is the vanishing, which means the gradients become smaller and convergence is difficult to achieve. In case it converges, a degradation problem sometimes results in poor performance.

These problems can be solved by the residual block as seen in Figure 6.4 and it is based on a skip connection that bypasses and connects the input and output, which means the block learns the residual instead of learning unreferenced functions and then very deep models can be trained reducing the risk of gradient degradation. It was presented by (He et al., 2015) at the 2015 Conference on Computer Vision and Pattern Recognition.

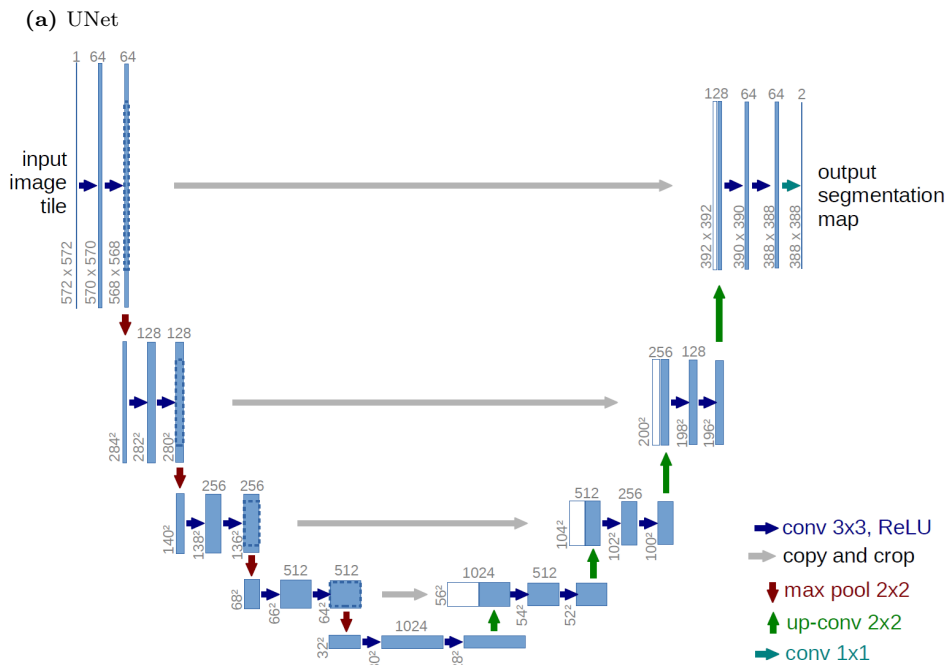
## UNet

The UNet architecture introduced by (Ronneberger et al., 2015) was designed for semantic segmentation, and it is constructed by a cascade of encoding blocks composed of convolutional and max-pooling layers. Next, a cascade of upsampling and deconvolutional layers decodes the compressed representation of the input and performs the prediction of its output. All convolutions are followed by non-linear ReLU activation functions and a pixel-wise softmax activation function at the output of the network. As shown in Figure 6.5, the UNet model describes a ‘U’-shape, hence



**Figure 6.4:** This is a description of the residual operation presented by (He et al., 2015). The block D1 represents a convolution with diation  $D = 1$ . The residual is a block that consists of a skip connection between the input and output, as a result, the overall block learns the residual. The block presented in the figure will be used as part of the modifications in the ablation study.

its name.



**Figure 6.5:** This figure is the Model presented by (Ronneberger et al., 2015). As stated by the author, each blue box is a multichannel feature map, and the number on top of the box indicates the number of channels.

UNet network was selected for an ablation study, and a detailed description of

this popular architecture, as well as different modifications to this architecture will be described in Chapter 8.

### 6.3.2 Training the DL models

The original WSI was resized by  $\frac{1}{4}$ , and the labeled region was divided into  $256 \times 256$  sections to prepare the training dataset. After data augmentation, 12,000 patches for training and validation of the models were obtained. For each of the three models, the hyperparameters were explored systematically during the training until the best accuracy performance was reached. Special care was taken into account to avoid overfitting. The training was performed completely from scratch, and as presented in table 6.2, this process required performing a complete training, validation, and testing process for each parameter's value selected. Initially, the parameters were selected randomly, which resulted in a poor performance. Then, the value was reduced and increased. Also, after the optimal value was found, the rest of the parameters were tuned again for this particular value to guarantee it was indeed the optimal value. For this reason, as well as for the limited computer resources available, only the optimal epochs were not selected for the rest of the experiments, but a lower number of epochs, as it takes many hours to train. Also, because one of the main objectives of this research is to propose a new model, the results of this part of the experimentation were used for the model selection for this proposal. The hyperparameters selected are explained next:

#### Mini-batch

size was tested for values 1, 2, 4, 8 and 16 during different training iterations, and the best results were obtained with a mini-batch of size 8.

#### Loss function

. The binary cross-entropy function was selected for the loss function.

#### L2 Regularization

was tested for values of 0, 0.0001 and 0.0004

## **Shuffle**

means the data is reordered before each epoch, and it was tested for both cases, with shuffle and not shuffle

## **Optimizer**

. Different optimizer methods were evaluated during the training: stochastic gradient descent with momentum (sgdm), root mean square propagation (rmsprop), and adaptive moment estimation (adam). Adam provided the best segmentation results.

## **Momentum**

was tested for values of 0.9 and 0.7.

## **Epochs**

Scenarios of 5, 10, 30, and 50 epochs were tested. Whilst 50 epochs provided the best results it took over 50 hours of training, for that reason 30 epochs were selected during the training phase.

A Computer with processor Intel core i7-7700k, 16 GB RAM, and CPU at 4.20 GHz with GPU with MATLAB 2020b was used to train the models. A summary of the accuracy obtained by the selected hyper-parameters is presented in table 6.2, the number shown in the table is the accuracy.

15 WSI that were reserved and not used during training were separated for testing the model. Each image is also split into  $256 \times 256$  patches, and after applying the DL model the resultant image is reassembled and rescaled to the original resolution image size, but only the labeled region by the pathologist was used for measuring the performance.

## **6.4 Results**

The three models were trained with the tuned selected hyper-parameters. Loss and accuracy measured during training indicate there was no overfitting. No pre-processing or post-processing was applied to the original patches for the metrics calculation. Figure 6.6 illustrates the results of the architectures with each class depicted in different colours, and in Figure 6.7, the correct and incorrect pixels are shown in white and black, respectively. A summary of the results obtained by the

Batch Size					
	1	2	4	8	16
Validation	0.52	0.52	0.53	0.69	0.67
Training	0.40	0.41	0.74	0.72	0.70

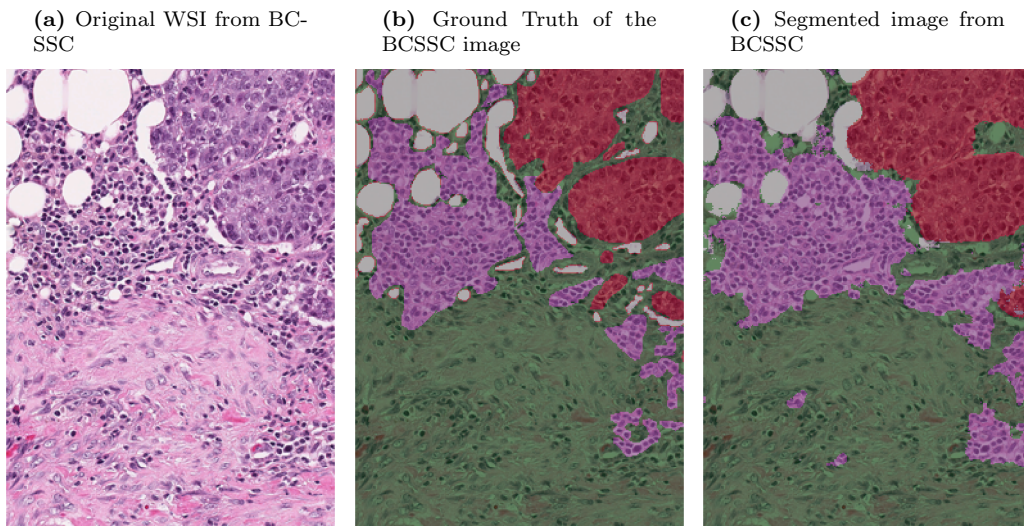
  

Loss function			
	sgdm	adam	rmsprop
Validation	0.69	0.76	0.74
Training	0.72	0.87	0.85

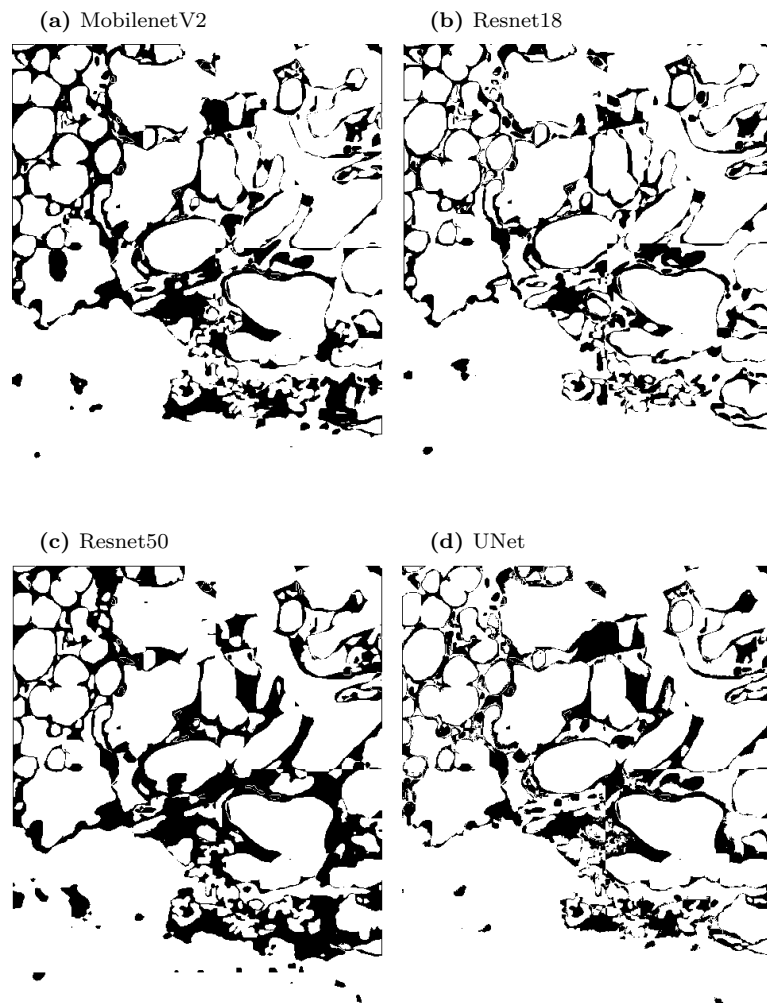
  

Epochs				
	5	10	30	50
Validation	0.63	0.69	0.76	0.90
Training	0.81	0.87	0.88	0.95

**Table 6.2:** Parameters evaluated during network training. These parameters were used for all the selected models, a different training was done for every single parameter, and accuracy was measured for the validation set after averaging on the overall tissue regions



**Figure 6.6:** Illustration of semantic segmentation into 5 classes of the image **TCGA-S3-AA15-DX1** with an overall accuracy of **0.8297**. (a) Original Image (b) Ground Truth (GT) indicated by shading (Tumour = red, Stroma = green, Inflammatory = purple, Necrosis = blue, Other = gray). (c) Segmentation with UNet.



**Figure 6.7:** Pixel to pixel comparison of GT and segmentation for: (a) MobilenetV2, (b) Resnet18, (c) Resnet50 and (d) UNet. Correctly classified pixels are labelled white and incorrect are black.

<b>Baseline</b>			
	<b>Mean AUC</b>	<b>Dice</b>	<b>Accuracy</b>
Overall	0.945	0.888	0.799
Tumour	0.941	0.851	0.804
Stroma	0.881	0.800	0.824
Inflammatory	0.917	0.712	0.743
Necrosis	0.864	0.723	0.872
Other	0.886	0.666	0.670
<b>Mobilenet V2</b>			
Overall	<b>0.951</b>	0.795	0.763
Tumour	<b>0.961</b>	0.873	0.837
Stroma	0.918	0.809	0.743
Inflammatory	0.967	0.839	0.875
Necrosis	0.972	0.812	0.678
Other	0.938	0.644	0.682
<b>Resnet18</b>			
	<b>Mean AUC</b>	<b>Dice</b>	<b>Accuracy</b>
Overall	0.950	0.827	<b>0.841</b>
Tumour	0.947	0.876	0.837
Stroma	<b>0.924</b>	0.823	0.760
Inflammatory	0.978	0.846	<b>0.944</b>
Necrosis	0.963	<b>0.883</b>	<b>0.876</b>
Other	<b>0.962</b>	<b>0.706</b>	<b>0.788</b>

**Table 6.3:** Comparison of the Baseline results (Amgad et al., 2019) and the results obtained in this work. Best results are highlighted in **bold**.

<b>Resnet50</b>			
	<b>Mean AUC</b>	<b>Dice</b>	<b>Accuracy</b>
Overall	0.950	0.788	0.763
Tumour	0.952	0.867	0.857
Stroma	0.916	0.779	0.668
Inflammatory	0.977	0.833	0.938
Necrosis	<b>0.978</b>	0.808	0.713
Other	0.926	0.653	0.640
<b>UNet</b>			
	<b>Mean AUC</b>	<b>Dice</b>	<b>Accuracy</b>
Overall	0.941	0.789	0.737
Tumour	0.960	<b>0.890</b>	<b>0.910</b>
Stroma	0.923	<b>0.831</b>	0.790
Inflammatory	<b>0.979</b>	<b>0.856</b>	0.830
Necrosis	0.936	0.808	0.745
Other	0.906	0.558	0.410

**Table 6.4:** Comparison of the Baseline results (Amgad et al., 2019) and the results obtained in this work. Best results are highlighted in **bold**.

different networks is presented in Tables 6.3 and 6.4. The metrics selected were Dice, Accuracy, and mean AUC. It can be observed that two networks (ResNet18 and UNet) outperform the baseline results reported by the challenge organizers (Amgad et al., 2019). The inflammatory and tumor regions, which include lymphocytes, plasma cells, and cancerous tissue respectively, achieved an accuracy greater than 0.9. In contrast, necrosis and stroma regions remained below this threshold. Our results suggest that necrotic areas are sometimes misclassified as stroma, possibly due to their similarly irregular shapes, which are characteristic of stromal tissue. The "other" category, which encompasses out-of-focus regions, noise artifacts, blood, fat, and other non-specific tissue, showed the lowest accuracy. This is likely due to the heterogeneous and noisy nature of the content within this class.

## 6.5 Summary

In this chapter, semantic segmentation of breast cancer (BC) images was studied for a better understanding of deep learning approach. The hyperparameters from three DL models were explored and adjusted to achieve optimal performance for the segmentation of BC tissue components. The results suggest that further improvements can be achieved by increasing the number of epochs, applying a higher data augmentation, and incorporating extra overlapping regions during patch extraction. Particular interest was given to the UNet model due to its performance compared to the other models. The results obtained with the UNet have encouraged us to further explore and this particular architecture, which will be discussed in detail in Chapter 8.

## Chapter 7

# Gleason Grading assignment for prostate cancer

This Chapter presents an experiment implemented to estimate the Gleason Grade (GG) on prostate cancer images. The approach is based on a multi-class segmentation task for the different GG regions, and a UNet model was selected, after the favorable results obtained in the previous experiments. Grading the severity of prostate cancer efficiently is a relevant task for early cancer detection. The images used in this task were described in Chapter 4, this dataset considers different colour variance from different type of scanners. Manipulation of the original WSI images from the Challenge is quite demanding in terms of computational complexity and memory. A special processing pipeline was developed for this task, which splits the WSI into patches and reconstructs the original image in a lower manageable resolution. The results were successfully submitted to the AGGC challenge contest.

### 7.1 Motivation: Prostate Cancer and the Gleason Grading score

Prostate cancer is quite relevant because it is the most frequently diagnosed cancer among males (Siegel et al., 2023). The estimated cases in 2023 were 288,300 and cancer deaths 34,700 cases worldwide, see also Figure 1.2. These numbers indicate a need for an adequate diagnosis and screening. Grading the disease is also important because it measures cancer severity and it helps the physicians to choose an adequate treatment (Bulten et al., 2020). The two common grading systems are the Gleason

Grading (GG) score and the ISUP (International Society of Urological Pathology) grading system. The GG score is a method based on visual inspection of the cell and the tumour in the microscope. The basic grading assigns the value 3 for cancerous cells that look similar to normal prostate cells, a scale of 4 for cells that look more abnormal, and grade 5 for highly abnormal cells. The final grading is calculated by adding two grades and the range starts from 6 ( $3 + 3$ ), score 7 (for  $3 + 4$  or  $4 + 3$ ), and score 10 ( $5 + 5$ ). Estimating this score is a time-consuming task done by a specialized pathologist. As emphasized by (Huo et al., 2024), this process is subjective and suffers from inter-observability differences between pathologists. For that reason, the use of computer tools based on AI and ML as a support for pathologists appears to be well accepted (Huo et al., 2024). An application called *Paige* was the first AI-based platform for digital pathology approved by the FDA in the US, which provides a solution for GG. However, as stated by (Huo et al., 2024) the challenge to train AI and ML models relies on a well-annotated dataset with a high number of images. Also, models are influenced by colour variation of staining due to the sample preparation done at the pathology laboratory, as well as colour variations from different WSI scanners. In this context, the Automated Gleason Grading Challenge (AGGC) dataset used in this experiment considers the different scanners and colour variations.

The main goal of the challenge is to detect different patterns in the images and generate binary masks to indicate stroma, Normal, Gleason Pattern 3 (GP 3), Gleason Pattern 4 (GP 4), and Gleason Pattern 5 (GP 5) respectively. Figure 7.1 shows an image sample of this dataset, it can be seen the regions annotated by the pathologist.

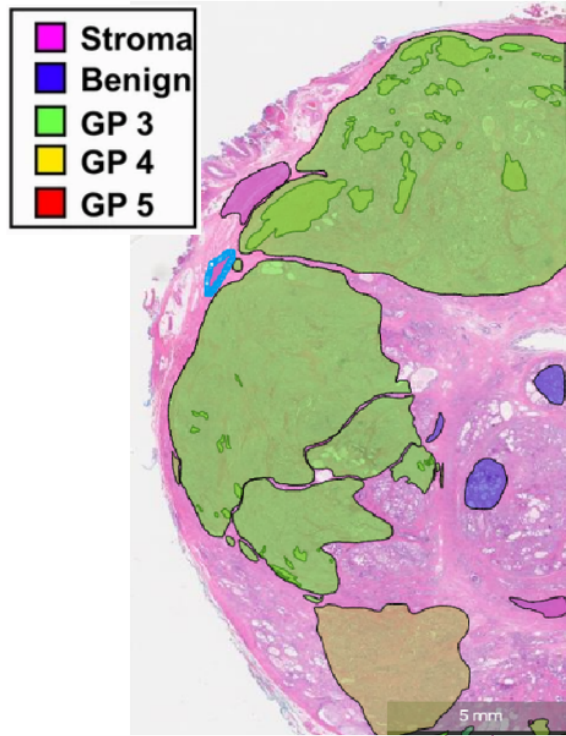
## 7.2 Materials

The Automatic Gleason Grading Challenge (AGGC) dataset consists of 286 training and 128 testing WSI with annotations for prostate cancer, which were divided into 3 subsets. It was obtained from 187 prostatectomy specimens and 156 biopsy specimens from 214 patients, done at the Department of Pathology of the National University Hospital (NUH) in Singapore (Huo et al., 2024). Images are H&E stained, and the dataset is described in Table 7.1.

An important characteristic of this dataset is that it was collected from differ-

<b>WSI from AGGC</b>	
Description	Parameter
Type of scanners	Akoya Biosciences, Olympus, Zeiss Leica, KFBio, Philips
Total prostatectomy specimens	187
Total biopsy specimens	156
Annotation area (prostatectomy)	22,148 $m^2$
Annotation area (biopsy)	2,224 $m^2$
<b>Annotation on prostatectomy per class</b>	
Area	$mm^2$
Stroma	3,816
Benign	3,182
GP 3	5,517
GP 4	8,889
GP 5	744
<b>Annotation on biopsy per class</b>	
Area	$mm^2$
Stroma	238
Benign	781
GP 3	278
GP 4	433
GP 5	494

**Table 7.1:** Characteristics of the AGGC dataset used in this work, the above table indicates WSI parameters of the original AGGC WSI images (Huo et al., 2024).



**Figure 7.1:** This is a sample image from the AGGC, source: (Huo et al., 2024). The annotated region selected by the pathologist is shown in the corresponding color. Purple for stroma, blue for benign, green for GP 3, Yellow for GP 4 and red for GP 5.

ent acquisition scanners, not previously considered in other datasets. The images were initially acquired using Vectra Polaris from Akoya Biosciences with bright-field imaging at  $0.5 \mu\text{m} \times 0.5 \mu\text{m}$  per pixel resolution. The dataset was presented at the Automatic Gleason Grading Challenge (AGGC) contest in 2022 to promote the development of AI models for automated GG. The images came with annotations for the different tissue regions and the GG scales: stroma, benign tissue, GP 3, GP 4, and GP 5, as seen in table 7.1. It is a clear imbalance between classes, as GP5 has the smallest annotated area. The total annotation area is  $22,148\text{mm}^2$  on 187 prostatectomy specimens and  $2,224\text{mm}^2$  on 156 biopsy specimens. Digitized images passed through a quality revision process by a processing tool called A!MagQCout, which tests out-of-focus, low contrast, saturation, artifacts, and texture uniformity. A cloud-based annotation platform named A!HistoClouds is used by pathologists to annotate the regions manually.

A preliminary analysis of the data was performed to improve our understanding of the GG assessment task, and the complete AGGC dataset was observed and statistically analyzed before training the model. Valuable statistical information for each image was revealed: the total labeled area, the non-labeled area, and the labeled

class distribution per image. This initial inspection indicated the total annotated area could generate nearly  $\frac{1}{2}$  million  $256 \times 256$  patches to train a DNN without any data augmentation; for that reason, a selective choice of the patches was needed. Also, the inspection revealed patches not completely annotated; for that reason, the patches were extracted based on certain annotation conditions, for example, the percentage of labeled area in the image patch, which means that only patches that satisfy the condition were extracted. Also, a particular WSI from any of the three subsets was selected for patch extraction. This was done according to the representative statistics.

Another assumption was that for overlapped annotated regions, according to (Huo et al., 2024), certain pixels might have been labeled with more than one class; the priority was assigned to regions with the highest GG, for example, GP5 was the priority, next by GP4, and finally, by GP3. This was due to the lack of enough annotated regions for the highest grades as revealed in the preliminary inspection.

Finally, from the original AGGC dataset, two datasets of patch images of size  $256 \times 256$  Pixels were extracted to train the UNet segmentation model based on the considerations, these datasets are described in the next subsections.

### **7.2.1 Extracted dataset 1**

The first dataset was extracted based on balanced classes (i.e. equal number of images per class). However, as there was a lack of images from the GP5 class, data augmentation was enhanced over this particular class for that purpose. Only images from subset 1 were selected, the reason for this decision is because it is the most representative subset in terms of statistics. In this dataset, no consideration was taken for the criteria of annotated area (i.e., patches with low annotated area were allowed). At the end, 19,578 training images and 4,895 validation images were obtained under this approach.

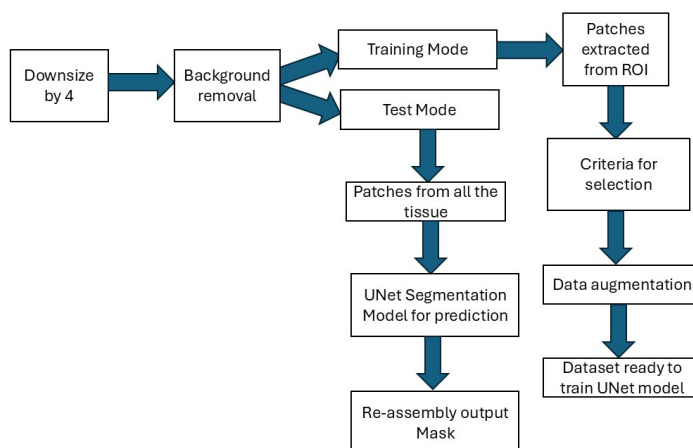
### **7.2.2 Extracted dataset 2**

For this dataset, different considerations were explored. Images from all subsets were selected with a criterion of over 90% of the annotation over the patches. Under the assumption that the performance could be improved if trained with more images, 80% of the images from the training dataset were used (the previous approach only

utilized the 25% from subset 1). This training dataset consists of 25,753 training images and 8,164 validation images.

### 7.3 Method

As a reference, the work by (Huo et al., 2024) presents a pipeline approach for automated analysis and GG based on an AI classification model that operates on patches extracted from original images, and different networks were trained: ResNet50, VGG16, and NasNet Mobile. The classification layers from the original networks were replaced with a weighted classification layer, and after a majority voting selection, the final grading was estimated. Two important conclusions are determined by the authors: AI is a support tool for pathologists that reduces the average examination time, and in some cases, accuracy is increased. In our work, we propose a multiclass segmentation approach for GG estimation, under the assumption that every pixel within the image belongs to a different tissue class or grade of the Gleason scale. Then, a multiclass segmentation model was proposed and trained for GG estimation. UNet neural network was selected and trained with the images prepared and extracted from the two datasets already prepared. UNet was chosen as a result of our previous experiment for the BC semantic segmentation task. A special pipeline for processing WSI was implemented, which is described next.

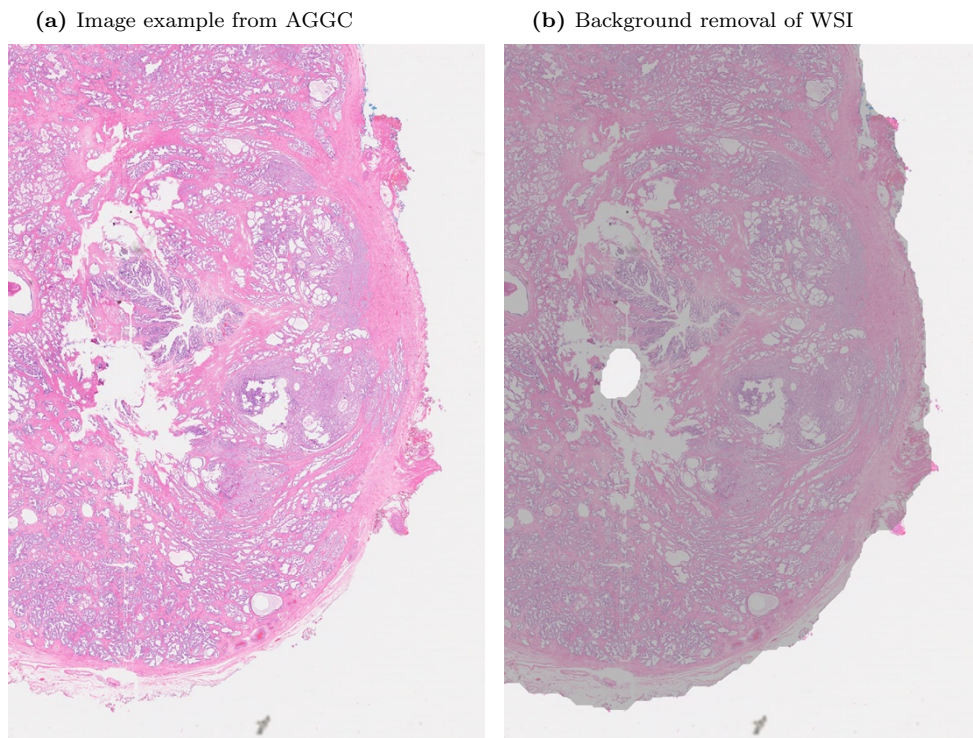


**Figure 7.2:** Pipeline developed for the AGGC assessment. There are two operating modes, one is to generate a training dataset, and the test mode is to estimate the GG over the test set of images.

### 7.3.1 Pipeline

The proposed pipeline automatically processes the complete set of WSI images. The pipeline operates in a training and testing mode. In the training mode, it extracts patches from the selected ROI of the images based on a certain criteria and in the test mode, all the tissue area from the WSI is split into patches and passed to the model, the segmented output image is reassembled to generate a segmentation mask that indicates the GG region.

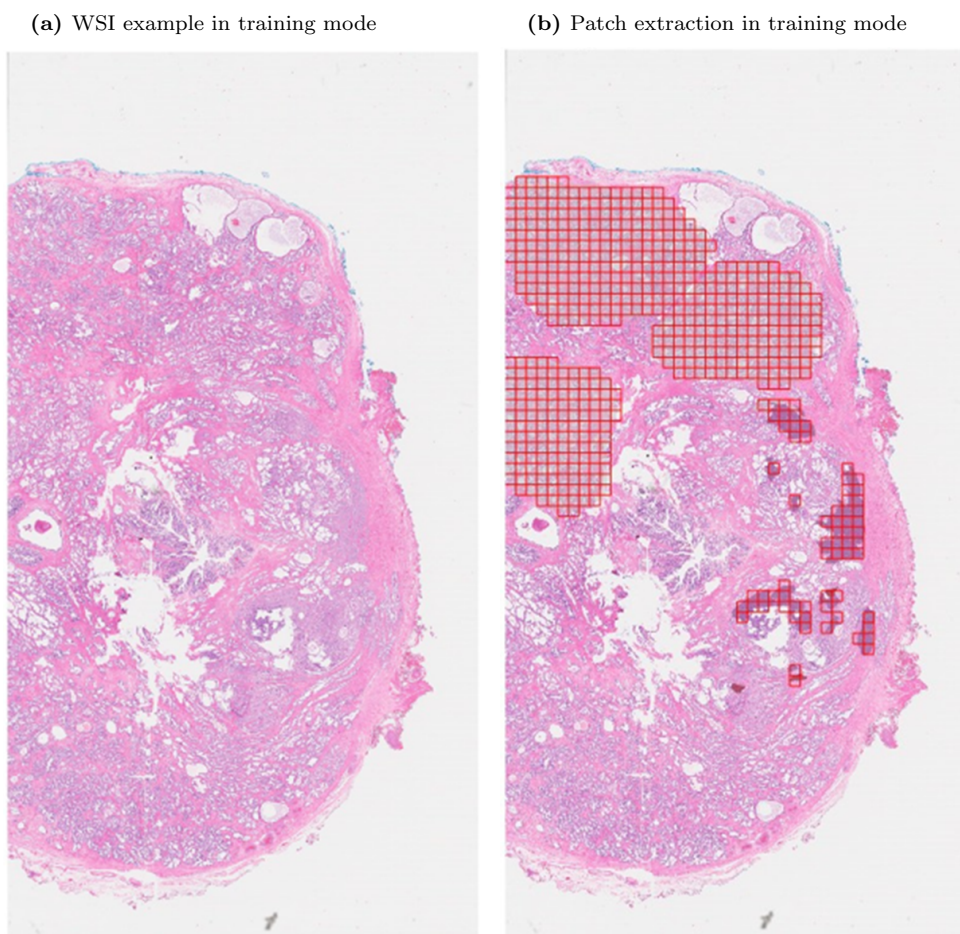
Initially, the images are downsized by 4 to visualize and manipulate them properly. It is important to mention that the images must be downsized by 10 times for submission within the contest. Next, the background region outside visible tissue is removed, and only the tissue area is passed to the splitting process. The background is calculated based on the Saturation and Hue of the WSI. Saturation is variable among the slides, especially within different scanners, and thus, a variable threshold based on Otsu's binarization was used. Hue is not affected, and values between 0.6 and 1 were used. Post-processing morphological functions are applied to obtain the final segmentation of the background mask as seen in Figure 7.3 (b).



**Figure 7.3:** Example of background removal, the image (a) is a full WSI image and (b) is the ROI selected for the tissue region, background removal is done automatically.

In the training mode, patches were extracted only from the annotated area as

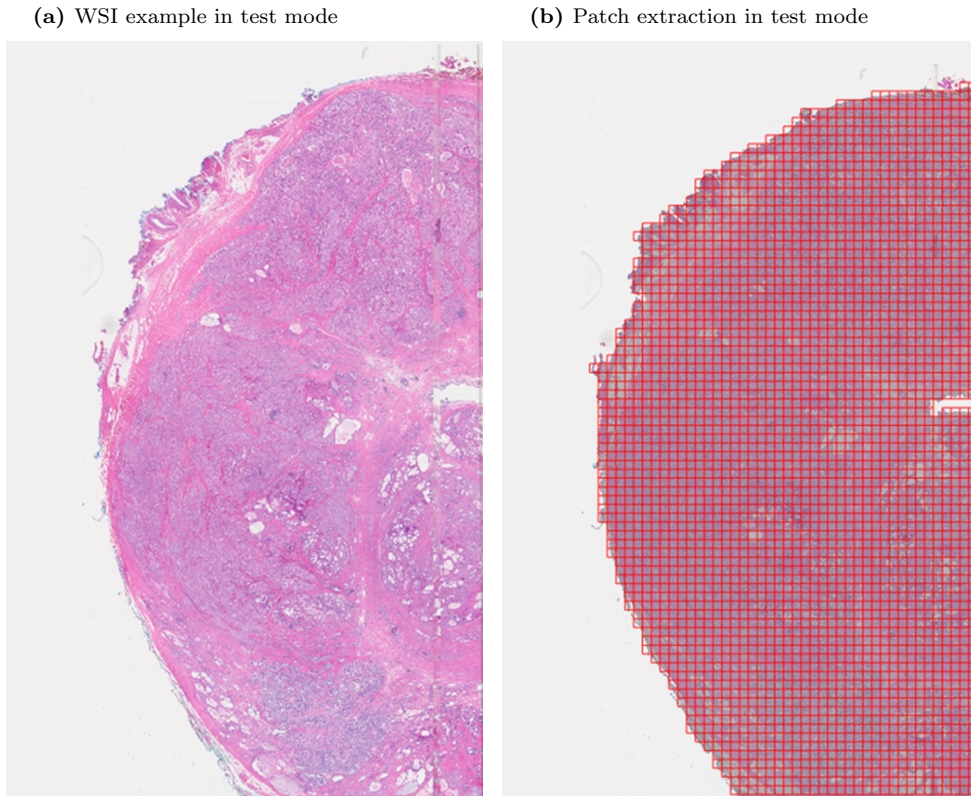
seen in red squares in Figure 7.4. In the test operating mode, the full image is processed. There is no information about the annotation area (the labeled region from the test images is unknown); for that reason, all the patches from the full tissue area are passed through a segmentation model, and an output segmentation mask is assembled; all the test images must be processed in that way. In Figure 7.5, we can see the WSI split into  $256 \times 256$  patches; all the tissue must be processed; for that reason, the testing mode is time-consuming. An efficient processing pipeline was achieved based on a low-resolution visualization combined with the high-resolution processing of the full test set.



**Figure 7.4:** A WSI from the training set (a), it was labeled by the pathologist, and only the labeled region was used to extract patches (b) during the pipeline training mode.

### 7.3.2 Training the DL model

With the use of the proposed pipeline, both datasets were successfully trained, although previous attempts to train and evaluate the test dataset crashed the computer, due to memory requirements. This was also another motivation for the selec-



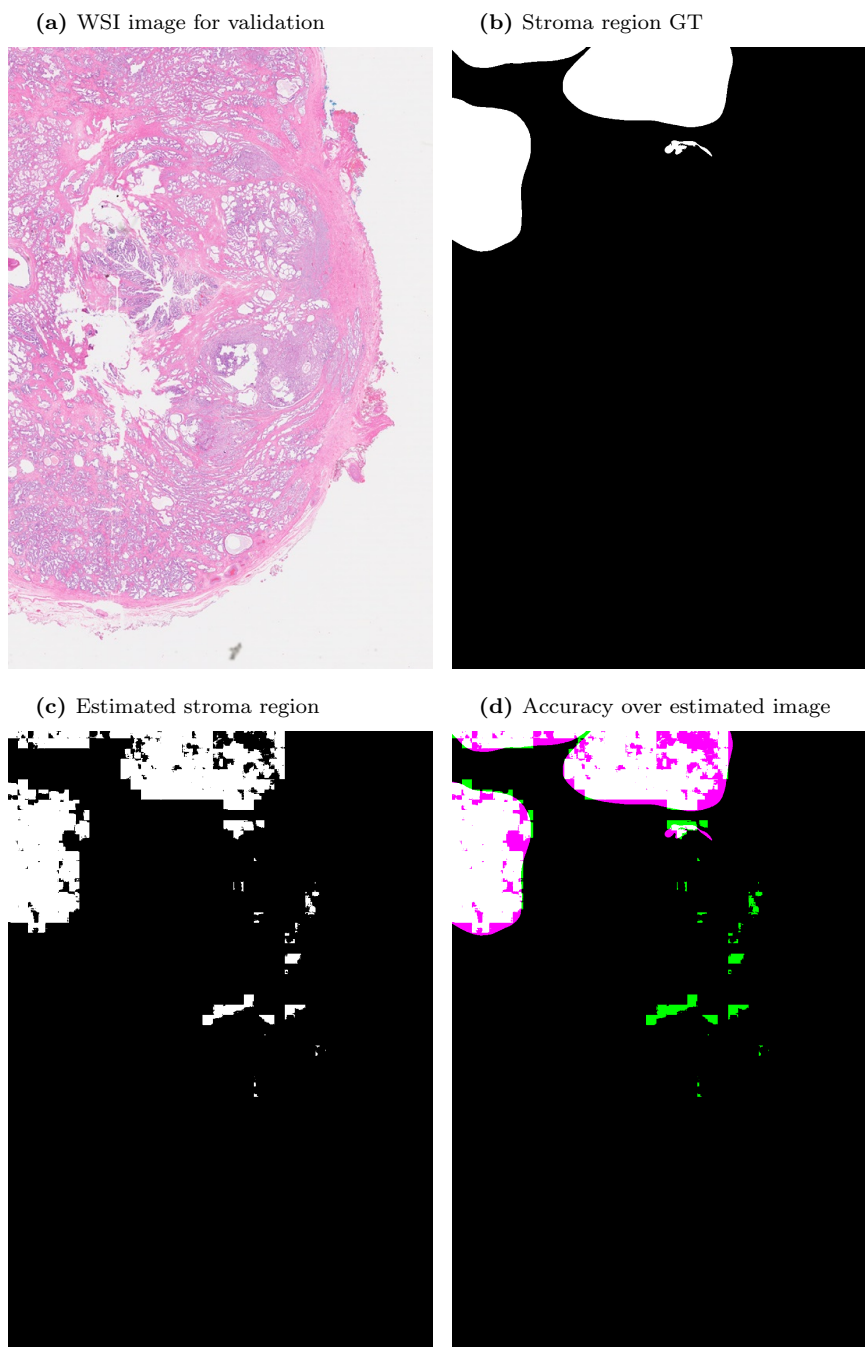
**Figure 7.5:** A WSI from the test set (a), the labeled region is unknown by participants and the full image needed to be analysed and all the extracted patches are processed by the model.

tive training dataset generation.

The hyperparameters selected and tuned for the BCSSC described in the previous chapter were also used and tuned during the first training attempts. The ADAM optimization method was selected, learning rate of 0.0001, Minibatch size of 16, Max epochs 7, and shuffle the data every epoch. Seven representative images from the training set were reserved for evaluating the model by the Accuracy and F1 over the annotated area given by the pathologist, as a Ground Truth. This was used for our test, and the results are presented in table 7.2.

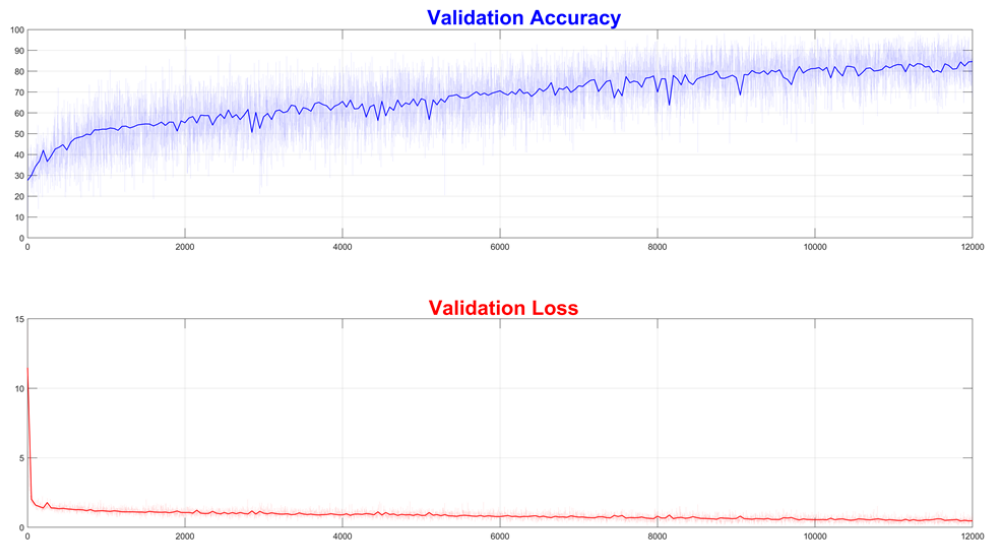
## 7.4 Results

The performance of the proposed approach was evaluated by accuracy and F-1 on the 7 selected test images without being used during training. From the seven selected images, Accuracy and F-1 results are presented in table 7.2 for the two trained datasets. From the table 7.2, it can be seen there is an improvement with the second dataset in all classes, except for stroma, and GP5, which is the class



**Figure 7.6:** An example of the GG test image processed by the model, a WSI (a) for validation from the training set with the corresponding Ground Truth GT for class stroma in (b), the segmented mask processed by the method is in (c) with the estimated stroma region, after this is validated with the labels by the pathologist, the result is in (d). TP is in white, TN in black, FP in green, and FN in pink.

(a) Validation accuracy and validation loss curves obtained in the training (dataset 1)



**Figure 7.7:** Validation accuracy is shown in light blue (dark blue is an average weighted curve), and validation loss similarly in red.

(a) Validation accuracy and validation loss curves obtained in the training (dataset 2)



**Figure 7.8:** Validation accuracy is shown in light blue (dark blue is an average weighted curve), and validation loss similarly in red. It can be seen a faster convergence compared to dataset1.

Dataset 1					
	Stroma	Normal	GP3	GP4	GP5
Accuracy	0.8882	0.6833	0.6033	0.5654	0.4574
F1	0.9395	0.8072	0.6897	0.6766	0.0021
Dataset 2					
	Stroma	Normal	GP3	GP4	GP5
Accuracy	0.8367	0.7793	0.6538	0.7020	0.4578
F1	0.9091	0.8744	0.7509	0.8067	0.0013

**Table 7.2:** Measurements of Accuracy and F1 was done over a set of selected images from the labeled by pathologists.

with the lowest annotated area. Only in the stroma class, F-1 measurement was over 90%, and also normal tissue has a high F-1, notice the high increment from 0.8072 to 0.8744, and also with GP4 from [0.5654, 0.6766] to [0.7020, 0.8067] for both metrics. This result indicates an improvement due to the assumptions made in the second training. On the other hand, the grade GP5 showed no improvement with F-1, and moreover, it still obtained the lowest performance. This implies that a high imbalance has a negative effect. Important conclusions were obtained from this experiment and are expressed in the summary chapter.

## 7.5 Summary

A semantic segmentation task was implemented to estimate the Gleason grade in prostate cancer images. Manipulation of gigabyte WSI was challenging, and a special pipeline to efficiently process the WSI testing images from the AGGC was implemented. From the original images, two datasets were derived, corresponding to weakly and strongly labeled annotations. The results indicate a level of improvement when annotations are of the strong type. Due to the challenges in processing WSIs and the scarcity of strong annotations, a different approach should be explored.

The results obtained thus far have motivated further exploration of the UNet architecture and propose improvements to the original model, which will be described in the next chapter.

## Chapter 8

# UNet analysis and modifications for an improved architecture

This chapter presents the analysis performed on the state-of-the-art UNet architecture for the multi-class segmentation task. Several modifications to the original model were implemented, and the overall performance was evaluated after each change. These modifications were based on two common techniques reported in the literature: residual connections and dilated convolutions. This is an ablation analysis of this architecture, with the main objective of having a better understanding of the insights of this model, and also to propose improvements in terms of its accuracy. This chapter describes this analysis and presents its results. The chapter starts with a general description of the UNet network, including its motivation, history, and architectural characteristics. A related work describing examples of successful improvements over this model is given. Next, the description of the modifications selected for the study is presented. Over 16 experiments were implemented with the changes, some associated with the reference-reported architectures. The dataset from the BCSSC was used for this study. The results provided valuable insights into the performance of the UNet model. As a result, a novel block was proposed and implemented, and its performance was compared with other variants, with an improvement in accuracy, particularly in tumour regions. The findings of this study, along with the proposed model, were also published in a scientific paper (Ortega-Ruíz et al., 2024).

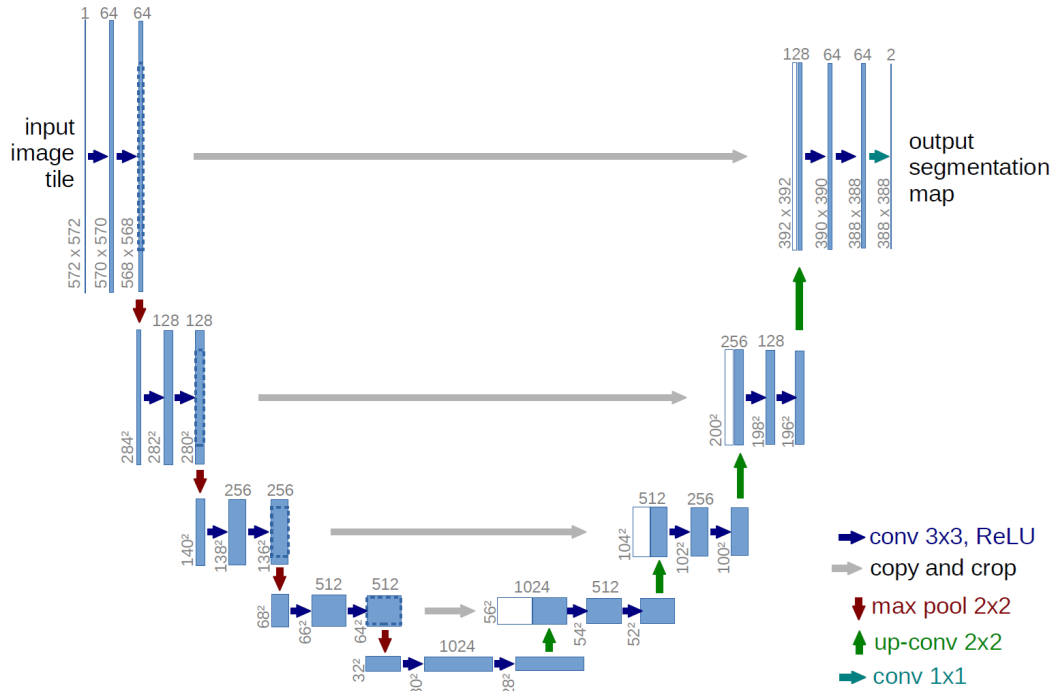
## 8.1 UNet an state-of-the-art segmentation network

### 8.1.1 UNet motivation

UNet was presented by (Ronneberger et al., 2015) in 2015 at the Medical Image Computing and Computer-Assisted Intervention (MICCAI) conference. Since then, numerous studies have been published recently presenting improvements in UNet implementation in medical image analyses. A PubMed query reveals nearly 5,000 scientific papers have been written about this network since 2018. As summarized in the review study by (Krithika alias AnbuDevi and Suganthi, 2022), this model has been used to segment the Brain, Retinal Vessel, Lung, Liver, Lymph nodes, Prostate, and Esophageal. The enormous interest in this network indicates its relevance within the digital pathology research community. The next paragraph presents a brief history of this network and a detailed description of its architecture.

### 8.1.2 Deconvolutional Neural Networks (DeCNN)

UNet is a Deconvolutional Neural Network type (DeCNN). It can be understood as an end-to-end architecture based on two operations, convolutions at the encoder step and deconvolutions at the decoder. It has been widely used in the segmentation task and image generation task. The encoder-decoder model captures context by downsampling the input and extracting features at various levels. The decoder path upsamples the feature maps to recover spatial details and provide a segmentation map. Multiresolution analysis on images was presented many decades ago, and Pyramids and Quadrees are examples of such analysis at different levels of detail based on computer vision techniques (Adelson et al., 1984a). The precursor of DeCNN is the Neocognitron proposed by Fukushima (Fukushima, 1975). It is based on multiple layers that progressively extract and recognize more abstract features from the input image. Also, Fully Convolutional Networks (FCN) incorporate at their output a spatial map where each pixel corresponds to a predicted class or value from its input, which is called spatially dense prediction. The work presented by (Shelhamer et al., 2017) is a DeCNN-type network that incorporates a dense prediction layer and a fully connected Conditional Random Field for the semantic segmentation task. The UNet model presented by (Ronneberger et al., 2015) is a DeCNN-type network as it uses a transposed convolution (deconvolutional) in the decoder. It is an efficient architecture that requires a small training dataset,



**Figure 8.1:** The UNet architecture as proposed by Ronneberger (Ronneberger et al., 2015). It consists of encoder-decoder sections and contains a series of downsampling steps (encoder) obtained by convolutions and downsampling operations, and then upsampling steps (decoder) formed by upsampling plus convolution operations.

providing precise pixel-level predictions for segmentation.

### 8.1.3 UNet architecture

The UNet model is shown in Figure 8.1, formed by a contracting path and expansive path, as called in the original paper by (Ronneberger et al., 2015); these are the corresponding encoder and decoder. The contracting path is a convolutional network that consists of two  $3 \times 3$  convolutions each followed by a rectified linear unit (ReLU) activation function as described in the blue left-to-right arrow in Figure 8.1 and a  $2 \times 2$  max pooling operation with stride 2 depicted in the red up-to-down arrow. The expansive path is formed by an upsampling of the feature map followed by a  $2 \times 2$  convolution shown in the green down-to-up arrow. There is the same number of decoding blocks as the encoding blocks, describing a *U* shape, hence its name. The key feature of the UNet is that it makes use of residual connections that concatenate the output of the  $i$ -th encoding block with its corresponding decoding counterpart. These residual connections serve two purposes: they mitigate the possibility of encountering vanishing gradient effects, and they increase the chance of exploiting visual patterns that might prove relevant for prediction but that could

have been overlooked by the encoding process (Ronneberger et al., 2015). At the end of the expansive path, there is a  $1 \times 1$  convolution used to map the feature vector to the desired number of classes.

## 8.2 Related work

In this section, we present relevant examples of UNet modifications and improvements over the original architecture (Krithika alias AnbuDevi and Suganthi, 2022; Ummadi, 2022). One of the first modifications of UNet was the inclusion of a residual block (Han, 2017) at the encoder and decoder, presented in the ResUNet network used for liver lesion segmentation. ResUNet achieved first place in the liver tumor segmentation challenge in 2017. The RMS UNet, a residual modification combined with dilated convolutions (Khan et al., 2022; Han, 2017), demonstrated high accuracy over different publicly available datasets. In brain segmentation, the stack Multi-Connection Simple Reducing Net (SMCSRNET) (Ding et al., 2019) is formed with stacked encoder-decoder blocks, improving segmentation with less training time, by using a reduced parameter model. MH-UNet (Ahmad et al., 2021), a multi-scaled network is formed with dense and residual blocks and used for multi-organ segmentation. MI-UNet (Zhang et al., 2021b) is also a modification that includes a procedure called brain parcellation to generate an input to UNet used for brain stroke segmentation under MRI. Spatial weighted UNet is used for 3D CT brain images (Sun et al., 2018) with residual-inception blocks densely connected, which reduces trainable parameters over the MRI dataset. For retinal vessel segmentation, the images require enhanced contrast for accurate vessel detection, and approaches based on residual attention and supervised UNet have been proposed (Yang et al., 2023; Wang et al., 2021). The network MI-UNet (Hu et al., 2019), which consists of two MI-UNets connected into one S-UNet, has been used for Brain Stroke Lesions. For nuclei and cell segmentation, RIC-UNet (Zeng et al., 2019), a network with residual blocks, multi-scale, and channel attention mechanism, demonstrated superior performance over traditional methods. Residual-Dense blocks have also been explored at the bottleneck connection to form D-UNet for lung vessel segmentation (Yuan et al., 2021). Also, generative networks have been explored using the UNet, RDA-UNET-WGAN (Negi et al., 2020) employs a Residual-Dilated-Attention-Gate-UNet as the generator network. The UNet3+, a full-scale skip connections and

deep supervision framework, has also been employed for 6 different architecture segmentation models (Alam et al., 2023a). Another important improvement is the modification of the convolution kernel size with an extended dilation (called Atrous convolution). The segmentation of objects at multiple scales, as well as for weakly supervised semantic segmentation, is presented by (Xu et al., 2021). The Purified and Residual UNet or P-ResUNet (Niu et al., 2022) is based on a Dilated Pyramid Block (DPB) and was used for brain tissue segmentation. This block consists of dilations of sizes  $D = 1, 2, 3$  in parallel.

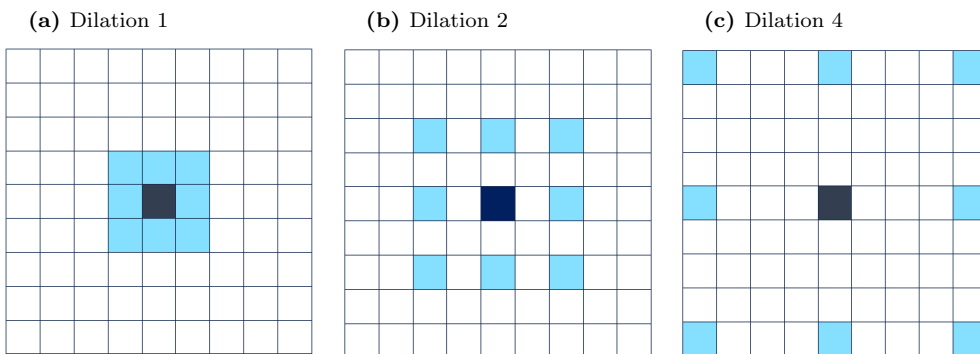
Many UNet modifications have also been implemented for breast cancer image analysis. First, relevant examples with mammograms are presented. The work by (Li et al., 2019) is based on a dense UNet combined with attention gates for mammogram segmentation. (Shen et al., 2019) presented a residual-aided network combined with a UNet to perform a joint segmentation and disease classification. The work by (Alom et al., 2018) for nuclei segmentation is based on a recurrent residual UNet. An architecture called Connected-UNets, using two UNets and additional modified skip connections, showed better visual results in segmenting the mass lesions over mammograms (Baccouche et al., 2021); which has been addressed by different UNet architectures, the ResUNet (Yue et al., 2022) estimates volumetric measurement of breast cancer on MRI. The modification presented by (Hossain, 2022) additional convolutions were added at the decoder, and the ReLu non-linear function was replaced by leaky ReLu for microcalcification segmentation from mammogram images.

For histopathological images, a dense network formed by a modified UNet (Zhou et al., 2019) with dense skip connections to bridge the semantic gap between encoder and decoder features was studied for breast nuclei segmentation (Alam et al., 2023b). A dense and residual network (DenseResUNet) for nuclei segmentation is based on a UNet with dense blocks in the last layers of the encoder with Atrous residual in the conventional skip connections (Kiran et al., 2022).

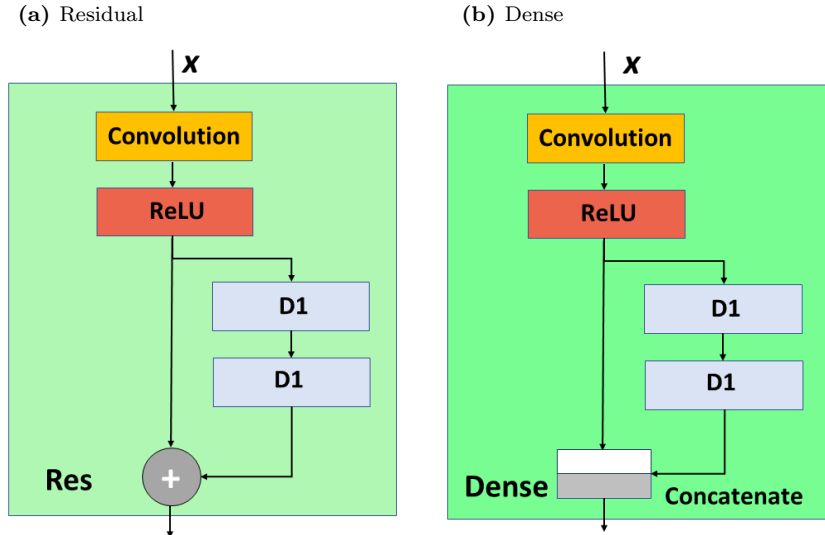
There are possibly hundreds of modifications of the original UNet model, and the examples presented in this section suggest that two important modifications must be analyzed: residual and dilated convolution. These operations will be described in the methods section. Both served as a base for the modifications implemented during the UNet analysis performed in this research.

### 8.3 Dilation, Residual and Dense (DRD) block motivation

The experiments performed on UNet presented in this research were selected from the different UNet versions described in the related work section. After the experimentation, the first results suggested that a deeper network performed better. On the other hand, the little improvement in the results observed after inserting the residual block itself indicated that this operation only supports convergence effects and has no influence on accuracy. In the case of Atrous or dilation convolution, this operation had a positive effect on segmentation accuracy, and the combination of residual+dilation was the best block modification over the original UNet. As a result, a deeper network with extra convolution, as well as an extra residual, was proposed, under the assumption that a deeper network in combination with the powerful residual+dilation might improve performance, but it was initially unsuccessfully trained. It could have been attributed to a convergence or memory problem after the computer had crashed several times during the training. The combination of Residual and Dense overcame the problem, as was also addressed by (Song et al., 2020) in the work for a super-resolution network; it suggests the importance of the combination of Residual and Dense. The proposed block, which we named DRD and will be described in the experiments section (see Figure 8.9), is our main contribution within this thesis research. This modification was also included in the experiments, and its results are compared with the rest of the models.



**Figure 8.2:** Illustration of dilated convolutions (Atrous Convolution) of different sizes and representation as blocks. Each square represents a pixel, the light blue pixels are the kernel size ( $3 \times 3$ ) and the dark blue is the resultant convolution. (a) Dilation  $D = 1$ . (b) Dilation  $D = 2$ . (c) Dilation  $D = 4$ . Image inspired from (Khan et al., 2022).



**Figure 8.3:** (a) The Residual Block introduces a skip connection between input and output; the model learns the difference between input and output (i.e., the residual), which is useful in deeper networks, reducing the vanishing gradient and making it easier to train. (b) The Dense operation combines the two branches with a concatenation, as opposed to the residual.

## 8.4 Methods

Modifications to the UNet architecture were systematically evaluated using the BCSS dataset (Amgad et al., 2019) for a multi-class semantic segmentation task, as previously described in Chapter 6. The same training set of  $256 \times 256$  patches was used for evaluating the UNet modifications, as well as the 15 WSI images for testing. The two main operations selected for the analysis of the UNet are: dilated convolutions and residual connections. Convolution is a fundamental operation that traverses an image and extracts the values of a neighborhood of pixels of a kernel size  $k$ . It has been extensively used for feature extraction. Convolution of an image is a 2-D digital filter, applied over a  $k \times k$ , normally contiguous pixels. A dilated convolution is the same operation over pixels separated by a factor of  $d$ , a larger area is covered as illustrated in Figures 8.2.b and 8.2.c. These dilated convolutions are sometimes called Atrous convolutions (Khan et al., 2022) and they produce an increase in the receptive field, which can be understood as the area in the original image that is seen at the input of the neuron. On the other hand, the essence of residual connections, also known as skip connections, is to add the output of a processing block with the input of that same block. In terms of branches, this is a parallel branch as opposed to a single sequential process, as illustrated in Figure 8.3. Residual connections were introduced at the computer vision and pattern

recognition (CVPR) 2016 conference (He et al., 2015), and have been demonstrated to minimize training errors, overfitting, and vanishing gradient effects (Han, 2017).

No	Name	Description	
1	'ImageInputLayer'	Image Input	images of size $256 \times 256 \times 3$
2	'Encoder-Stage-1-Conv-1'	2-D Convolution	64 filters of $3 \times 3 \times 3$
3	'Encoder-Stage-1-ReLU-1'	ReLu	ReLu
4	'Encoder-Stage-1-Conv-2'	2-D Convolution	64 filters of $3 \times 3 \times 64$
5	'Encoder-Stage-1-ReLU-2'	ReLu	ReLu
6	'Encoder-Stage-1-MaxPool'	2-D Max Pooling	$2 \times 2$ max pooling with stride [2 2]
7	'Encoder-Stage-2-Conv-1'	2-D Convolution	128 filters of $3 \times 3 \times 64$
8	'Encoder-Stage-2-ReLU-1'	ReLu	ReLu
9	'Encoder-Stage-2-Conv-2'	2-D Convolution	128 filters of $3 \times 3 \times 128$
10	'Encoder-Stage-2-ReLU-2'	ReLu	ReLu
11	'Encoder-Stage-2-MaxPool'	2-D Max Pooling	$2 \times 2$ max pooling with stride [2 2]
12	'Encoder-Stage-3-Conv-1'	2-D Convolution	256 filters of $3 \times 3 \times 128$
13	'Encoder-Stage-3-ReLU-1'	ReLu	ReLu
14	'Encoder-Stage-3-Conv-2'	2-D Convolution	256 filters of $3 \times 3 \times 256$
15	'Encoder-Stage-3-ReLU-2'	ReLu	ReLu
16	'Encoder-Stage-3-DropOut'	Dropout	50% dropout
17	'Encoder-Stage-3-MaxPool'	2-D Max Pooling	$2 \times 2$ max pooling with stride [2 2]
18	'Bridge-Conv-1'	2-D Convolution	512 filters of $3 \times 3 \times 256$
19	'Bridge-ReLU-1'	ReLu	ReLu
20	'Bridge-Conv-2'	2-D Convolution	512 filters of $3 \times 3 \times 512$
21	'Bridge-ReLU-2'	ReLu	ReLu
22	'Bridge-DropOut'	Dropout	50% dropout

**Table 8.1:** Definition of the Layers within MATLAB code for a UNET of depth 3, in total, there are 46 layers. The first 22 layers are for the encoder side.

#### 8.4.1 Original UNet Model used in the experiments

UNET of depth 3 was used for the study, and the encoder and decoder were systematically modified and replaced with the reference blocks. Each modification is a complete experiment represented by training and testing the modified UNet architecture and evaluating the performance. The modifications were implemented in the

No	Name	Description	
23	'Decoder-Stage-1-UpConv'	2-D Transposed Convolution	256 $2 \times 2 \times 512$ transposed convolutions with stride [2 2]
24	'Decoder-Stage-1-UpReLU'	ReLU	ReLU
25	'Decoder-Stage-1-DepthConcatenation'	Depth concatenation	Depth concatenation of 2 inputs
26	'Decoder-Stage-1-Conv-1'	2-D Convolution	256 filters of $3 \times 3 \times 512$
27	'Decoder-Stage-1-ReLU-1'	ReLU	ReLU
28	'Decoder-Stage-1-Conv-2'	2-D Convolution	256 filters of $3 \times 3 \times 256$
29	'Decoder-Stage-1-ReLU-2'	ReLU	ReLU
30	'Decoder-Stage-2-UpConv'	2-D Transposed Convolution	256 $2 \times 2 \times 512$ transposed convolutions with stride [2 2]
31	'Decoder-Stage-2-UpReLU'	ReLU	ReLU
32	'Decoder-Stage-2-DepthConcatenation'	Depth concatenation	Depth concatenation of 2 inputs
33	'Decoder-Stage-2-Conv-1'	2-D Convolution	256 filters of $3 \times 3 \times 512$
34	'Decoder-Stage-2-ReLU-1'	ReLU	ReLU
35	'Decoder-Stage-2-Conv-2'	2-D Convolution	256 filters of $3 \times 3 \times 256$
36	'Decoder-Stage-2-ReLU-2'	ReLU	ReLU
37	'Decoder-Stage-3-UpConv'	2-D Transposed Convolution	256 $2 \times 2 \times 512$ transposed convolutions with stride [2 2]
38	'Decoder-Stage-3-UpReLU'	ReLU	ReLU
39	'Decoder-Stage-3-DepthConcatenation'	Depth concatenation	Depth concatenation of 2 inputs
40	'Decoder-Stage-3-Conv-1'	2-D Convolution	256 filters of $3 \times 3 \times 512$
41	'Decoder-Stage-3-ReLU-1'	ReLU	ReLU
42	'Decoder-Stage-3-Conv-2'	2-D Convolution	64 filters of $3 \times 3 \times 64$
43	'Decoder-Stage-3-ReLU-2'	ReLU	ReLU
44	'Final-ConvolutionLayer'	2-D Convolution	5 filters of $1 \times 1 \times 64$
45	'Softmax-Layer'	Softmax	Softmax
46	'Segmentation-Layer'	Pixel Classification Layer	Cross-entropy loss with 'other', 'tumor', and 3 other classes

**Table 8.2:** Definition of the Layers within MATLAB code for a UNET of depth 3. The last 24 layers are for the decoder side and segmentation output.

encoder as seen in figure 8.4 represented in light green blocks and the decoder on the right side as shown in stronger green color blocks. Tables 8.1 and 8.2 show all the encoder and decoder layers of the original UNet of depth 3 as it was implemented in MATLAB code. The tables show the sequence of layers and the total number of convolutional filters.

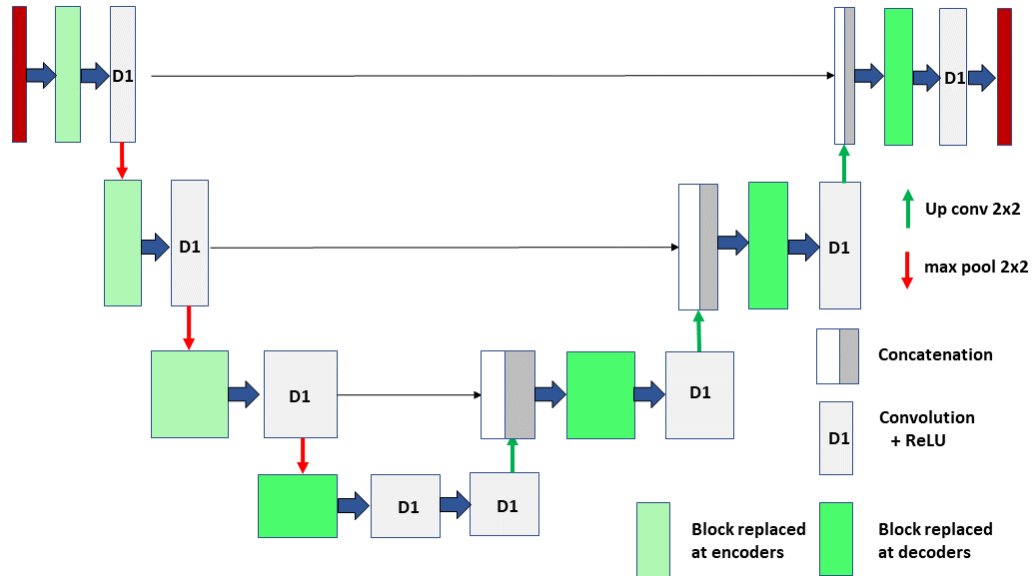
No	Experiment	Param. $1 \times 10^6$	Layers
1	<b>UNet-3</b>	7.69	46
2	<b>ResUNet-i</b>	7.84	62
3	<b>ResUNet-e</b>	8.65	66
4	<b>ResUNet</b>	15.28	75
5	<b>ResUNet-BN</b>	15.28	89
6	<b>P-ResUNet</b>	15.36	96
7	<b>DPB3-i</b>	7.99	62
8	<b>DPB3-e</b>	10.09	74
9	<b>DPB3-a</b>	21.41	107
10	<b>DPB5-i</b>	8.29	74
11	<b>DPB3-i+Res</b>	11.82	80
12	<b>DPB4-i+Res</b>	11.94	86
13	<b>ResUNet-a</b>	13.88	119
14	<b>Series-i</b>	7.84	60
15	<b>Series-e/d</b>	10.93	75
16	<b>RMS-UNet</b>	14.03	85
17	<b>DRD-UNet</b>	15.40	130

**Table 8.3:** List of all experiments implemented in UNet, experiment 17 is our contribution.

#### 8.4.2 Experiments over the UNet

16 different models were tested and evaluated for the UNet analysis for BC histopathology images. The first five architectures were implemented with the inclusion of residual connections (Han, 2017). The next 4 models evaluate the dilated convolution, and next, different combinations of residual and dilation were also tested, as well as the model named RMS-UNet. Two different combinations of dilation in a model

with blocks in series were also implemented, and finally, our proposal, the DRD, was also evaluated. The full list of experiments can be seen in Table 8.3, which indicates the total filter parameters and the number of layers. A brief description of each experiment is presented in the next section.



**Figure 8.4:** This is the original UNet architecture. The three blocks of the input in light green represent the encoder section, which is responsible for convolutional operations and downsamples the input image. The three blocks in dark green on the right perform the Deconvolutional operation and they represent the decoder section. The modifications performed in this study were done in coder and decoder blocks.

### ResUNet-i

In this experiment, the residual block as described in Figure 8.3 was inserted at the input layer of the encoder, which is the first green box just after the input, thus the suffix  $-i$ . The total number of layers of this model is 62.

### ResUNet-e

The residual block was inserted at all layers of the encoder (then the suffix  $-e$ ), which are the light-green boxes shown on the left in Figure 8.4. The total layers of this model is 66.

### ResUNet

In this model, the residual block was added to the green blocks (dark and light) from Figure 8.4 at both the encoder and decoder, which corresponds to the ResUNet

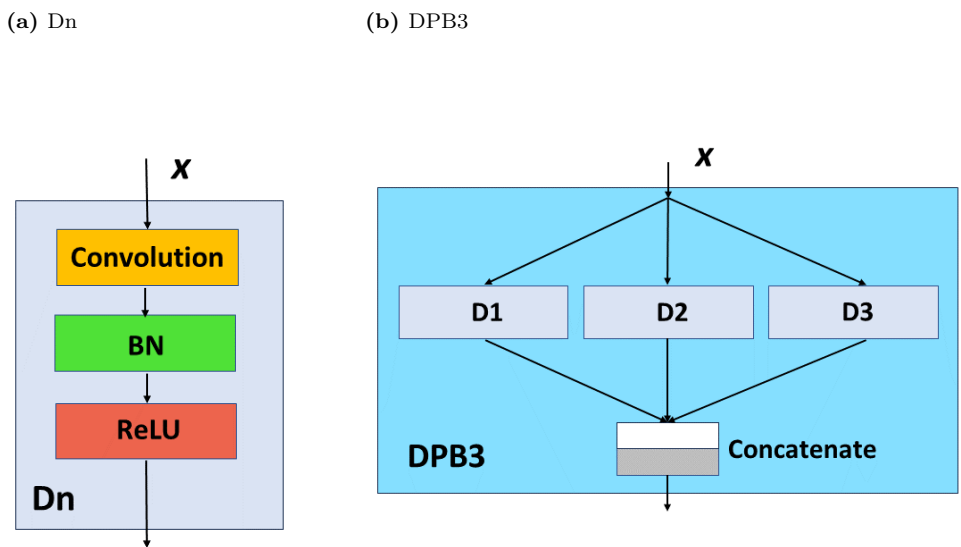
proposal given by (He et al., 2015), with 75 layers in total.

### ResUNet-BN

A modified ResUNet with added batch normalization after the convolutional layer was evaluated in the model ResUNet-BN. It can be noticed that the original UNet does not include the BN layer. The number of layers reached 89.

### DPB3-i, DPB3-e and DPB3-a

Dilated Convolution was evaluated by inserting the DPB block presented by (Niu et al., 2022) at the input layer of the encoder, thus the suffix  $-i$ . This model has 62 layers. Also, DPB was tested at all the encoder blocks, with 74 layers, represented in light-green in the Figure 8.4, and at the encoder and decoder blocks (DPB3-a) with a total amount of 107 layers. The DPB operation is shown in Figure 8.5, and it consists of the parallel combinations of convolutions of dilation 1,2, and 3.



**Figure 8.5:** We named the combination of operations Convolution + BN + ReLU as Dn block. The DPB3 block was presented by(Niu et al., 2022) in the Purified UNet. Inside DPB3, each Dn block is the dilated version of the block on the left image.

### DPB5-i

A dilation of  $D = 5$  was also tested at the input layer of the encoder (DPB5-i); in this experiment, special interest was the receptive view of the input image, for that

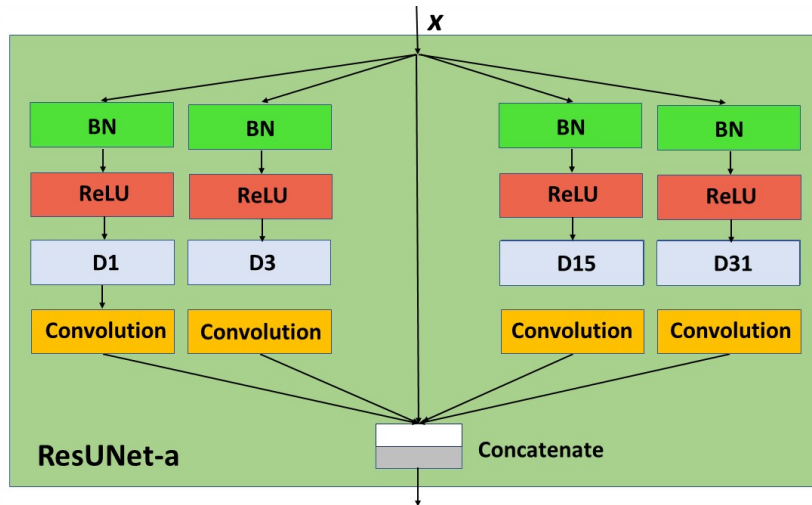
reason, a higher dilation factor was used in this experiment.

### DPB3-i+Res and DPB4-i+Res

Combinations of residual connections and Atrous convolutions were also evaluated in these two experiments. Dilations of size  $D = 3$  and  $D = 4$  at the input layer of the encoder with residual connection over all the layers of both the encoder and decoder were implemented. In these experiments, the total number of layers is 80 and 86, respectively.

### ResUNet-a

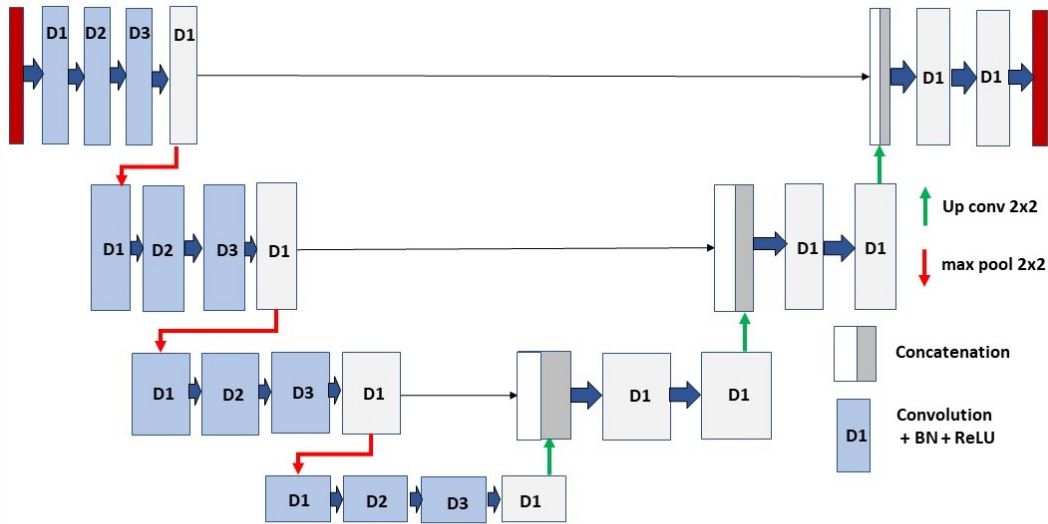
The network ResUNet-a, as presented by (Diakogiannis et al., 2020) for the segmentation of remote sense data, was used with dilations up to size  $D = 31$ , and was evaluated at the encoder. This model has the highest number of layers.



**Figure 8.6:** ResUNet-a, this is an arrangement of several parallel Convolutions with Dilations  $D = 1, 3, 15, 31$  as proposed by (Diakogiannis et al., 2020).

### Series-i and Series-e/d

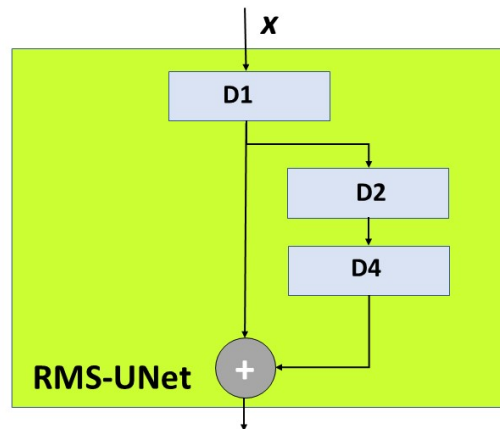
Dilation in series was also evaluated at the input (Series-i) block and the encoder and decoder blocks (Series-e/d). In this particular experiment, the UNet was modified to arrange the convolutions in series at every encoder, as it can be seen in Figure 8.7.



**Figure 8.7:** This is the Atrous Convolution in series inserted at every encoder, it performs the sequence Convolution with Dilation  $D = 1, D = 2$  and  $D = 3$  at the encoder and at the UNet backbone

## RMS-UNet

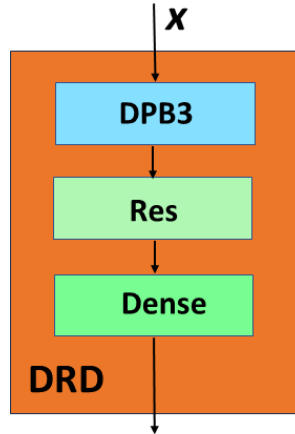
RMS-UNet presented by (Khan et al., 2022) combines residual connections and Atrous convolution with  $D = 2, 4$  inserted in the residual connection (Figure 8.8), as stated by (Khan et al., 2022), this configuration achieves high performance with minimal loss and low computational cost, and for that reason, it was also selected for our study.



**Figure 8.8:** RMS block as described by (Niu et al., 2022), it is a dilated version of the Residual, the blocks  $D2$  and  $D4$  indicate Convolution with Dilation of size  $D = 2$  and  $D = 4$

## DRD

The last experiment employs a model composed of a Dilation DPB3 block, Residual, and Dense block named (DRD) as seen in Figure 8.9, this is our contribution and its motivation was previously presented in this Chapter.



**Figure 8.9:** Our proposal, a block integrated with a DPB3 + Residual + Dense (DRD). The DPB3 block corresponds to figure 8.5b, the block Res to figure 8.3a, and the block Dense is represented in figure 8.3b.

Table 8.3 presents the total amount of trainable parameters from the corresponding Convolutional filters given in  $1 \times 10^6$  and the total amount of layers. The model with the largest amount of parameters is DPB3-a ( $21.4e6$ ), and DRD-UNet contains the higher number of layers (130).

### 8.4.3 Hyper-parameters selection

The hyper-parameters previously selected for the multi-class segmentation experiment presented in chapter 6 were used in this study with minor adjustments for the training and were maintained for all architectures. Mini-batch of size 8, 10 epochs with a random shuffle of images at every epoch, initial learning rate of  $1e-3$ , adaptive moment (ADAM) estimation optimization algorithm. The loss function employed was the Cross-Entropy for  $k$  Mutually Exclusive Classes (Prince, 2023). The total number of parameters to be trained for each model varies from  $7.69 \times 10^6$  in the baseline UNet up to  $21.41 \times 10^6$  for the DPB3-a, which is a model with modifications at

all levels of the encoder and decoder. This is computed by counting the total number of filter coefficients from all the 2-D convolution layers, given in their corresponding definition tables, for example, from tables 8.1 and 8.2 for UNet depth-3, at the layer 2, encoder-stage-1-Conv-1, there are  $64 \times 3 \times 3 \times 3 = 1728$ . On the other hand, the total number of iterations is 12,996.

#### 8.4.4 Implementation Details

All the programming was performed in a computer with processor Intel core i7-7700k, 16 GB RAM, and CPU at 4.20 GHz with an Intel graphics P4000 GPU. The platform used was Matlab<sup>®</sup> version 2023a (The Mathworks<sup>™</sup>, Natick, MA, USA) with the deep learning toolbox. All the models were trained on this computer, which took from a few hours to dozens of hours for training the largest models. The code is publicly available in the GitHub repository: <https://github.com/mauOrtRuiz/DRD-UNet>.

## 8.5 Results

The performance of the architectures proposed in this work was assessed by pixel classification, from a binary classification each pixel has one category: True Positive ( $TP$ ), True Negative ( $TN$ ), False Positive ( $FP$ ), and False Negative ( $FN$ ) when compared against the ground truth. For the multi-class segmentation all these have been calculated on a *per – class* basis, the sub-index  $i$  indicates the corresponding class, as described in Appendix D. From these, the following pixel-wise, and class-wise metrics were calculated: dice coefficient, Jaccard similarity index (also known as intersection over union), specificity, and sensitivity. According to (Müller et al., 2022), sensitivity and specificity are not proper segmentation metrics if they are not correctly interpreted. For that reason, our results are based only on Accuracy, dice, and Jaccard index. Specificity and sensitivity values are presented in Appendix C for future reference.

The sixteen experiments correspond to variants of the base model, and the performance for each modification is evaluated. Each new proposal over the model was built, and after a model compilation with no errors, it was trained and validated with the selected dataset, for that reason, each modification corresponds to a complete full ablation experiment over the base UNet. The trained models were used to per-

Model	Jaccard Similarity Index				
	Other	Tumor	Stroma	Inflammatory	Necrosis
UNet 3	0.33±0.29	0.73±0.10	0.60±0.17	0.61±0.21	0.15±0.23
ResUNet-i	0.29±0.27	0.70±0.12	0.51±0.20	0.48±0.22	0.14±0.22
ResUNet-e	0.28±0.23	0.70±0.13	0.61±0.16	<b>0.68±0.10</b>	0.19±0.28
ResUNet	0.31±0.27	0.73±0.12	0.59±0.17	0.62±0.18	0.16±0.25
ResUNet-BN	0.27±0.23	0.72±0.13	0.59±0.17	0.61±0.19	0.16±0.27
P-ResUNet	0.30±0.25	0.67±0.18	0.60±0.17	0.62±0.19	0.17±0.26
DPB3-i	0.31±0.25	0.71±0.14	0.60±0.16	0.62±0.19	0.19±0.28
DPB3-e	0.30±0.28	0.73±0.14	0.59±0.18	0.61±0.15	0.15±0.25
DPB3-a	0.31±0.29	0.73±0.15	<b>0.62±0.19</b>	0.62±0.20	<b>0.23±0.31</b>
DPB5-i	0.31±0.25	<b>0.75±0.10</b>	0.59±0.17	0.65±0.12	0.20±0.30
DPB3-i+Res	0.32±0.28	0.66±0.19	0.60±0.16	0.59±0.20	0.21±0.30
DPB4-i+Res	0.26±0.25	0.72±0.13	0.48±0.19	0.53±0.17	0.11±0.19
ResUNet-a	0.31±0.26	0.73±0.14	0.60±0.17	0.60±0.23	0.16±0.25
Series-i	0.28±0.27	0.73±0.13	0.48±0.19	0.57±0.20	0.10±0.17
Series-e/d	0.32±0.28	0.70±0.16	0.61±0.16	0.62±0.20	0.20±0.29
RMS-UNet	0.30±0.25	0.74±0.12	0.60±0.17	0.63±0.20	0.20±0.28
<b>DRD-UNet</b>	<b>0.34±0.28</b>	<b>0.75±0.11</b>	<b>0.62±0.17</b>	0.64±0.20	0.18±0.27

**Table 8.4: Performance of sixteen deep learning architectures and the proposed DRD-UNet.** Per-class performance is measured with Jaccard Similarity Index. Best results are highlighted in **bold**

Model	Dice Coefficient				
	Other	Tumor	Stroma	Inflammatory	Necrosis
UNet 3	0.43±0.33	0.84±0.08	0.73±0.16	0.73±0.22	0.20±0.31
ResUNet-i	0.39±0.32	0.82±0.09	0.65±0.20	0.62±0.23	0.20±0.29
ResUNet-e	0.40±0.29	0.82±0.09	0.74±0.16	<b>0.80±0.07</b>	0.24±0.35
ResUNet	0.41±0.32	0.84±0.09	0.72±0.16	0.75±0.16	0.22±0.33
ResUNet-BN	0.38±0.29	0.83±0.09	0.73±0.16	0.73±0.21	0.20±0.34
P-ResUNet	0.41±0.30	0.79±0.15	0.74±0.16	0.74±0.22	0.22±0.33
DPB3-i	0.42±0.30	0.82±0.11	0.74±0.16	0.74±0.22	0.25±0.35
DPB3-e	0.40±0.33	0.83±0.10	0.73±0.17	0.75±0.12	0.20±0.31
DPB3-a	0.41±0.32	0.84±0.12	0.74±0.19	0.74±0.22	<b>0.29±0.38</b>
DPB5-i	0.42±0.31	<b>0.85±0.07</b>	0.72±0.16	0.78±0.10	0.25±0.36
DPB3-i+Res	0.43±0.33	0.78±0.16	0.73±0.17	0.72±0.21	0.26±0.37
DPB4-i+Res	0.36±0.30	0.83±0.10	0.62±0.20	0.68±0.15	0.16±0.25
ResUNet-a	0.42±0.31	0.84±0.11	0.74±0.17	0.71±0.25	0.21±0.32
Series-i	0.37±0.32	0.84±0.09	0.63±0.20	0.70±0.19	0.15±0.24
Series-e/d	0.42±0.33	0.81±0.13	0.74±0.16	0.74±0.22	0.25±0.36
RMS-UNet	0.40±0.30	<b>0.85±0.08</b>	0.74±0.16	0.75±0.22	0.25±0.35
<b>DRD-UNet</b>	<b>0.44±0.33</b>	<b>0.85±0.08</b>	<b>0.75±0.16</b>	0.75±0.22	0.24±0.35

**Table 8.5: Performance of sixteen deep learning architectures and the proposed DRD-UNet.** Per-class performance is measured with the Dice coefficient. Best results are highlighted in **bold**

Model	Accuracy
UNet 3	0.80±0.05
ResUNet-i	0.73±0.01
ResUNet-e	0.78±0.06
ResUNet	0.79±0.06
ResUNet-BN	0.79±0.05
P-ResUNet	0.78±0.07
DPB3-i	0.79±0.06
DPB3-e	0.78±0.07
DPB3-a	<b>0.81±0.06</b>
DPB5-i	0.80±0.05
DPB3-i+Res	0.78±0.06
DPB4-i+Res	0.72±0.11
ResUNet-a	0.80±0.07
Series-i	0.72±0.11
Series-e/d	0.79±0.07
RMS-UNet	0.80±0.05
<b>DRD-UNet</b>	<b>0.81±0.05</b>

**Table 8.6: Performance of sixteen deep learning architectures and the proposed DRD-UNet.** Performance is measured with the Accuracy. Best results are highlighted in **bold**

form semantic segmentation of the fifteen ROIs, which had been selected as the test set. The performance of each architecture was measured as described previously, for Dice, Jaccard, and accuracy. Sensitivity and specificity as stated by (Müller et al., 2022), are not suitable for measuring segmentation. For that reason, we present tables for the Jaccard similarity index (Table 8.4), Dice coefficient (Table 8.5), and accuracy (Table 8.6). Results of the semantic segmentation of selected regions of the images are presented in Figures 8.11 and 8.12 for all the architectures.

From the tables, we can notice that adding residual connections (experiments 2-4) increases the number of parameters, but seems to reduce the overall performance, except for the class *Inflammatory*, where the performance increases. This behavior holds regardless of the section of the UNet where the residual connections are placed. Additionally, notice that Batch Normalization (ResUNet-BN) seems to have a random impact, sometimes increasing and sometimes lowering the performance, although just marginally.

The performance rises for classes *stroma* and *necrosis* when the dilated convolution operation is added to the architecture (DPB3, DPB5). This shows that residual connections with dilated convolutions are a suitable combination. Moreover, when the dilated convolution is used as part of the DPB module, it has a null contribution, except for class *necrosis*. Such results show that the dilated convolution by itself brings little contribution, as opposed to its combination with the residual connections.

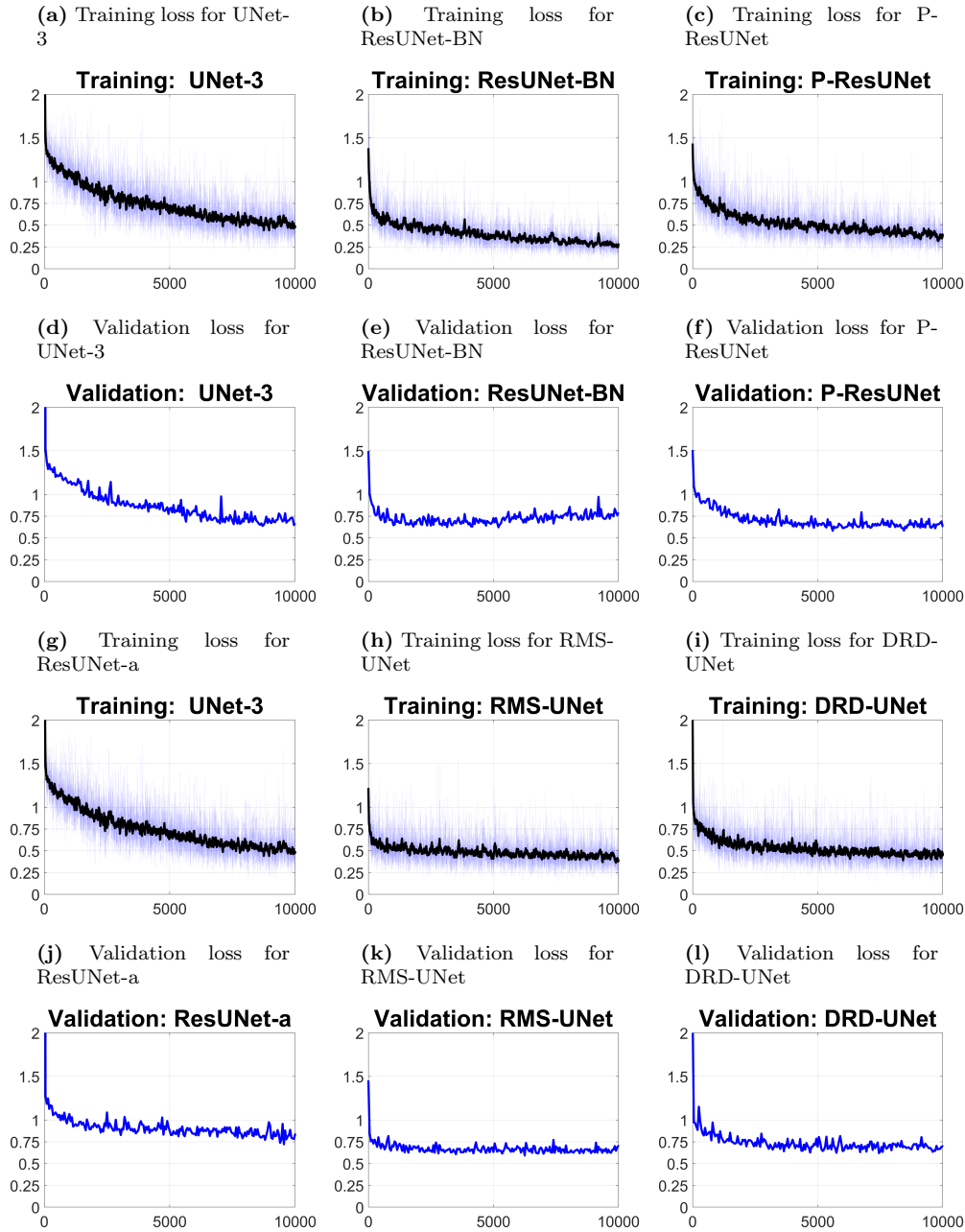
Regarding the architectures where the dilation of the convolutions is gradually increased (Series and RMS), it seems that these architectures also improve and lower the performance depending on particular classes but fail to remain consistently better for all cases. Only when the integration of all variants in a robust architecture (DRD-UNet) it achieves the highest performance for three classes and on average for all five classes.

Figures 8.11 and 8.12 illustrate the semantic segmentation results with selected sections of the images that contain a representative region of each of the five classes in each column and rows correspond to the architectures. TP are labeled in white; TN are labeled in black; FP are labeled in green, and FN are labeled in pink. The best results of DRD-UNet are visible in the smaller green and pink regions for tumor, stroma and other, which correspond to the values of Jaccard and Dice shown in tables 8.5 and 8.4. For inflammatory and necrosis, the areas are slightly higher

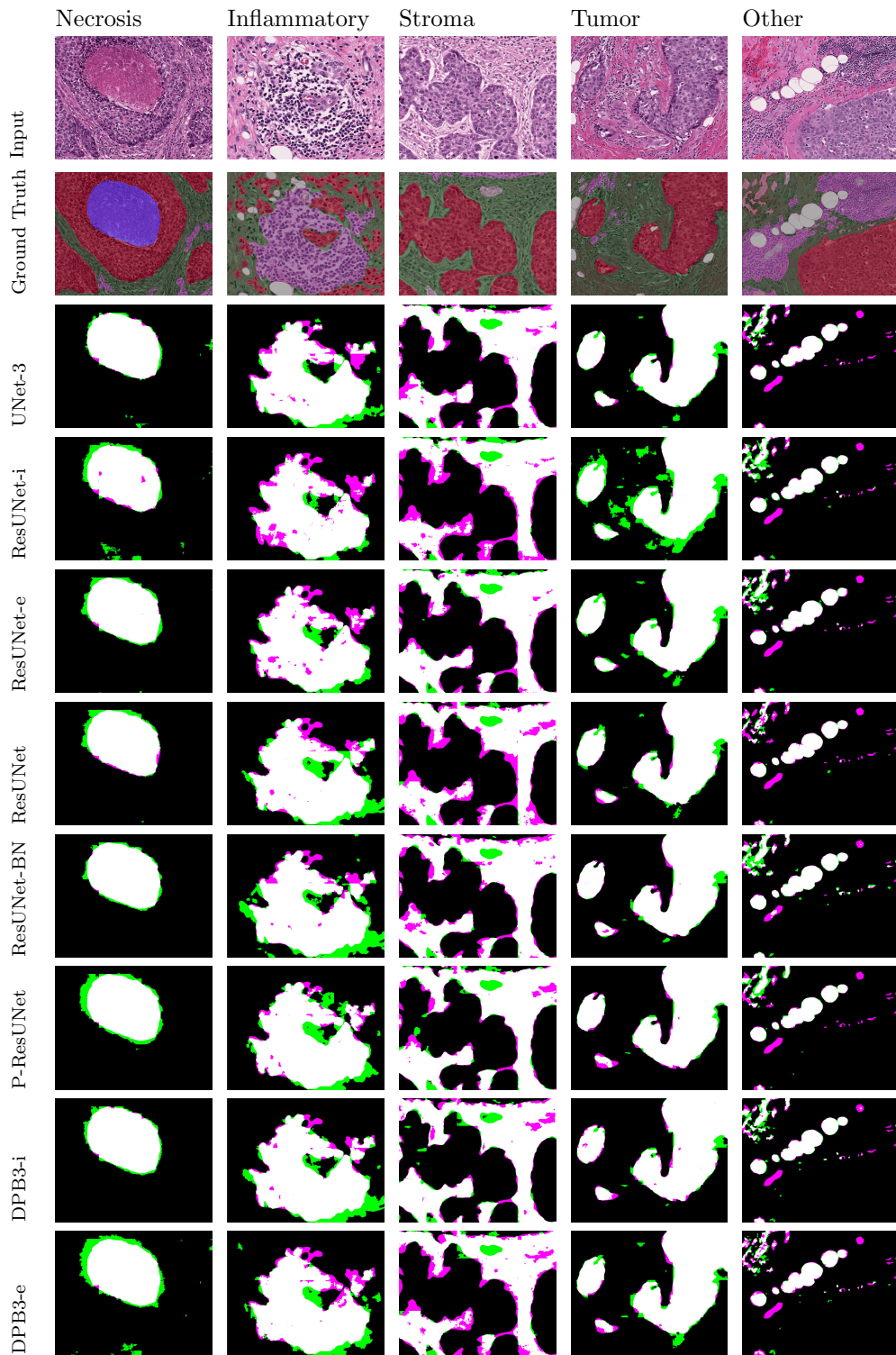
than other architectures, but since tumor and inflammatory are the most common classes, these are the ones that have a greater impact on the overall accuracy giving DRD-UNet an advantage.

## 8.6 Summary

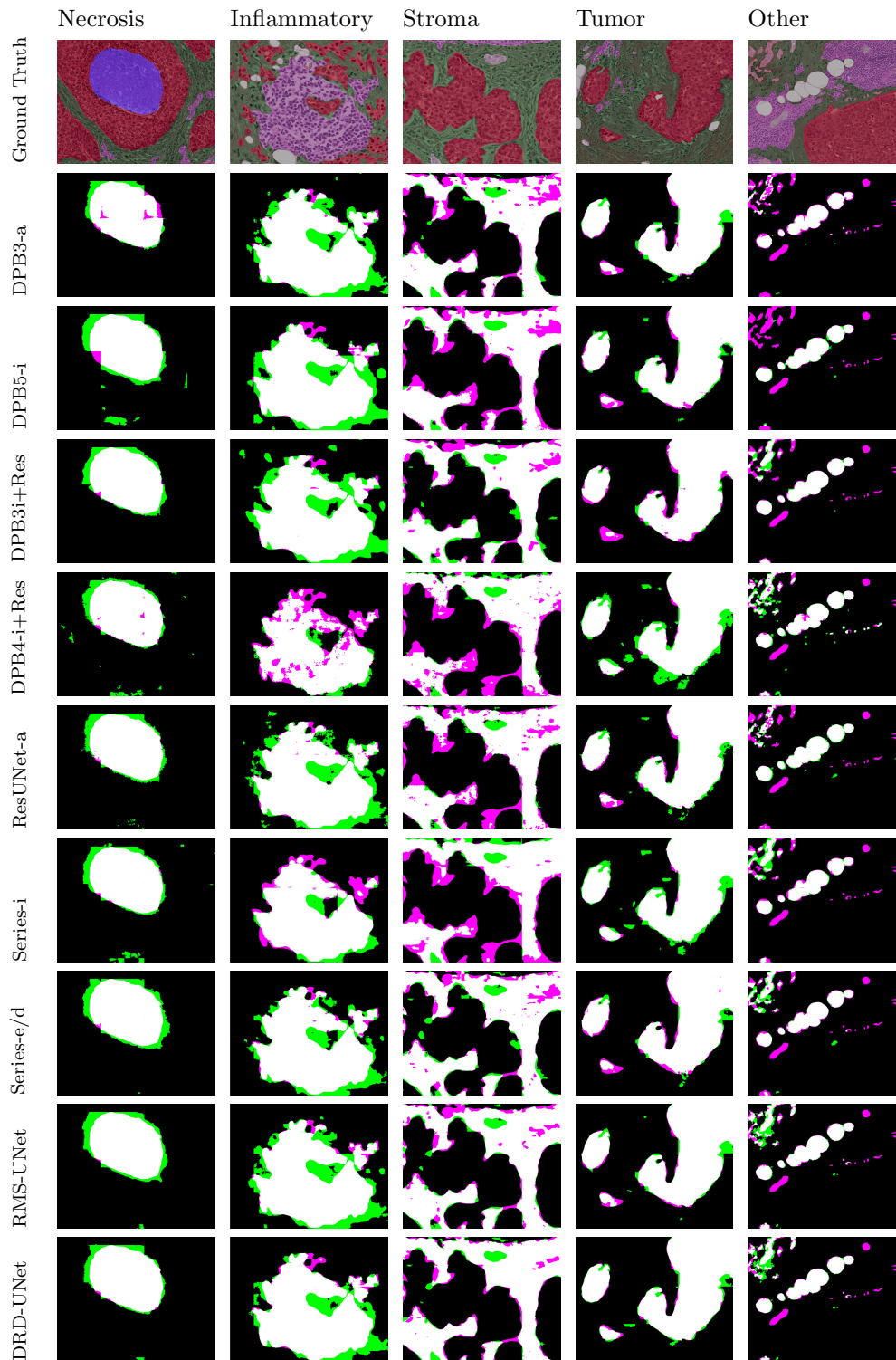
One of the main contributions of the thesis is presented in this chapter. A modification of the UNet architecture is proposed. It is based on inserting a dilation + residual + dense block. An ablation analysis of UNet was performed by replacing blocks and adding branches and layers mainly based on dilation and residual operations. Our proposal was achieved after the interpretation of the 16 experiments, and its performance improves more classes in our model than in the other results. The proposed DRD UNet is formed by the insertion of a full processing unit block named DRD which is located in each step at the encoder side. A more recent research, performed several months after our work was published, confirms the relevance of the combination of Dense + Residual + Dilation. The work by (Osman and Tamam, 2023) presents the named DD-Res-UNet, used for contrast-enhanced T1 image synthesis from contrast-free image(s) in 3D MRI images. This modification is different from our proposal, because it is in a different position inside the UNet architecture; the residual and dense are separate modules and with a different position in the original UNet. However, as demonstrated by (Osman and Tamam, 2023) it shows excellent capabilities in the generation of synthetic contrast-enhanced T1 images from contrast-free MRI images. This indicates the relevance of the combination of blocks. Also, the result presented by (Kim and Lee, 2024) in a network called N-Net, is a pre-trained DenseNet-121 is inserted before the actual encoder section of the UNet (thus it becomes an N shape rather than U) with direct connections at each encoder level. The dense section is at the actual input of this network, and the dilation and residual are in the encoder and backbone of the UNet. It is not a full processing unit, but it outperforms other segmentation methods used for Colonoscopy image segmentation. This result also confirms the potential of a Dense+Residual+Dilation combination. These findings motivate us to continue analysing this combination applied in different datasets, as well as for other tasks different than segmentation.



**Figure 8.10: Comparison of training loss and validation loss curves for the architectures that have been published** UNet-3 (Ronneberger et al., 2015), ResUNet-BN (Han, 2017), P-ResUNet (Niu et al., 2022), ResUNet-a (Diakogiannis et al., 2020), RMS-UNet (Khan et al., 2022), and DRD-UNet (ours). (a-f) correspond to training loss curves, (g-l) correspond to validation loss curves. For all cases, the training loss is shown in blue with transparency and a weighted average thicker black line is overlaid. DRD-UNet shows a rapid decrease in training loss, similar to RMS-UNet and there is no indication of overfitting in the validation curves. The case with slowest decrease is ResUNet-a, which ends above UNet-3. ResUNet-BN shows a fast decrease in the training, with probably the lowest values, but on the validation, there is an increase, which suggest some overfitting.



**Figure 8.11: Semantic segmentation results of five different classes with the first 8 models evaluated.** The first row shows representative input patches that focus on a region where each class is present. The second row shows the classes of the ground truth overlaid on the input patch with colors: necrosis (dark purple), inflammatory (light purple), stroma (green), tumor (red), and other (gray label). Rows 3 to 10 are the results of the first 8 models. TP are labeled in white; TN are labeled in black; FP are labeled in green, and FN are labeled in pink.



**Figure 8.12: Semantic segmentation results of five different classes with models 9 to 17.** The first row shows representative input patches that focus on a region where each class is present. The second row shows the classes of the ground truth overlaid on the input patch with colors: necrosis (dark purple), inflammatory (light purple), stroma (green), tumor (red), and other (gray label). Rows 2 to 10 are the results of the evaluated models. TP are labeled in white; TN are labeled in black; FP are labeled in green, and FN are labeled in pink.

## Chapter 9

# Conclusions and Future Work

In this thesis, methodologies for the assessment of tumour cellularity, tissue segmentation of breast cancer, and Gleason grading of prostate cancer from histopathological images were presented. The methodologies included both traditional image processing algorithms as well as deep learning architectures. Digital pathology for Breast cancer image analysis has been explored in this research through the assessment of tumour cellularity and segmentation of histopathology tissue. The segmentation approach was also applied to grade the Gleason score of prostate cancer images. The proposed methodology is based on the general workflow for digital pathology described in the literature. From the segmentation results, a deeper study of the UNet model was implemented, resulting in an improved segmentation model based on the dense, residual, and dilation.

This chapter presents a summary of the thesis, the major contributions and conclusions, and a proposed future work.

### 9.1 Summary

The overview of this research, followed by a general background and the relevant research previously done on breast cancer histopathology image analysis, are all described in Chapters 1 to 3. Next, the datasets used in our research are presented in Chapter 4.

In chapter 5, a digital pathology method for automated processing of H&E histopathology images was implemented for the clinical task of cellularity assessment. This method followed these steps: colour separation, cell segmentation, feature extraction, and tissue classification. Each of the steps of this workflow was

explored and implemented by selecting representative algorithms: K-means clustering, Otsu-binarization, watershed segmentation, and morphological analysis of binary images. Supported vector machine methods were selected, for their relevance seen also after comparison with other methods, for the classification of the tissue.

In chapter 6, three DL models for multi-class segmentation of H&E were compared. The performance of the segmentation over 4 different tissue regions was evaluated (tumour, stroma, inflammatory, and necrosis). In this task, a deep learning approach was selected, which contrasted with the hand-crafted methods, particularly in terms of software programming complexity, in which DL software code and programs are less complex but require a larger dataset for the training. The first step in the implementation was to select and adjust the hyper-parameters, for that reason, multiple trainings of the model were implemented and compared for different values to find the optimal parameters. The segmentation results were compared using accuracy, Dice, and mean AUC. Three neural networks were explored: ResNet, Mobilenet V2, and UNet. ResNet 18 and UNet demonstrated better performance, and UNet was selected for further experiments.

In chapter 7, the UNet model was trained and applied to the Gleason Grading of prostate cancer images. Dataset images from the automated Gleason grading challenge 2022 were used, which are from multiple slide scanners to consider colour variations. Images are of giga-byte size and hard to process; the images and the training required many computer resources and are quite time-consuming from the data preparation until the testing phase. After both experiments, UNet showed better performance for relevant classes (like tumour), and was selected for an ablation analysis for an improved model.

In chapter 8, an analysis was performed over the proposed UNet network to obtain an improved model for multi-class segmentation. After a literature review, two main operations were introduced into the UNet architecture at the encoder and decoder: the residual and dilation. 16 different experiments were performed during this study, each experiment was a training and validation of the corresponding model modification, and next, the performance was measured through accuracy, dice, and Jaccard. The comparison allowed for assessing different assumptions, explained in the results section, that yield a model proposal, formed by the combination of dilation, residual, and a dense operation. The performance of the proposed model outperforms several of the models studied.

## 9.2 Main contributions

1. A methodology for tumour cellularity assessment on BC images was proposed, and the results were published in a scientific paper. Within this, an efficient cell segmentation method was proposed for the analysis of H&E histopathology images based on nuclei enhancement. This is based on a combination of the gamma function and median filters, which results in a nuclei region with higher contrast in the grayscale histopathology image. This algorithm required the colour mask components (blue and pink) obtained by the colour separation done in the workflow. Also, a detection method for the tubules and ducts from breast images was proposed. The difference between the colour masks reveals the duct elements of the image.
2. Participation in the AGGC 2022. The proposed approach for the GG assessment was based on a segmentation model, and the test set for the contest was successfully processed. The results were ranked among the finalists of the contest. More than 750 participants, with approximately 1050 total submissions, were in the challenge. Our participation consisted of two successful submissions to the contest. The results achieved the 10th place among 20 finalists reported in the final ranking.
3. An improved segmentation UNet model was proposed. An analysis of the original UNet model revealed an efficient block in terms of accuracy based on three operations: dilation, residual, and dense. A scientific publication was published with this proposal.

## 9.3 Main conclusions

Conclusions on the assessment of tumour cellularity. The methodology processed a training set of breast cancer images under neo-adjuvant treatment, and the results indicated that 22 key morphological parameters are strongly correlated with cellularity. Interesting results were revealed from the correlation analysis of the morphological parameters, in which the strongest related parameter was stroma density, in agreement with (Beck et al., 2011), which states that the histology of stroma correlates with prognosis in breast cancer.

The conclusions for the proposed DRD block, for the UNet modification, provided improved results in 15 selected test images of the BCSS challenge. Whilst the set is relatively small, the results are encouraging and are worth considering in future experimentation. It can be noted that the importance of dilated convolutions to increase the receptive field, and we chose the purified DPB3 reported by (Niu et al., 2022) with an Atrous convolution for the ablation study. Experimentation revealed there was no significant improvement after adding several residual blocks, inside the different UNet levels. Under the assumption that we can extend performance by adding a deeper layer, a dense block was proposed after the residual, which resulted in the proposed model. Besides accuracy performance, the training loss of the proposed model indicates a rapid computational convergence in terms of training time.

Our analysis was performed on a single dataset of breast cancer cases. Whilst this is a limitation in the diversity of cancer cases, we consider that to compare a series of architectures, the conclusions in terms of performance should still be valid.

## 9.4 Suggestions for future work

From the participation in the AGGC, an important lesson was obtained from the annotated dataset. A well-annotated ground truth set of images is needed for the training, which is sometimes difficult to generate. The annotations from the GG experiment are not well delimited, which are sometimes called *weak* annotations (Dietterich et al., 1997). The results revealed that a *stronger* annotation improved the performance. This was observed after an application of the extraction criteria on the pathologist’s annotations, which means the ground truth becomes stronger. This approach demanded high computer resources, which is a machine learning method based on individual labelled instances. A better approach should be explored in the future, like Multiple Instance Learning (MIL), as presented by (Dietterich et al., 1997). This method overcomes the weak label problem, which is solved by a model that directly learns from weak label annotations. This is done by a dataset organized into bags of instances, the bag is labeled as positive if at least one instance is positive, otherwise it is negative. For the pathology image analysis, a bag can be a whole slide image, but the training requires a large number of images, then MIL overcomes the problem of *weak* annotations.

From the proposed model modification over the UNet model, further work can be done to evaluate the operation over different datasets and other types of cancers and staining. A preliminary test showed favorable results over the ACROBAT dataset, in which the model was tested, however, this was only after a visual inspection, as this set has no labeled reference for a quantitative analysis. Further studies to confirm the performance of DRD-UNet with larger datasets and with other cancers would be implemented to demonstrate the capabilities of the architecture. A closely related observation is that all the images were stained with H&E. Further experimentation with other staining should be done to explore the capabilities of the architectures compared here. This is an important and challenging issue in a deep learning model, the generalization or the ability to perform well on unseen data. This must also be explored in future research.

A recent framework approach for segmentation is called nnUnet (no new UNet), by (Isensee et al., 2021). This framework has become popular recently, as it has demonstrated its capabilities by being used by the winners of different medical segmentation challenges, for example, BraTS, LiTS, and KiTS. It can be adapted to any dataset and reproduced. Version 2 was released in 2023.

A recent framework approach for segmentation is called nnUnet (no new UNet), by (Isensee et al., 2021) and should be explored. This framework has become popular recently, as it has demonstrated its capabilities by being used by the winners of different medical segmentation challenges: BraTS, LiTS, and KiTS; and it can be adapted and reproduced to any dataset. Version 2 was released in 2023.



# Appendix A

## Digital images in medicine

Digital images have a wide range of applications, and this Appendix presents relevant examples in the field of medical imaging. Almost all medical fields require different types of images captured by specialized instruments, and in some cases, the specialist performs a visual inspection of organs of the body itself. A dermatologist for example observes anomalies on the skin, an ophthalmologist analyses the eye condition with specialized instruments to diagnose glaucoma or cataracts. Even a cardiologist uses an ultrasound or electrocardiography image that gives support to evaluate the structure and function of the heart. The main characteristics of medical images are described in this appendix.

### A.1 Digital images

An image is a two-dimensional function, say  $I(m, n)$ , of the pair of coordinates  $m$  and  $n$ , that represent the row and column of the respective pixel, it can be defined as

$$I(m, n) \rightarrow [0, 1]^c \tag{A.1}$$

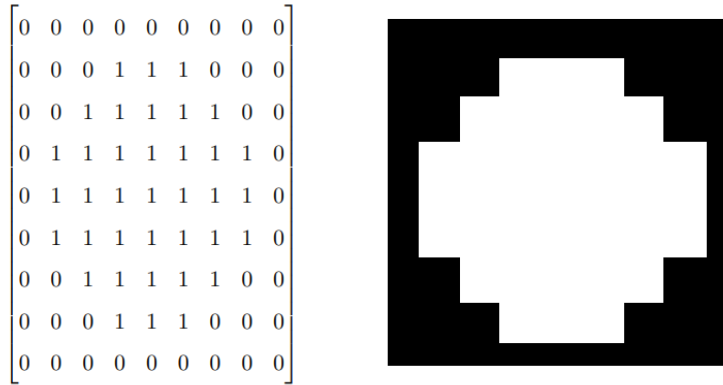
$I(m, n)$  represents the image mapped to a  $c$ -dimensional color space. For grayscale and binary images,  $c = 1$ , while for color images,  $c = 3$ . The coordinates  $(m, n)$ , correspond to the pixel location, where  $m = \{0, 1, 2, \dots, M - 1\}$  denotes the row index and  $n = \{0, 1, 2, \dots, N - 1\}$  denotes the column index.

### A.1.1 Binary images

A binary image can be defined as,

$$I_{BW}(m, n) \rightarrow \{0, 1\} \quad (\text{A.2})$$

where the output of the mapping belongs to a binary number either 1 or 0. Binary images are used for representation of objects in images, and in the segmentation context, the binary image is used as a mask representation of the segmented region. The object of interest is assigned a value of 1, while the background is assigned a value of 0. The image in Figure A.1 was extracted from a mask of a cell image in a pathology image, it has a circle shape of a diameter size of 7 pixels.



**Figure A.1:** A binary image is represented of two values 1 for an object of interest and 0 for the background, as seen in the matrix representation on the left. Values 1 are shown as white pixels and 0 as black pixels on the right.

### A.1.2 Grayscale images

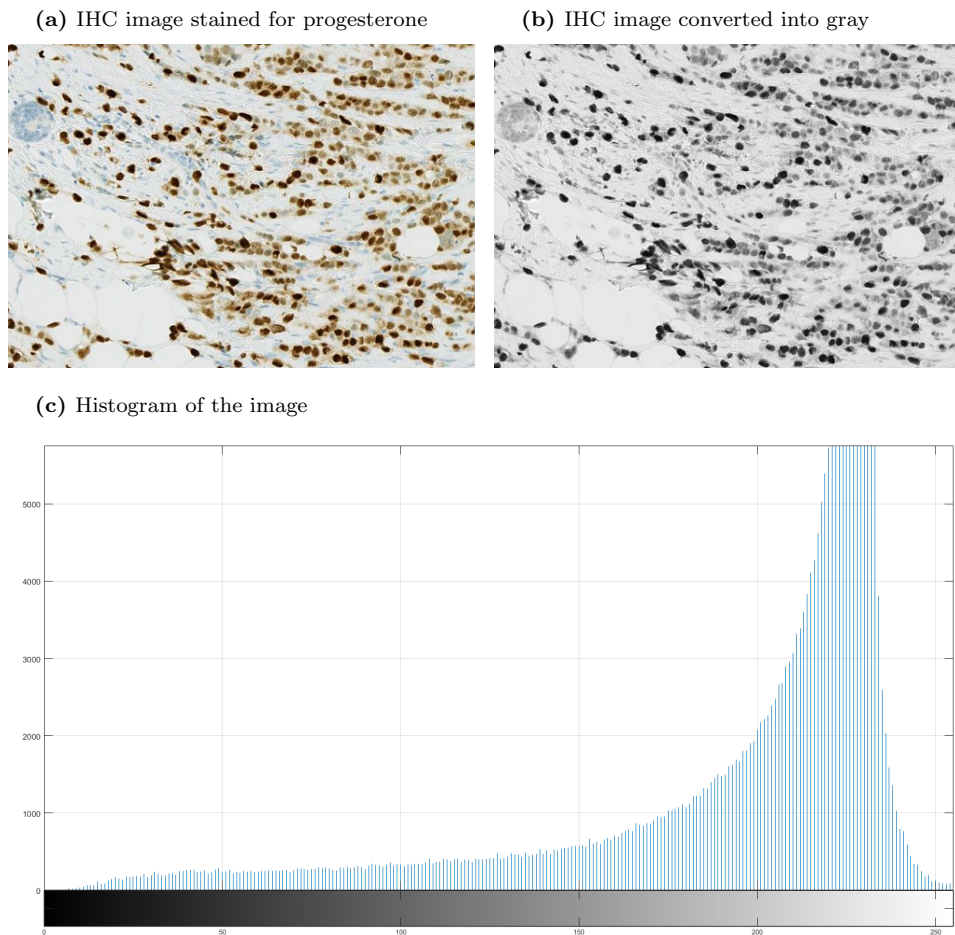
Gray images are represented by

$$I_g : U \rightarrow [0, 1] \quad (\text{A.3})$$

the function is mapped to a one-dimensional,  $c = 1$ , output and it is in the range  $[0, 1]$ , in which 0 represents black and 1 is white. For example, let  $a_{m,n}$  be the entries of a grayscale image, the image matrix can be expressed as (Gonzalez, 2009)

$$I = \begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0,N} \\ a_{1,0} & a_{1,1} & & a_{1,N} \\ \cdot & & & \\ \cdot & & & \\ a_{M,0} & a_{M,1} & \dots & a_{M,N} \end{bmatrix} \quad (\text{A.4})$$

In figure A.2 an example with a grayscale IHC image is presented. It is an IHC image stained for the progesterone (PR) protein receptor, this particular test, cell intensity is critical because it reflects the amount of PR protein present in the cells and an intensity score is assigned by the pathologist (weak, moderate, or strong staining); the gray level can provide information about intensity of a selected cell.



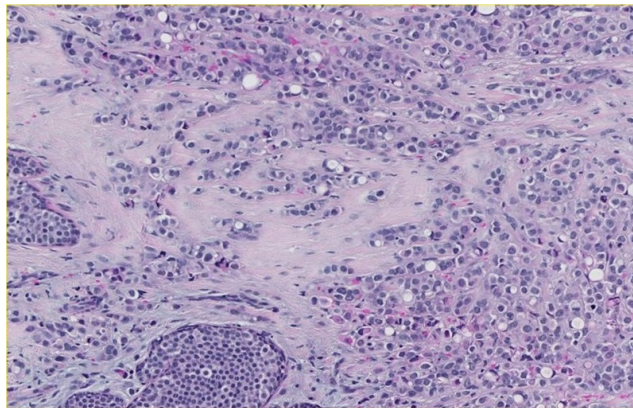
**Figure A.2:** The image in (a) is an IHC stained sample for progesterone (PGR) protein receptor, it is converted to grayscale image, and presented in (b). The histogram of the image is shown in (c). Image extracted from the ACROBAT dataset (Weitz et al., 2022).

### A.1.3 Colour images

In a colour image, the output of the mapping is a three-dimensional vector that represents the normalized colour in the range  $[0, 1]$ :

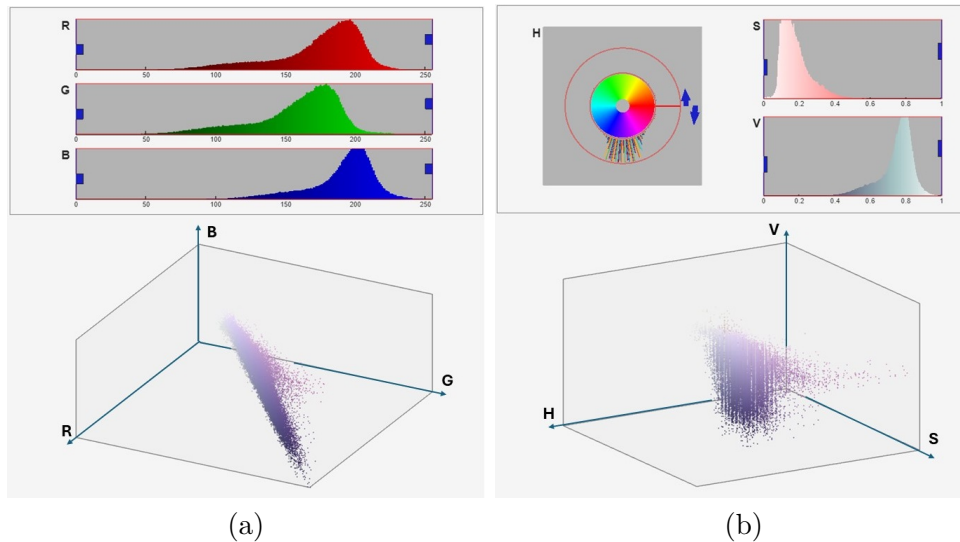
$$I(m, n) = \left( \frac{R}{255}, \frac{G}{255}, \frac{B}{255} \right) \quad (\text{A.5})$$

Each entry of the colour matrix is a number that indicates the image colour based on a 3D space representation. Two common representations are RGB and HSV. In an RGB space, each colour is formed by the union of the three main colours: red, green, and blue; In an HSV (Hue, Saturation, and Value) colour representation, each pixel is described by the three components: the Hue that is the colour itself and covers  $360^\circ$ , Saturation is the intensity of the colour given by a percentage (100% means a pure colour, and 0% is a black or no colour) and the brightness of the colour is given by the value. Also, there is the L\*a\*b colour space in which L\* is the brightness, a\* is the red to green chromatic component and b\* is for the blue to yellow one. This representation is useful for colour correction methods as is not dependent on the device like a camera or printer. There are two types of histopathology images based on the staining (H&E and IHC), these are presented in RGB and HSV colour space representations in Figures A.3 to A.6. In the RGB space representation the three colour histograms indicate each individual colour distribution. It is interesting to observe the HSV histogram, which is presented in a circular graph, for the Hue  $360^\circ$  span, and the colour peaks are presented in radial lines at the respective colour values.



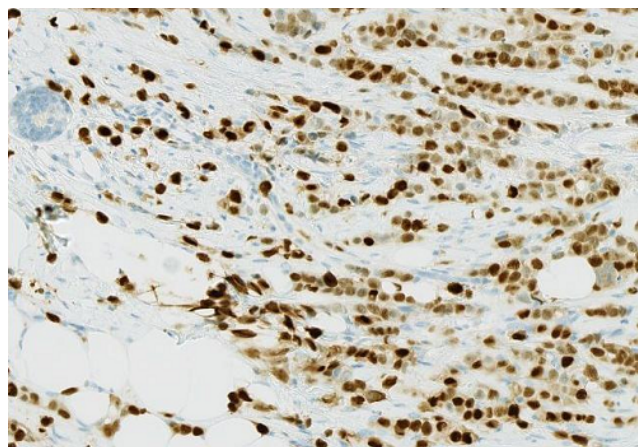
**Figure A.3:** H&E image formed of blue and pink components, HSV and RGB colour spaces are shown in Figure A.4.

These two colour representations are greatly employed in medical imaging and

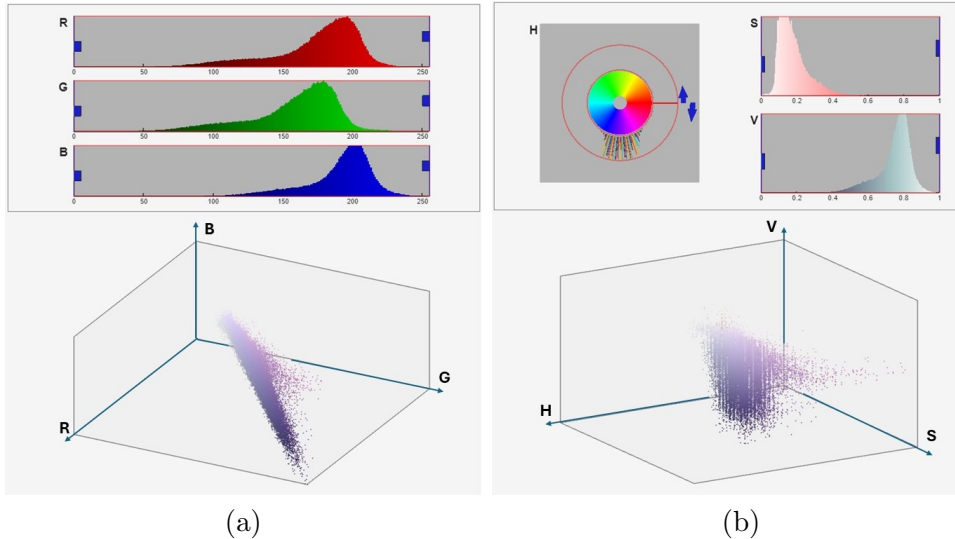


**Figure A.4:** Colour components of the H&E image from Figure A.3. The RGB colour space is in (a) and HSV colour space in (b). The upper plot are the corresponding histograms in each colour space. A colour histogram represents the colour saturation from the darker (on the left) to the lighter (on the right). In the case of HSV colour space Hue histogram is shown in circular mode. The 3D representation in the bottom of each map is useful to visualize the color distribution in the color space and is an indicator for their separability. Image from ACROBAT dataset (Weitz et al., 2022).

were used in this research work. Understanding the colour characteristics of pathological images is part of the computer pathology workflow. Colour variation in pathology images can be produced during tissue preparation, staining, and also due to the different scanners. Colour normalization in digital pathology is the process to ensure a reliable and consistent image analysis. Traditional methods of colour normalization are based on matching the histogram to a target reference slide (Reinhard et al., 2001b).



**Figure A.5:** H&E image formed of blue and brown colour components, HSV and RGB colour spaces are shown in Figure A.6.



**Figure A.6:** Colour components of an IHC image from Figure A.5. The RGB colour representation is in (a) and HSV colour representation in (b). Image from ACROBAT dataset (Weitz et al., 2022).

## A.2 Images in the frequency spectrum

This section presents examples of medical images categorized by their electromagnetic frequency, describing the types of images generated based on their primary energy source. Medical imaging modalities span a broad range of frequencies, starting from lower frequencies (longer wavelengths) such as infrared and visible light, to higher frequencies including ultraviolet, X-rays, and gamma rays.

### Gamma ray images

Gamma-ray images are produced by injecting the patient with a radioactive isotope that emits gamma rays as it decays. These emissions are detected by gamma-ray detectors to generate images used for visualizing organ function and tissue activity.

### X-Ray images

X-ray applications are widely used in medical diagnosis, particularly for detecting bone lesions. X-rays are generated by a tube containing a cathode and an anode, with the patient positioned between the X-ray source and an X-ray-sensitive film. The intensity recorded on the film changes based on the absorption of X-rays as they pass through the patient. Digital radiography produces X-ray images either by digitizing traditional X-ray films or by using a phosphor screen. While bones are easily visualized with X-rays, other organs can also be examined through radiographic

techniques.

Angiography involves injecting an X-ray contrast medium into the area or organ under study to visualize blood vessels. Additionally, 3D images produced by Computed Tomography (CT or CAT) are generated using X-ray signals.

### **Computer Tomography**

In CT imaging, cross-sectional images containing more detailed information than regular X-rays are produced. The CT scanner rotates around the body, acquiring X-ray measurements from multiple angles. These measurements are then processed by a computer system to generate 3D visualizations composed of “slices” of the body. Three standard views can be displayed: axial, coronal, and sagittal.

### **Mammography**

Mammography is a widely used X-ray application important for breast cancer diagnosis. This technique employs a low dose of X-rays to generate images of breast tissue called mammograms. These images reveal the internal structure of the tissue, allowing radiologists to identify masses and distortions associated with breast cancer. Mammography is typically used to detect cancer at an early stage, often before it can be felt by physical examination, which significantly improves survival rates. The American Cancer Society (ACS) recommends annual mammography screening starting at age 45. Although mammography is the preferred diagnostic tool, it can be less effective in women with dense breast tissue, where ultrasound is often used as a complementary method.

### **Ultraviolet Light Images**

Fluorescence microscopy is based on ultraviolet (UV) light. The phenomenon of fluorescence was first observed in the mid-20th century when the mineral fluor spar was found to fluoresce under UV light. Although ultraviolet light itself is invisible, it excites electrons in fluorescent materials. When these excited electrons return to their normal state, they emit visible light, often in the red region of the spectrum. A fluorescence microscope uses this excitation light to observe naturally fluorescent materials as well as specimens treated with specific chemicals that cause them to fluoresce. This type of microscope is widely used to visualize cellular structures and to monitor certain live biological processes.

### **Radio wave images**

At the opposite end of the frequency spectrum from gamma rays are radio waves, which are used in medicine to produce Magnetic Resonance Images (MRI). During an MRI scan, the patient is placed inside a powerful magnet to align hydrogen atoms within the body. Radio wave pulses are then transmitted through the patient's body, temporarily disturbing this alignment. The response to these pulses generates detailed two-dimensional images of cross-sections of the body. This technique can produce images in any plane. Sometimes, a contrast agent is injected to enhance visualization of specific regions or tumors. MRI is especially useful for detecting breast cancer when mammography results are inconclusive, such as in young women with dense breast tissue.

### **Visible light images**

Visible light encompasses the range of the electromagnetic spectrum between violet and red wavelengths, approximately 400 to 700 nm, which can be directly detected by the human eye. Images of the skin, retina, and those obtained from optical microscopes fall within this category. For example, pathology images of tissue samples analyzed under a microscope are included here. Our research primarily focuses on this type of image.

## Appendix B

# Wide Slide Scanners

Examples of WSI scanners are:

- a Akoya PhenoImager HT is an automated imaging system for brightfield or fluorescent whole slide imaging (WSI), it scans tissue sections with an 80-slide capacity (<https://www.akoyabio.com/phenoimager/instruments/>).
- b Optrascan OS-Ultra is a high-speed 40× whole slide digital pathology scanner that has one of the largest 3 × 2 in slide scanning capacities with continuous slide feeding capacity (<https://www.optrascan.com/scan/os-ultra-brightfield-scanner>).
- c 3D Histech Panoramic 250 Flash is a scanner that allows continuous loading to accommodate large volumes of slides. The model is capable of producing whole slide images at 20× and 40× magnifications (<https://www.3dhistech.com/diagnostics/pannoramc-diagnostic-scanners/>).
- d Leica Aperio GT450DX is designed for high-throughput pathology laboratories. Its high capacity and continuous loading capability scans a large number of slides quickly and automatically with minimal input from the operator. The scanner contains an integrated image algorithm providing real-time notification of suboptimal scanning via a touchscreen interface (<https://www.leicabiosystems.com/digital-pathology/scan/aperio-gt-450-dx/>).
- e Hamamatsu NanoZoomer S360 A high-throughput scanner capable of scanning 360 slides in one batch (<https://www.hamamatsu.com/jp/en/product/life-science-and-medical-systems/digital-slide-scanner/C13220-01.html>).



## Appendix C

# Sensitivity and specificity performance of DRD-UNet model

Model	Sensitivity Index				
	Other	Tumor	Stroma	Inflammatory	Necrosis
UNet 3	0.42±0.31	<b>0.94±0.06</b>	0.62±0.20	0.51±0.25	0.55±0.14
ResUNet-i	0.44±0.31	0.88±0.12	0.73±0.14	0.70±0.18	0.71±0.12
ResUNet-e	0.44±0.31	0.88±0.12	0.73±0.14	0.70±0.18	0.71±0.12
ResUNet	0.45±0.32	0.88±0.12	0.73±0.14	0.73±0.22	0.56±0.17
ResUNet-BN	<b>0.56±0.33</b>	0.82±0.14	0.71±0.14	0.87±0.08	0.54±0.32
P-ResUNet	0.48±0.32	0.76±0.21	0.77±0.13	0.83±0.09	0.69±0.15
DPB3-i	0.42±0.30	0.80±0.17	<b>0.78±0.12</b>	0.81±0.11	0.65±0.12
DPB3-e	0.42±0.31	0.86±0.14	0.74±0.15	0.70±0.18	0.74±0.12
DPB3-a	0.36±0.30	0.84±0.17	0.75±0.16	0.86±0.10	0.71±0.15
DPB5-i	0.44±0.33	0.87±0.10	0.71±0.15	0.80±0.17	0.78±0.19
DPB3-i+Res	0.43±0.32	0.72±0.21	0.76±0.13	<b>0.91±0.06</b>	0.69±0.10
DPB4-i+Res	0.39±0.32	0.90±0.12	0.60±0.21	0.55±0.18	0.71±0.13
ResUNet-a	0.44±0.33	0.84±0.15	0.76±0.13	0.81±0.21	0.59±0.18
Series-i	0.32±0.30	0.91±0.11	0.59±0.19	0.61±0.23	0.77±0.09
Series-e/d	0.45±0.31	0.77±0.19	<b>0.78±0.12</b>	0.85±0.11	0.71±0.17
RMS-UNet	0.48±0.32	0.84±0.12	0.72±0.13	0.87±0.08	<b>0.79±0.15</b>
<b>DRD-UNet</b>	0.43±0.32	0.87±0.11	0.76±0.13	0.83±0.11	0.71±0.14

**Table C.1: Performance of sixteen deep learning architectures and the proposed DRD-UNet.** Per-class performance is measured with a Sensitivity Index. Best results are highlighted in **bold**

Model	Specificity Index				
	Other	Tumor	Stroma	Inflammatory	Necrosis
UNet 3	0.98±0.02	0.72±0.18	0.88±0.11	<b>0.99±0.00</b>	0.97±0.04
ResUNet-i	0.98±0.01	0.83±0.13	0.86±0.10	0.98±0.01	<b>0.98±0.02</b>
ResUNet-e	0.98±0.01	0.83±0.13	0.86±0.10	0.98±0.01	0.98±0.02
ResUNet	0.98±0.01	0.83±0.14	0.87±0.08	0.97±0.01	<b>0.99±0.00</b>
ResUNet-BN	0.96±0.03	0.89±0.10	<b>0.88±0.06</b>	0.95±0.04	<b>0.99±0.00</b>
P-ResUNet	0.97±0.01	0.91±0.10	0.83±0.10	0.96±0.03	0.97±0.02
DPB3-i	0.98±0.00	0.90±0.09	0.83±0.09	0.96±0.02	0.98±0.01
DPB3-e	0.98±0.01	0.86±0.11	0.86±0.09	0.98±0.01	0.96±0.04
DPB3-a	<b>0.99±0.00</b>	0.87±0.17	0.87±0.07	0.95±0.04	0.98±0.01
DPB5-i	0.98±0.01	0.85±0.14	<b>0.88±0.06</b>	0.96±0.03	0.97±0.02
DPB3-i+Res	0.98±0.01	<b>0.93±0.07</b>	0.82±0.10	0.93±0.06	0.98±0.01
DPB4-i+Res	0.98±0.01	0.80±0.15	0.86±0.13	<b>0.99±0.00</b>	0.93±0.06
ResUNet-a	0.42±0.01	0.88±0.12	0.85±0.10	0.96±0.03	0.98±0.01
Series-i	0.99±0.00	0.81±0.13	0.88±0.11	<b>0.99±0.01</b>	0.91±0.06
Series-e/d	0.98±0.01	0.92±0.07	0.83±0.10	0.96±0.03	0.97±0.02
RMS-UNet	0.97±0.02	0.90±0.08	<b>0.88±0.06</b>	0.96±0.03	0.97±0.01
DRD-UNet	0.98±0.00	0.88±0.10	<b>0.88±0.06</b>	0.96±0.03	0.98±0.01

**Table C.2: Performance of sixteen deep learning architectures and the proposed DRD-UNet.** Per-class performance is measured with the Specificity. Best results are highlighted in **bold**



# Appendix D

## Deep Learning Background

### D.1 Neural Network Structure

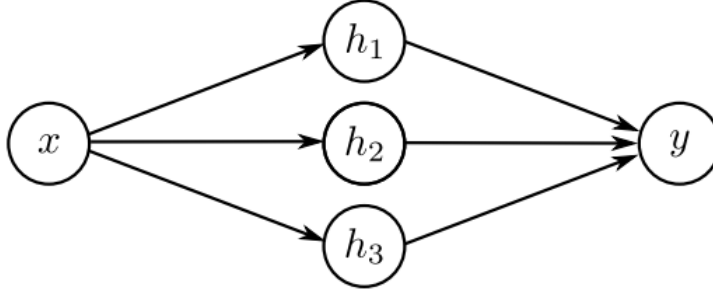
An ANN is a computational model that processes information, based on the way a biological neuron from the human brain recognizes patterns (Lin, 2017). The ANN, as seen in Figure 2.13, has an input  $\mathbf{x} = (x_1, \dots, x_M)$ , and an output  $y$  which is a linear combination of inputs with a certain set of weights  $\mathbf{w} = (w_1, w_2, \dots, w_M)$ , and a bias  $b = w_0$ , which are scalars that enhance certain features from the input signal. The output of the neuron is given by a non-linear activation function  $h(x)$ . Mathematically the neuron can be expressed by

$$y = h(\mathbf{w}^T \mathbf{x} + w_0) \tag{D.1}$$

A linear combination of multiple basic neurons, also called hidden units (Prince, 2023), increases the network capabilities. Then, an intermediate layer is formed by  $N$  different neurons, also known as a hidden layer (a layer between input and output). The number of hidden units  $N$  determines the network's capacity, for example, any continuous function  $f$  of dimension 1-D can be approximated by the combination of several neurons (Cybenko, 1989), let  $\hat{f}$  be the approximation, then it can be expressed by:

$$\hat{f} = \sum_{i=0}^{N-1} \phi_i h(\mathbf{w}^T \mathbf{x} + w_0) \tag{D.2}$$

The activation functions are no-linear functions and the Rectified Linear output



**Figure D.1:** Multiple neurons in the hidden layer example. Each neuron is also called a hidden unit as it is inserted in the hidden layer, multiple hidden units are combined with a parameter  $\phi$  that increases the neuron capability, this is an example for  $N = 3$  hidden units. Source (Prince, 2023)

(ReLU) is the most common which is defined as:

$$h_{ReLU} [z] = \begin{cases} 0 & \text{for } z < 0 \\ z & \text{for } z \geq 0 \end{cases} \quad (\text{D.3})$$

Another activation function is the Sigmoid, defined as

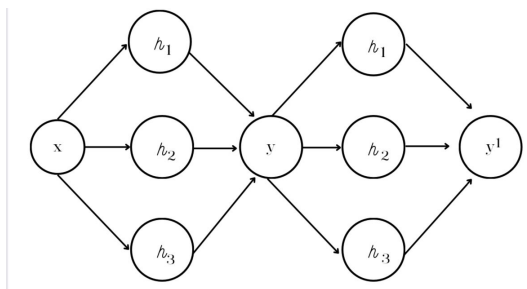
$$h_{Sigmoid} [z] = \frac{1}{1 + e^{-z}} \quad (\text{D.4})$$

and also hyperbolic Tangent.

$$h_{\text{hyperbolic tangent}} [z] = \tanh(z) \quad (\text{D.5})$$

In terms of layers, we can define the first layer as the input layer, the center layers are the hidden layers and the final layers at the right side are the output layers. This simple case is a shallow network. If we add more hidden layers, in which the output of the first layer becomes the input of the next layer we get a more complex neural network named deep neural network, which can improve the descriptive functionality up to high dimension (images) complex input functions. Figure D.1 shows the case of a network with two hidden layers. In this case, the overall effect is that an extra hidden layer with another set of units increases the linear regions (Cybenko, 1989) of a given analysed function.

Current networks can have more than 100 layers and thousands of hidden units at each layer. The number of hidden layers is the depth of the network and the number of hidden units is the width of the network, which determines the network's capacity. All these elements described so far define the neural network operation it-



**Figure D.2:** Multiple hidden layers are constructed with internal layers connected to the next layer, this figure shows a network with two hidden layers. Image adapted from (Prince, 2023).

self, however, some extra layers and operations are needed to implement a functional neural network, as described in the next section for the case of networks applied for image manipulation tasks.

### D.1.1 Input Layer

The input layer provides the input to the network and is defined by the size of the input image. It is formed by a series of connections to the second layer, which corresponds to the first hidden internal layer.

### D.1.2 Fully connected Layer

A layer is defined as fully connected when all the neurons of the previous layer are connected to all of the neurons of the next layer.

### D.1.3 Convolutional Layer

Convolutional Neural Network is a network used mainly in image analysis tasks.

The convolutional layer is formed by the Convolution operation (sometimes called filter) which is a dot matrix multiplication between the image and a kernel or small window of smaller size, say  $3 \times 3$  pixels. The resultant pixel is  $P = a_{1,1}w_{1,1} + a_{1,2}w_{1,2} + a_{1,3}w_{1,3} + a_{2,1}w_{2,1} + \dots + a_{3,3}w_{3,3}$ , where  $a_{m,n}$  is the pixel at location  $m, n$  and  $w_{m,n}$  are the weights, then the window kernel is moved along the original image. Some of the important hyper-parameters for a convolution operation

- a Kernel size, it is recommended to use small sizes like 1, 3, 5
- b Stride determines the number of pixels that the kernel window is moved along the operation, with a stride greater than 1 the original image is reduced.

- c Padding refers the extra pixels out of the border required to preserve the same image size, usually are filled with zeros.
- d Dilation refers to the separation between pixels selected for the multiplication, 1 means are the contiguous pixels, 2 means every two pixels are selected.
- e Number of channels, this is the number of convolutions that are computed to determine different features that the convolutional layer needs to identify.

#### **D.1.4 Maxpool Layer**

This layer generally comes after the convolution layer and is used to reduce the size of the convolution output. This operation is done by downsampling the image input, and only the most representative features are preserved. Generally, the maximum value is used.

#### **D.1.5 SoftMax Layer**

The softmax layer is generally at the end of a network and it assigns a probability  $p$ , where  $p \rightarrow [0, 1]$ , to each input, for example in a multi-class problem.

### **D.2 Network functionality: forward propagation**

The purpose of the model is to generate an outcome when input data is applied to the network. For example, a colour image, defined as a  $c = 3$  dimensional, inputs each individual neuron at the first layer, next, every hidden layer extracts general features that pass through subsequent layers until the last layers identify more complex tasks like object recognition. The final layer compiles all the information and the outcome represents the final prediction based on the type of task. The most relevant task in the image processing techniques are: binary Image classification, multi-class classification, object detection within an image, object localization, semantic segmentation in which every pixel is assigned a category, and image generation in generative models.

### **D.3 Network performance and loss function**

The output prediction  $\hat{y}$  is compared to reference data  $y$  to evaluate the model performance. This is done by the loss function  $L(y, \hat{y})$  which returns a single value,

when this is minimum the model performance is good. Some of the most common loss functions are the following:

### D.3.1 Mean Absolute Error MAE

This metric is the average magnitude of the difference between predictions and the actual value. It is the simplest error measurement

$$L_{MAE} = \frac{1}{N} \sum |y_i - \hat{y}_i| \quad (D.6)$$

### D.3.2 Mean Square Error MSE

The mean square error is the average square of the difference between predictions and the actual value and is mainly used in regression tasks. It is computed as

$$L_{MSE} = \frac{1}{N} \sum (y_i - \hat{y}_i)^2 \quad (D.7)$$

### D.3.3 Binary cross entropy loss BCE

This is the negative of the sum of the probabilities of the predictions against the total number of samples

$$L_{BCE} = -\frac{1}{M} \sum [y_i * \log(\hat{y})] \quad (D.8)$$

It is a measure of the distance between the true values and the probability of the predictions. It is widely used for binary classification models.

## D.4 Fitting the model and training

Training or fitting the model is the process of determining those parameters that minimize the corresponding loss function. This process is done by processing the error from the output to the input and computing the gradient of the loss function for each weight. This process is called back-propagation. Weights of the network are updated to minimize the loss function, one method to do this is Gradient Descent, an optimization algorithm to find the local minimum of a differentiable function.

### D.4.1 Gradient Descent

It starts with a training set of data of known input/output pairs  $(x_i, y_i)$  and seeks for parameters  $w_i$  that map the inputs  $x_i$  to the outputs  $y_i$ . The optimizer algorithm is used to find the set of  $w_i$  that minimizes the loss function. It starts from an initial set of parameter values  $\mathbf{w} = [w_0 \ w_1 \ w_2 \ \dots \ w_M]$ . The method iterates in two steps

a Compute the gradient of the loss

b Update the parameters

The gradient of the loss is calculated by

$$\frac{\partial L}{\partial \mathbf{W}} = \begin{bmatrix} \frac{\partial L}{\partial W_0} \\ \frac{\partial L}{\partial W_1} \\ \frac{\partial L}{\partial W_2} \\ \cdot \\ \cdot \\ \frac{\partial L}{\partial W_M} \end{bmatrix} \quad (\text{D.9})$$

Parameters are updated by

$$\mathbf{w} \leftarrow \mathbf{w} - \alpha \frac{\partial L}{\partial \mathbf{W}} \quad (\text{D.10})$$

In this equation, the parameter  $\alpha$  is a scalar that determines the magnitude of the change and is known as the Learning Rate. This is an iterative process until the point of the minimum of the loss function or gradient is zero (when it gets a flat surface), which occurs when the change of parameters is small enough, and the algorithm finishes. The final destination of the gradient depends on the starting point, but a local minima problem might arise, which is a point with gradient zero, if we move in any direction Loss increases but is not the overall minima, and the algorithm might end here.

### D.4.2 Stochastic Gradient Descent

Some random can be added to the algorithm to have different starting points at each iteration, this is done by selecting a random subset of training data which is known as a mini-batch. During this process, the algorithm might eventually find an

initiation point close to the global minima. This method has some advantages, it is less computationally when the algorithm works with a small subset of data at each iteration, and if it is close to the global minima, the gradient is low no matter which batch is chosen, also it responds well to new input data.

After the algorithm uses all the training data, there is a single pass through all the data which is called an epoch.

### D.4.3 Gradient descent with momentum

The change at the current step is a combination of previous weights and the current gradient. Then the term  $\mathbf{m}$  is called momentum and is added to the algorithm through the following equation

$$\mathbf{m}_{t+1} \leftarrow \beta \mathbf{m}_t + (1 - \beta) \sum_i \frac{\partial l_i[\mathbf{w}_t]}{\partial \mathbf{w}} \quad (\text{D.11})$$

$$\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \alpha \frac{\check{\mathbf{m}}_{t+1}}{\sqrt{\check{\mathbf{v}}_{t+1} + \varepsilon}} \quad (\text{D.12})$$

$i \in \mathbb{Z} \times \mathbb{Z}$  is a set containing the input/output pairs in current batch,  $\mathbf{m}$  represents an infinite weighted sum of all previous gradients and  $\beta$  controls the rate in which gradient is smoothed.

### D.4.4 Adaptive Moment estimation or ADAM

The motivation of this method is that in some cases, the gradient might be larger in a certain direction than another, and for this reason, it might be difficult to find an adequate learning rate, then the idea of normalizing the gradient. The method named adaptive moment estimation is a gradient descent method with momentum that normalizes gradient and the result is that it better converges to the overall minimum. Gradient normalization can be done by adding the quadratic gradient term  $\mathbf{v}$

$$\mathbf{m}_{t+1} \leftarrow \beta \mathbf{m}_t + (1 - \beta) \sum_i \frac{\partial l_i[\mathbf{w}_t]}{\partial \mathbf{w}} \quad (\text{D.13})$$

$$\mathbf{v}_{t+1} \leftarrow \gamma \mathbf{v}_t + (1 - \gamma) \sum_i \left( \frac{\partial l_i[\mathbf{w}_t]}{\partial \mathbf{w}} \right)^2 \quad (\text{D.14})$$

as initial values of  $\mathbf{m}_t$  and  $\mathbf{v}_t$  are zero, first calculations of  $\mathbf{m}$  and  $\mathbf{v}$  at initial iterations are biased towards zero, then this can be corrected by

$$\check{\mathbf{m}}_{t+1} \leftarrow \frac{\mathbf{m}_{t+1}}{1 - \beta^{t+1}} \quad (\text{D.15})$$

$$\check{\mathbf{v}}_{t+1} \leftarrow \frac{\mathbf{v}_{t+1}}{1 - \gamma^{t+1}} \quad (\text{D.16})$$

finally, the update of the parameters is done by

$$\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \alpha \frac{\check{\mathbf{m}}_{t+1}}{\sqrt{\check{\mathbf{v}}_{t+1} + \varepsilon}} \quad (\text{D.17})$$

The correct selection of the learning algorithm, the batch size, and the learning rate will affect the final model performance. As all these are different from the internal network parameters are named hyper-parameters.

## D.5 Model Performance

Measuring the performance of a ML model is fundamental because it gives information about its performance. Moreover, according to the specific task, an adequate metric need to be properly selected, for example, within the segmentation task, dice, sensitivity and specificity measurement help to identify if the segmented section is appropriate because it affects the overall diagnosis (Vlăsceanu et al., 2024). In addition, measurement for multi-class segmentation might result in biased results, specially when classes are imbalanced and is recommended to compute the metrics individually for each class (Müller et al., 2022). Metrics related with image segmentation are presented in this section.

### D.5.1 Binary and multi-class classification

From a binary classification basis, each pixel has one category: True Positive ( $TP$ ), True Negative ( $TN$ ), False Positive ( $FP$ ), and False Negative ( $FN$ ) when compared against the ground truth.

In the multi-class context (Müller et al., 2022), a pixel whose class was correctly predicted was counted as correct, this is the True Positive ( $TP_i$ ) and True Negative ( $TN_i$ ) per class in which  $i$  is the class. Otherwise, incorrect pixels included False

Positive ( $FP_i$ ) and False Negative ( $FN_i$ ) per class.

### D.5.2 Accuracy

Accuracy represents the proportion of all classifications that are correct, i.e., the true classifications over the total classifications, and is given by

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (\text{D.18})$$

for the multi-class case is calculated by

$$\text{Accuracy}_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (\text{D.19})$$

### D.5.3 Precision

Precision is the proportion of all the model's positive classifications that are actually positives, given by

$$\text{Precision} = \frac{TP}{TP + FP} \quad (\text{D.20})$$

### D.5.4 Sensitivity

Sensitivity indicates the proportion of true positives that the model identified correctly, given by

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (\text{D.21})$$

### D.5.5 Specificity

Specificity indicates the percentage of true negative classifications that the model identified successfully.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (\text{D.22})$$

### D.5.6 Dice coefficient

Is a measure between the similarity between two samples, for the segmentation task is used to measure the degree of agreement between the predicted and the ground truth, it is given by

$$\text{Dice coefficient} = \frac{2TP}{2TP + FP + FN} \quad (\text{D.23})$$

### D.5.7 Jaccard index

Jaccard index compares the referenced ground truth with the predicted segmentation considering an overlap between both, which means the Intersection Over the Union ( $IoU$ ), calculated as

$$\text{Jaccard Similarity Index} = IoU = \frac{TP}{TP + FP + FN}, \quad (\text{D.24})$$

### D.5.8 Receiver Operating Characteristic

The Receiver Operating Characteristic (ROC) curve followed the true positive rate (also known as recall) calculated as

$$\text{TPR} = \frac{TP_i}{TP_i + FP_i} \quad (\text{D.25})$$

against the false positive rate

$$\text{FPR} = \frac{FN}{TN + FP} \quad (\text{D.26})$$

from which the area under the curve (AUC) is calculated. A higher AUC means it is a perfect classification whereas an AUC of 0.5 means it is a random guessing.

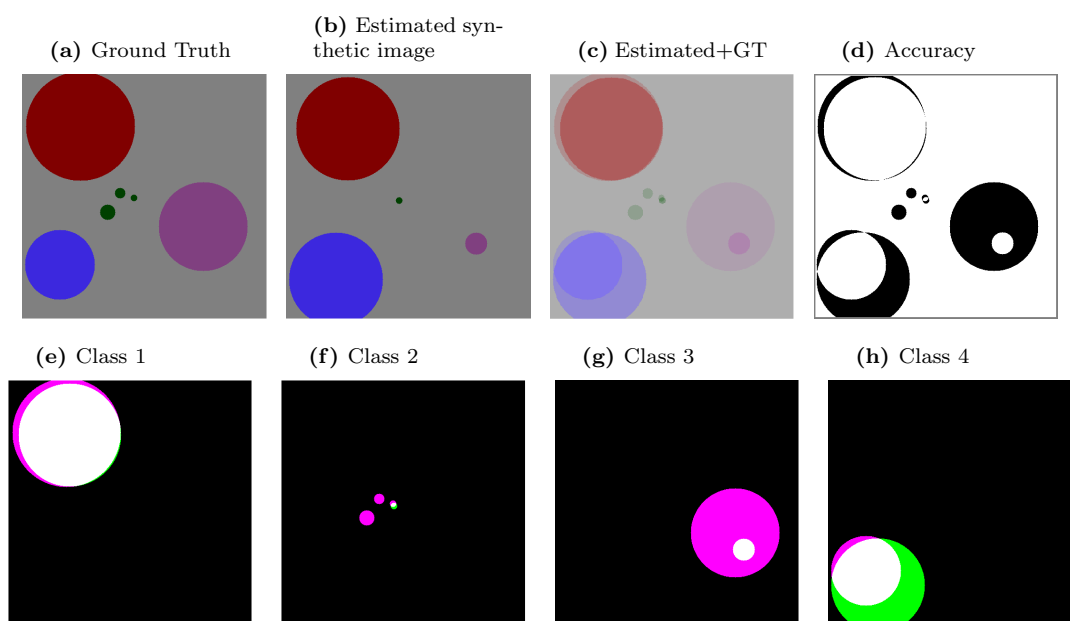
### D.5.9 F-measure

F-score or F-1 score is a measure that emphasizes that neither recall or precision is overemphasized, and is calculated by

$$F_1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (\text{D.27})$$

### D.5.10 Multi-class measurement example

For a better understanding of the multi-class performance, a synthetic example is presented in Figure D.3. In this example 4 classes are represented in colors in the synthetic image: red, green, purple, and blue. The ground truth is shown in (a) and the estimated image in (b), and the difference in the estimated is visible in the overlaid image (c). The overall accuracy is seen in (d) in white color, and the per class accuracy is seen in white in each figures (e-h). In these figures  $TP$  are white,  $FP$  are green,  $FN$  are pink, and  $TN$  are black.



**Figure D.3:** Metrics with a synthetic image with 4 classes. (a) Ground truth. The four classes are represented by colors: red, green, purple and blue. (b) Estimated image. (c) Ground truth and estimated images overlaid. True Positives (TP) are white, True Negatives (TN) are black, False Positives (FP) are green, False Negatives (FN) are pink. (d) Illustration of overall accuracy; pixels correctly estimated are shown in white and incorrect are shown in black. (e-h) Illustration of per-class TP, TN, FP, FN.



# References

- M. Peikari, S. Salama, S. Nofech-Mozes, and A. L. Martel, “Automatic cellularity assessment from post-treated breast surgical specimens,” *Cytometry. Part A: The Journal of the International Society for Analytical Cytology*, vol. 91, no. 11, pp. 1078–1087, 2017.
- Z. Guo, H. Liu, H. Ni, X. Wang, M. Su, W. Guo, K. Wang, T. Jiang, and Y. Qian, “A Fast and Refined Cancer Regions Segmentation Framework in Whole-slide Breast Pathological Images,” *Scientific Reports*, vol. 9, no. 1, p. 882, Jan. 2019, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s41598-018-37492-9>
- P. Weitz, M. Valkonen, L. Solorzano, C. Carr, K. Kartasalo, C. Boissin, S. Koivukoski, A. Kuusela, D. Rasic, Y. Feng *et al.*, “Acrobat—a multi-stain breast cancer histological whole-slide-image data set from routine diagnostics for computational pathology,” *arXiv preprint arXiv:2211.13621*, 2022.
- A. G. Howard, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.
- S. Akbar, M. Peikari, S. Salama, A. Y. Panah, S. Nofech-Mozes, and A. L. Martel, “Automated and Manual Quantification of Tumour Cellularity in Digital Slides for Tumour Burden Assessment,” *Scientific Reports*, vol. 9, no. 1, pp. 1–9, Oct. 2019. [Online]. Available: <https://www.nature.com/articles/s41598-019-50568-4>
- P. Dey, *Basic and Advanced Laboratory Techniques in Histopathology and Cytology*, 2nd ed. Singapore: Springer Nature Singapore, 2022.
- M. Veta, J. P. W. Pluim, P. J. v. Diest, and M. A. Viergever, “Breast Cancer Histopathology Image Analysis: A Review,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1400–1411, May 2014.

- H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review-current status and future potential," *IEEE reviews in biomedical engineering*, vol. 7, pp. 97–114, 2013, publisher: IEEE.
- M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009, conference Name: IEEE Reviews in Biomedical Engineering. [Online]. Available: <https://ieeexplore.ieee.org/document/5299287>
- M. S. Hosseini, B. E. Bejnordi, V. Q.-H. Trinh, L. Chan, D. Hasan, X. Li, S. Yang, T. Kim, H. Zhang, T. Wu, K. Chinniah, S. Maghsoudlou, R. Zhang, J. Zhu, S. Khaki, A. Buin, F. Chaji, A. Salehi, B. N. Nguyen, D. Samaras, and K. N. Plataniotis, "Computational pathology: A survey review and the way forward," *Journal of Pathology Informatics*, vol. 15, p. 100357, Dec. 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2153353923001712>
- A. H. Song, G. Jaume, D. F. K. Williamson, M. Y. Lu, A. Vaidya, T. R. Miller, and F. Mahmood, "Artificial intelligence for digital and computational pathology," *Nature Reviews Bioengineering*, vol. 1, no. 12, pp. 930–949, Dec. 2023, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s44222-023-00096-8>
- E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001.
- E. Michael, H. Ma, H. Li, F. Kulwa, and J. Li, "Breast Cancer Segmentation Methods: Current Status and Future Potentials," *BioMed Research International*, vol. 2021, p. e9962109, Jul. 2021, publisher: Hindawi. [Online]. Available: <https://www.hindawi.com/journals/bmri/2021/9962109/>
- L. E. Boucheron, *Object-and spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer*. University of California at Santa Barbara, 2008.
- H. H. Rashidi, N. K. Tran, E. V. Betts, L. P. Howell, and R. Green, "Artificial

- Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods,” *Academic Pathology*, vol. 6, p. 2374289519873088, Jan. 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2374289521001573>
- S. Shalev-Shwartz and S. Ben-David, *Clustering*. Cambridge University Press, 2014, p. 264–277.
- S. Takano, “Chapter 3 - machine learning and its implementation,” in *Thinking Machines*, S. Takano, Ed. Academic Press, 2021, pp. 49–66. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B978012818279600013X>
- I. Goodfellow, *Deep learning*. MIT press, 2016, vol. 196.
- D. Komura and S. Ishikawa, “Machine Learning Methods for Histopathological Image Analysis,” *Computational and Structural Biotechnology Journal*, vol. 16, pp. 34–42, Jan. 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2001037017300867>
- A. Madabhushi and G. Lee, “Image analysis and machine learning in digital pathology: Challenges and opportunities,” *Medical Image Analysis*, vol. 33, pp. 170–175, 2016.
- A. H. Beck, A. R. Sangoi, S. Leung, R. J. Marinelli, T. O. Nielsen, M. J. van de Vijver, R. B. West, M. van de Rijn, and D. Koller, “Systematic analysis of breast cancer morphology uncovers stromal features associated with survival,” *Science Translational Medicine*, vol. 3, no. 108, p. 108ra113, Nov. 2011.
- A. B. Pardee and G. S. Stein, *The biology and treatment of cancer: understanding cancer*. Hoboken, N.J: Wiley-Blackwell, 2009.
- R. I. Aqeilan, N. Zanasi, and C. M. Croce, “Environmental, Genetic, and Viral Causes of Cancer,” in *The Biology and Treatment of Cancer*. John Wiley & Sons, Ltd, 2008, pp. 35–56.
- B. Weigelt, F. C. Geyer, and J. S. Reis-Filho, “Histological types of breast cancer: how special are they?” *Molecular oncology*, vol. 4, no. 3, pp. 192–208, 2010.
- R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, “Cancer statistics, 2023,” *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17–48, 2023,

\_eprint: <https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21763>.  
[Online]. Available: <https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21763>

Y.-n. Zhang, K.-R. Xia, C.-Y. Li, B.-L. Wei, and B. Zhang, “Review of breast cancer pathological image processing,” *BioMed research international*, vol. 2021, no. 1, p. 1994764, 2021.

E. Michael, H. Ma, H. Li, F. Kulwa, and J. Li, “Breast Cancer Segmentation Methods: Current Status and Future Potentials,” *BioMed Research International*, vol. 2021, p. 9962109, Jul. 2021, publisher: Hindawi. [Online]. Available: <https://doi.org/10.1155/2021/9962109>

H. Su, F. Liu, Y. Xie, F. Xing, S. Meyyappan, and L. Yang, “Region segmentation in histopathological breast cancer images using deep convolutional neural network,” in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 55–58.

B. M. Priego-Torres, D. Sanchez-Morillo, M. A. Fernandez-Granero, and M. Garcia-Rojo, “Automatic segmentation of whole-slide H&E stained breast histopathology images using a deep convolutional neural network architecture,” *Expert Systems with Applications*, vol. 151, p. 113387, Aug. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417420302116>

J. L. Champaign and G. J. Cederbom, “Advances in Breast Cancer Detection with Screening Mammography,” *Ochsner Journal*, vol. 2, no. 1, pp. 33–35, Jan. 2000, publisher: Ochsner Journal Section: Featured Articles. [Online]. Available: <https://www.ochsnerjournal.org/content/2/1/33>

P. Sasieni, R. Smittenaar, E. Hubbell, J. Broggio, R. D. Neal, and C. Swanton, “Modelled mortality benefits of multi-cancer early detection screening in England,” *British Journal of Cancer*, vol. 129, no. 1, pp. 72–80, Jul. 2023. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10307803/>

F. Theissig, K. Kunze, G. Haroske, and W. Meyer, “Histological grading of breast cancer: interobserver, reproducibility and prognostic significance,” *Pathology-Research and Practice*, vol. 186, no. 6, pp. 732–736, 1990.

- D. Crosby, N. Lyons, E. Greenwood, S. Harrison, S. Hiom, J. Moffat, T. Quallo, E. Samuel, and I. Walker, “A roadmap for the early detection and diagnosis of cancer,” *The Lancet Oncology*, vol. 21, no. 11, pp. 1397–1399, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1470204520305933>
- M. A. Ortega-Ruiz, C. Karabağ, V. G. Garduño, and C. C. Reyes-Aldasoro, “Morphological estimation of cellularity on neo-adjuvant treated breast cancer histological images,” *Journal of Imaging*, vol. 6, no. 10, 2020. [Online]. Available: <https://www.mdpi.com/2313-433X/6/10/101>
- O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *arXiv:1505.04597 [cs]*, May 2015, arXiv: 1505.04597. [Online]. Available: <http://arxiv.org/abs/1505.04597>
- M. A. Ortega-Ruiz, C. Karabağ, E. Roman-Rangel, and C. C. Reyes-Aldasoro, “Drd-unet, a unet-like architecture for multi-class breast cancer semantic segmentation,” *IEEE Access*, vol. 12, pp. 40 412–40 424, 2024.
- J. Guinebretiere, E. Menet, A. Tardivon, P. Cherel, and D. Vanel, “Normal and pathological breast, the histological basis,” *European journal of radiology*, vol. 54, no. 1, pp. 6–14, 2005.
- S. Pandya and R. G. Moore, “Breast development and anatomy,” *Clinical obstetrics and gynecology*, vol. 54, no. 1, pp. 91–95, 2011.
- T. de Bel, G. Litjens, J. Ogony, M. Stallings-Mann, J. M. Carter, T. Hilton, D. C. Radisky, R. A. Vierkant, B. Broderick, T. L. Hoskin *et al.*, “Automated quantification of levels of breast terminal duct lobular (tdlu) involution using deep learning,” *NPJ breast cancer*, vol. 8, no. 1, p. 13, 2022.
- H. Zolfagharnasab, S. Bessa, S. Oliveira, P. Faria, J. Teixeira, J. Cardoso, and H. Oliveira, “A regression model for predicting shape deformation after breast conserving surgery,” *Sensors*, vol. 18, p. 167, 01 2018.
- M. Thomas, E. D. Kelly, J. Abraham, and M. Kruse, “Invasive lobular breast cancer: A review of pathogenesis, diagnosis, management, and future directions of early stage disease,” in *Seminars in oncology*, vol. 46, no. 2. Elsevier, 2019, pp. 121–132.

- J. Wang, B. Li, M. Luo, J. Huang, K. Zhang, S. Zheng, S. Zhang, and J. Zhou, “Progression from ductal carcinoma in situ to invasive breast cancer: molecular features and clinical significance,” *Signal Transduction and Targeted Therapy*, vol. 9, no. 1, p. 83, 2024.
- W. B. Coleman, “Breast Ductal Carcinoma in Situ: Precursor to Invasive Breast Cancer,” *The American Journal of Pathology*, vol. 189, no. 5, pp. 942–945, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0002944019301920>
- A. Campilho, F. Karray, and B. t. H. Romeny, *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings*. Springer, Jun. 2018, google-Books-ID: YHJgDwAAQBAJ.
- R. H. Grobstein, *The breast cancer book: what you need to know to make informed decisions*. New Haven: Yale University Press, 2005.
- R. Iacob, E. R. Iacob, E. R. Stoicescu, D. M. Ghenciu, D. M. Cocolea, A. Constantinescu, L. A. Ghenciu, and D. L. Manolescu, “Evaluating the Role of Breast Ultrasound in Early Detection of Breast Cancer in Low- and Middle-Income Countries: A Comprehensive Narrative Review,” *Bioengineering*, vol. 11, no. 3, p. 262, Mar. 2024. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10968105/>
- S. H. Jafari, Z. Saadatpour, A. Salmaninejad, F. Momeni, M. Mokhtari, J. S. Nahand, M. Rahmati, H. Mirzaei, and M. Kianmehr, “Breast cancer diagnosis: Imaging techniques and biochemical markers,” *Journal of cellular physiology*, vol. 233, no. 7, pp. 5200–5213, 2018.
- R. Sood, A. F. Rositch, D. Shakoor, E. Ambinder, K.-L. Pool, E. Pollack, D. J. Mollura, L. A. Mullen, and S. C. Harvey, “Ultrasound for Breast Cancer Detection Globally: A Systematic Review and Meta-Analysis,” *Journal of Global Oncology*, vol. 5, p. JGO.19.00127, Aug. 2019. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6733207/>
- K. C. Oeffinger, E. T. Fontham, R. Etzioni, A. Herzig, J. S. Michaelson, Y.-C. T. Shih, L. C. Walter, T. R. Church, C. R. Flowers, S. J. LaMonte *et al.*, “Breast can-

- cer screening for women at average risk: 2015 guideline update from the american cancer society,” *Jama*, vol. 314, no. 15, pp. 1599–1614, 2015.
- J. E. Aldrich, “Basic physics of ultrasound imaging,” *Critical care medicine*, vol. 35, no. 5, pp. S131–S137, 2007.
- R. Bartrum Jr, “Ultrasound instrumentation.” *Critical reviews in diagnostic Imaging*, vol. 25, no. 3, pp. 279–303, 1986.
- G. F. Tadesse, E. M. Tegaw, and E. K. Abdisa, “Diagnostic performance of mammography and ultrasound in breast cancer: a systematic review and meta-analysis,” *Journal of Ultrasound*, vol. 26, no. 2, pp. 355–367, 2023.
- S. E. Harms, “Mri in breast cancer diagnosis and treatment,” *Current problems in diagnostic radiology*, vol. 25, no. 6, pp. 192–215, 1996.
- R. M. Mann, N. Cho, and L. Moy, “Breast mri: state of the art,” *Radiology*, vol. 292, no. 3, pp. 520–536, 2019.
- Y. Gao, B. Reig, L. Heacock, D. L. Bennett, S. L. Heller, and L. Moy, “Magnetic resonance imaging in screening of breast cancer,” *Radiologic Clinics of North America*, vol. 59, no. 1, p. 85, 2020.
- R. Kradin, V. Deshpande, and A. J. Iafrate, “General principles in the diagnosis of infection,” *Diagnostic pathology of infectious disease*, p. 3, 2017.
- H. A. Alturkistani, F. M. Tashkandi, and Z. M. Mohammedsaleh, “Histological stains: a literature review and case study,” *Global journal of health science*, vol. 8, no. 3, p. 72, 2015.
- A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, “Hematoxylin and eosin staining of tissue and cell sections,” *Cold spring harbor protocols*, vol. 2008, no. 5, pp. pdb–prot4986, 2008.
- H. M. Hussaini, B. Seo, and A. M. Rich, “Immunohistochemistry and immunofluorescence,” in *Oral Biology: Molecular Techniques and Applications*. Springer, 2022, pp. 439–450.
- S. Maleki, O. Dorokhova, J. Sunkara, K. Schlesinger, M. Suhrland, and M. H. Oktay, “Estrogen, progesterone, and her-2 receptor immunostaining in cytology: the

- effect of varied fixation on human breast cancer cells,” *Diagnostic Cytopathology*, vol. 41, no. 10, pp. 864–870, 2013.
- C. Bertrand, S. Chan-Lang, S. Petit, I. Villa, R. Fick, and S. B. Hadj, “Tumoral aware deep learning algorithm for automatic ki67 scoring,” *Journal of Pathology Informatics*, vol. 13, p. 100050, 2022.
- J. W. Prichard, “Overview of automated immunohistochemistry,” *Archives of pathology & laboratory medicine*, vol. 138, no. 12, pp. 1578–1582, 2014.
- M. E. Vandenberghe, M. L. Scott, P. W. Scorer, M. Söderberg, D. Balcerzak, and C. Barker, “Relevance of deep learning to facilitate the diagnosis of her2 status in breast cancer,” *Scientific reports*, vol. 7, no. 1, p. 45938, 2017.
- M. P. Humphries, D. Kaye, G. Stankeviciute, J. Halliwell, A. I. Wright, D. Bansal, D. Brettle, and D. Treanor, “Development of a multi-scanner facility for data acquisition for digital pathology artificial intelligence,” *The Journal of Pathology*, vol. 264, no. 1, pp. 80–89, 2024.
- M. Amgad, H. Elfandy, H. Hussein, L. A. Atteya, M. A. T. Elsebaie, L. S. Abo Elnasr, R. A. Sakr, H. S. E. Salem, A. F. Ismail, A. M. Saad, J. Ahmed, M. A. T. Elsebaie, M. Rahman, I. A. Ruhban, N. M. Elgazar, Y. Alagha, M. H. Osman, A. M. Alhusseiny, M. M. Khalaf, A.-A. F. Younes, A. Abdulkarim, D. M. Younes, A. M. Gadallah, A. M. Elkashash, S. Y. Fala, B. M. Zaki, J. Beezley, D. R. Chittajallu, D. Manthey, D. A. Gutman, and L. A. D. Cooper, “Structured crowdsourcing enables convolutional segmentation of histology images,” *Bioinformatics*, vol. 35, no. 18, pp. 3461–3467, Sep. 2019. [Online]. Available: <https://doi.org/10.1093/bioinformatics/btz083>
- P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman, J. A. James, M. Salto-Tellez, and P. W. Hamilton, “QuPath: Open source software for digital pathology image analysis,” *Scientific Reports*, vol. 7, no. 1, p. 16878, Dec. 2017, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s41598-017-17204-5>
- M. Rasoolijaberi, M. Babaei, A. Riasatian, S. Hemati, P. Ashrafi, R. Gonzalez, and H. R. Tizhoosh, “Multi-Magnification Image Search in Digital Pathology,” *IEEE*

- Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4611–4622, Sep. 2022, conference Name: IEEE Journal of Biomedical and Health Informatics. [Online]. Available: <https://ieeexplore.ieee.org/document/9793694>
- A. Goode, B. Gilbert, J. Harkes, D. Jukic, and M. Satyanarayanan, “Openslide: A vendor-neutral software foundation for digital pathology,” *Journal of pathology informatics*, vol. 4, no. 1, p. 27, 2013.
- MATLAB, *version 9.14.0 (R2023a)*. Natick, Massachusetts: The MathWorks Inc., 2023.
- S. J. Prince, *Understanding Deep Learning*. MIT Press, 2023. [Online]. Available: <http://udlbook.com>
- G. Karalis, “Decision trees and applications,” *GeNeDis 2018: Computational Biology and Bioinformatics*, pp. 239–242, 2020.
- C. Li and B. Wang, “Fisher linear discriminant analysis,” *CCIS Northeastern University*, vol. 6, 2014.
- M. Bansal, A. Goyal, and A. Choudhary, “A comparative analysis of k-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning,” *Decision Analytics Journal*, vol. 3, p. 100071, 2022.
- C. Cortes, “Support-vector networks,” *Machine Learning*, 1995.
- J. Elith, J. R. Leathwick, and T. Hastie, “A working guide to boosted regression trees,” *Journal of animal ecology*, vol. 77, no. 4, pp. 802–813, 2008.
- D. R. Cox, “The regression analysis of binary sequences,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 20, no. 2, pp. 215–232, 1958.
- G. Palm, “Warren mcculloch and walter pitts: A logical calculus of the ideas immanent in nervous activity,” in *Brain Theory*, G. Palm and A. Aertsen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1986, pp. 229–230.
- F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychological Review*, vol. 65, no. 6, pp. 386–408,

- Nov. 1958, publisher: American Psychological Association. [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=pdh&AN=1959-09865-001&site=ehost-live&scope=site&custid=s1089299>
- M. Minsky and S. Papert, “An introduction to computational geometry,” *Cambridge tiass.*, *HIT*, vol. 479, no. 480, p. 104, 1969.
- K. Fukushima, “Cognitron: A self-organizing multilayered neural network,” *Biological Cybernetics*, vol. 20, no. 3, pp. 121–136, Sep. 1975. [Online]. Available: <https://doi.org/10.1007/BF00342633>
- D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- A. Kurenkov, “A brief history of neural nets and deep learning,” *Skynet Today*, 2020.
- G. E. Hinton, S. Osindero, and Y.-W. Teh, “A Fast Learning Algorithm for Deep Belief Nets,” *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006. [Online]. Available: <https://doi.org/10.1162/neco.2006.18.7.1527>
- Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, conference Name: Proceedings of the IEEE. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/726791>
- A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)
- A. Heyden, D. Laurendeau, M. Felsberg, and M. Borga, “Proceedings. 22nd international conference on pattern recognition icpr 2014, 24-28 august 2014, stockholm, sweden,” in *22nd International Conference on Pattern Recognition ICPR 2014, 24-28 August 2014, Stockholm, Sweden*. IEEE conference proceedings, 2014.
- R. Ludovic, R. Daniel, L. Nicolas, K. Maria, I. Humayun, K. Jacques, C. Frédérique, G. Catherine *et al.*, “Mitosis detection in breast cancer histological images an icpr 2012 contest,” *Journal of pathology informatics*, vol. 4, no. 1, p. 8, 2013.

- F. Dong, H. Irshad, E.-Y. Oh, M. F. Lerwill, E. F. Brachtel, N. C. Jones, N. W. Knoblauch, L. Montaser-Kouhsari, N. B. Johnson, L. K. F. Rao, B. Faulkner-Jones, D. C. Wilbur, S. J. Schnitt, and A. H. Beck, “Computational Pathology to Discriminate Benign from Malignant Intraductal Proliferations of the Breast,” *PLOS ONE*, vol. 9, no. 12, p. e114885, Dec. 2014, publisher: Public Library of Science. [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114885>
- A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, and A. Madabhushi, “Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks,” in *Medical Imaging 2014: Digital Pathology*, vol. 9041. SPIE, Mar. 2014, p. 904103. [Online]. Available: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9041/904103/Automatic-detection-of-invasive-ductal-carcinoma-in-whole-slide-images/10.1117/12.2043872.full>
- H. Chen, Q. Dou, X. Wang, J. Qin, and P. Heng, “Mitosis detection in breast cancer histology images via deep cascaded networks,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, no. 1, 2016.
- A. Janowczyk and A. Madabhushi, “Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases,” *Journal of Pathology Informatics*, vol. 7, no. 1, p. 29, Jan. 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2153353922005478>
- B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol *et al.*, “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.
- D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, “Deep Learning for Identifying Metastatic Breast Cancer,” Jun. 2016, arXiv:1606.05718 [cs, q-bio]. [Online]. Available: <http://arxiv.org/abs/1606.05718>
- S. Mehta, E. Mercan, J. Bartlett, D. Weaver, J. G. Elmore, and L. Shapiro, “Y-net: joint segmentation and classification for diagnosis of breast biopsy images,” in

*Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II 11.* Springer, 2018, pp. 893–901.

J. G. Elmore, G. M. Longton, P. A. Carney, B. M. Geller, T. Onega, A. N. Tosteson, H. D. Nelson, M. S. Pepe, K. H. Allison, S. J. Schnitt *et al.*, “Diagnostic concordance among pathologists interpreting breast biopsy specimens,” *Jama*, vol. 313, no. 11, pp. 1122–1132, 2015.

M. Pereira, “Breakhis-breast cancer histopathological database,” *Mendeley Data*, vol. 1, 2023.

F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using convolutional neural networks,” in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 2560–2567.

Y. Yari, T. V. Nguyen, and H. T. Nguyen, “Deep learning applied for histological diagnosis of breast cancer,” *IEEE Access*, vol. 8, pp. 162 432–162 448, 2020.

Y. Benhammou, B. Achchab, F. Herrera, and S. Tabik, “BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights,” *Neurocomputing*, vol. 375, pp. 9–24, Jan. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231219313128>

G. Aresta, T. Araújo, S. Kwok, S. S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, G. Fernandez, J. Zeineh, M. Kohl, C. Walz, F. Ludwig, S. Braunewell, M. Baust, Q. D. Vu, M. N. N. To, E. Kim, J. T. Kwak, S. Galal, V. Sanchez-Freire, N. Brancati, M. Frucci, D. Riccio, Y. Wang, L. Sun, K. Ma, J. Fang, I. Kone, L. Boulmane, A. Campilho, C. Eloy, A. Polónia, and P. Aguiar, “Bach: Grand challenge on breast cancer histology images,” *Medical Image Analysis*, vol. 56, pp. 122–139, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841518307941>

O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec. 2015. [Online]. Available: <https://doi.org/10.1007/s11263-015-0816-y>

- A. A. Balasubramanian, S. M. A. Al-Heejawi, A. Singh, A. Breggia, B. Ahmad, R. Christman, S. T. Ryan, and S. Amal, “Ensemble Deep Learning-Based Image Classification for Breast Cancer Subtype and Invasiveness Diagnosis from Whole Slide Image Histopathology,” *Cancers*, vol. 16, no. 12, p. 2222, Jun. 2024. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11201924/>
- G. Aresta, T. Araújo, S. Kwok, S. S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, G. Fernandez, J. Zeineh, M. Kohl, C. Walz, F. Ludwig, S. Braunewell, M. Baust, Q. D. Vu, M. N. N. To, E. Kim, J. T. Kwak, S. Galal, V. Sanchez-Freire, N. Brancati, M. Frucci, D. Riccio, Y. Wang, L. Sun, K. Ma, J. Fang, I. Kone, L. Boulmane, A. Campilho, C. Eloy, A. Polónia, and P. Aguiar, “BACH: Grand challenge on breast cancer histology images,” *Medical Image Analysis*, vol. 56, pp. 122–139, Aug. 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841518307941>
- Y. Celik, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, “Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images,” *Pattern Recognition Letters*, vol. 133, pp. 232–239, May 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167865520300891>
- H. Bolhasani, E. Amjadi, M. Tabatabaeian, and S. J. Jassbi, “A histopathological image dataset for grading breast invasive ductal carcinomas,” *Informatics in Medicine Unlocked*, vol. 19, p. 100341, Jan. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352914820300757>
- C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- V. Andrearczyk, V. Oreiller, S. Boughdad, C. C. L. Rest, H. Elhalawani, M. Jreige, J. O. Prior, M. Vallières, D. Visvikis, M. Hatt *et al.*, “Overview of the hecktor challenge at miccai 2021: automatic head and neck tumor segmentation and outcome prediction in pet/ct images,” in *3D head and neck tumor segmentation in PET/CT challenge*. Springer, 2021, pp. 1–37.
- T. Liu, Y. Su, J. Zhang, T. Wei, and Z. Xiao, “3D U-Net Applied to Simple Attention Module for Head and Neck Tumor Segmentation in PET and CT Images,” in

- Head and Neck Tumor Segmentation and Outcome Prediction*, V. Andrearczyk, V. Oreiller, M. Hatt, and A. Depeursinge, Eds. Cham: Springer International Publishing, 2022, pp. 99–108.
- Y. Yuan, “Automatic Head and Neck Tumor Segmentation in PET/CT with Scale Attention Network,” in *Head and Neck Tumor Segmentation*, V. Andrearczyk, V. Oreiller, and A. Depeursinge, Eds. Cham: Springer International Publishing, 2021, pp. 44–52.
- X. Huo, K. H. Ong, K. W. Lau, L. Gole, D. M. Young, C. L. Tan, X. Zhu, C. Zhang, Y. Zhang, L. Li, H. Han, H. Lu, J. Zhang, J. Hou, H. Zhao, H. Gan, L. Yin, X. Wang, X. Chen, H. Lv, H. Cao, X. Yu, Y. Shi, Z. Huang, G. Marini, J. Xu, B. Liu, B. Chen, Q. Wang, K. Gui, W. Shi, Y. Sun, W. Chen, D. Cao, S. J. Sanders, H. K. Lee, S. S.-S. Hue, W. Yu, and S. Y. Tan, “A comprehensive AI model development framework for consistent Gleason grading,” *Communications Medicine*, vol. 4, no. 1, pp. 1–13, May 2024, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s43856-024-00502-1>
- N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, p. 62–66, Jan 1979.
- J. Ghosh and A. Liu, “K-means,” in *The top ten algorithms in data mining*. Chapman and Hall/CRC, 2009, pp. 35–50.
- F. Meyer, “Topographic distance and watershed lines,” *Signal processing*, vol. 38, no. 1, pp. 113–125, 1994.
- W. F. Symmans, F. Peintinger, C. Hatzis, R. Rajan, H. Kuerer, V. Valero, L. Assad, A. Poniecka, B. Hennessy, M. Green *et al.*, “Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy,” *Journal of Clinical Oncology*, vol. 25, no. 28, pp. 4414–4422, 2007.
- I. Fondón, A. Sarmiento, A. I. García, M. Silvestre, C. Eloy, A. Polónia, and P. Aguiar, “Automatic classification of tissue malignancy for breast carcinoma diagnosis,” *Computers in Biology and Medicine*, vol. 96, pp. 41–51, May 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010482518300544>

P. Jaccard, “Étude comparative de la distribution florale dans une portion des alpes et des jura,” *Bulletin del la Soci  t   Vaudoise des Sciences Naturelles*, vol. 37, p. 547–579, 1901.

R. C. Gonzalez, *Digital image processing*. Pearson education india, 2009.

B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A Training Algorithm for Optimal Margin Classifiers,” in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*. ACM Press, 1992, pp. 144–152.

E. Fix, *Discriminatory analysis: nonparametric discrimination, consistency properties*. USAF school of Aviation Medicine, 1985, vol. 1.

Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

Z.-H. Zhou, *Machine learning*. Springer nature, 2021.

A. M. Khan, H. El-Daly, E. Simmons, and N. M. Rajpoot, “Hymap: A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images,” *Journal of pathology informatics*, vol. 4, no. 2, p. 1, 2013.

B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, J. A. W. M. van der Laak, and a. t. C. Consortium, “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” vol. 318, no. 22, pp. 2199–2210, eprint: [https://jamanetwork.com/journals/jama/articlepdf/2665774/jama\\_ehteshami\\_bejnordi\\_2017\\_oi\\_170](https://jamanetwork.com/journals/jama/articlepdf/2665774/jama_ehteshami_bejnordi_2017_oi_170) [Online]. Available: <https://doi.org/10.1001/jama.2017.14585>

M.-A. Khalil, Y.-C. Lee, H.-C. Lien, Y.-M. Jeng, and C.-W. Wang, “Fast segmentation of metastatic foci in h&e whole-slide images for breast cancer diagnosis,” *Diagnostics*, vol. 12, no. 4, p. 990, 2022.

Y. Benmabrouk, M. Gasmi, H. Bendjenna, and A. Nadjah, “Semantic segmentation of breast cancer histopathology images using deep learning,” in *2022 4th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*. IEEE, 2022, pp. 1–7.

- F. Demir, "Deepbreastnet: A novel and robust approach for automated breast cancer detection from histopathological images," *Biocybernetics and biomedical engineering*, vol. 41, no. 3, pp. 1123–1139, 2021.
- Y. Liu, T. Zhen, Y. Fu, Y. Wang, Y. He, A. Han, and H. Shi, "Ai-powered segmentation of invasive carcinoma regions in breast cancer immunohistochemical whole-slide images," *Cancers*, vol. 16, no. 1, p. 167, 2023.
- K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015, arXiv:1512.03385 [cs]. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- H. Aljuaid, N. Alturki, N. Alsubaie, L. Cavallaro, and A. Liotta, "Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning," *Computer Methods and Programs in Biomedicine*, vol. 223, p. 106951, 2022.
- Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- W. Bulten, H. Pinckaers, H. van Boven, R. Vink, T. de Bel, B. van Ginneken, J. van der Laak, C. Hulsbergen-van de Kaa, and G. Litjens, "Automated deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study," *The Lancet Oncology*, vol. 21, no. 2, pp. 233–241, 2020.
- M. Krithika alias AnbuDevi and K. Suganthi, "Review of Semantic Segmentation of Medical Images Using Modified Architectures of UNET," *Diagnostics*, vol. 12, no. 12, p. 3064, Dec. 2022, number: 12 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/2075-4418/12/12/3064>
- E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA engineer*, vol. 29, no. 6, pp. 33–41, 1984.
- E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, Apr. 2017, conference Name: IEEE

- Transactions on Pattern Analysis and Machine Intelligence. [Online]. Available: <https://ieeexplore.ieee.org/document/7478072>
- V. Ummadi, “U-Net and its variants for Medical Image Segmentation : A short review,” Apr. 2022, arXiv:2204.08470 [cs, eess]. [Online]. Available: <http://arxiv.org/abs/2204.08470>
- X. Han, “Automatic Liver Lesion Segmentation Using A Deep Convolutional Neural Network Method,” *Medical Physics*, vol. 44, no. 4, pp. 1408–1419, Apr. 2017, arXiv:1704.07239 [cs]. [Online]. Available: <http://arxiv.org/abs/1704.07239>
- R. A. Khan, Y. Luo, and F.-X. Wu, “RMS-UNet: Residual multi-scale UNet for liver and lesion segmentation,” *Artificial Intelligence in Medicine*, vol. 124, p. 102231, Feb. 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0933365721002244>
- Y. Ding, J. Sohn, M. Kawczynski, H. Trivedi, R. Harnish, N. Jenkins, D. Lituiev, T. Copeland, M. Aboian, C. Aparici, S. Behr, R. Flavell, S.-Y. Huang, K. Zalocusky, L. Nardo, Y. Seo, R. Hawkins, M. Pampaloni, D. Hadley, and B. Franc, “A deep learning model to predict a diagnosis of Alzheimer disease by using  $^{18}\text{F}$ -FDG PET of the brain,” *Radiology*, vol. 290, no. 3, pp. 456–464, mar 2019.
- P. Ahmad, H. Jin, R. Alroobaea, S. Qamar, R. Zheng, F. Alnajjar, and F. Aboudi, “Mh unet: A multi-scale hierarchical based architecture for medical image segmentation,” *IEEE Access*, vol. 9, pp. 148 384–148 408, 2021.
- Y. Zhang, J. Wu, Y. Liu, Y. Chen, E. X. Wu, and X. Tang, “MI-UNet: Multi-Inputs UNet Incorporating Brain Parcellation for Stroke Lesion Segmentation From T1-Weighted Magnetic Resonance Images,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 2, pp. 526–535, Feb. 2021, conference Name: IEEE Journal of Biomedical and Health Informatics. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9099078>
- R. Sun, E. Limkin, M. Vakalopoulou, L. Dercie, S. Champiat, S. Rong Han, L. Verlingue, D. Brandao, A. Lancia, S. Ammari, A. Hollebecque, J.-Y. Scoazec, A. marabelle, C. Massard, J.-C. Soria, C. Robert, N. Paragios, E. Deutsch, and C. Ferté, “A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retro-

- spective multicohort study,” *The Lancet Oncology*, vol. 19, no. 9, pp. 1180–1191, sep 2018.
- Y. Yang, W. Wan, S. Huang, X. Zhong, and X. Kong, “RADCU-Net: residual attention and dual-supervision cascaded U-Net for retinal blood vessel segmentation,” *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 5, pp. 1605–1620, May 2023. [Online]. Available: <https://doi.org/10.1007/s13042-022-01715-3>
- J. Wang, X. Li, P. Lv, and C. Shi, “SERR-U-Net: Squeeze-and-Excitation Residual and Recurrent Block-Based U-Net for Automatic Vessel Segmentation in Retinal Image,” *Computational and Mathematical Methods in Medicine*, vol. 2021, p. e5976097, Aug. 2021, publisher: Hindawi. [Online]. Available: <https://www.hindawi.com/journals/cmmm/2021/5976097/>
- J. Hu, H. Wang, S. Gao, M. Bao, T. Liu, Y. Wang, and J. Zhang, “S-UNet: A Bridge-Style U-Net Framework With a Saliency Mechanism for Retinal Vessel Segmentation,” *IEEE Access*, vol. 7, pp. 174 167–174 177, 2019, conference Name: IEEE Access. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8842560>
- Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, “RIC-Unet: An Improved Neural Network Based on Unet for Nuclei Segmentation in Histology Images,” *IEEE Access*, vol. 7, pp. 21 420–21 428, 2019, conference Name: IEEE Access. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8632907>
- H. Yuan, Z. Liu, Y. Shao, and M. Liu, “ResD-Unet Research and Application for Pulmonary Artery Segmentation,” *IEEE Access*, vol. 9, pp. 67 504–67 511, 2021, conference Name: IEEE Access. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9408588>
- A. Negi, A. N. J. Raj, R. Nersisson, Z. Zhuang, and M. Murugappan, “RDA-UNET-WGAN: An Accurate Breast Ultrasound Lesion Segmentation Using Wasserstein Generative Adversarial Networks,” *Arabian Journal for Science and Engineering*, vol. 45, no. 8, pp. 6399–6410, Aug. 2020. [Online]. Available: <https://doi.org/10.1007/s13369-020-04480-z>
- T. Alam, W.-C. Shia, F.-R. Hsu, and T. Hassan, “Improving Breast Cancer

- Detection and Diagnosis through Semantic Segmentation Using the Unet3+ Deep Learning Framework,” *Biomedicines*, vol. 11, no. 6, p. 1536, Jun. 2023, number: 6 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/2227-9059/11/6/1536>
- L. Xu, H. Xue, M. Bennamoun, F. Boussaid, and F. Sohel, “Atrous convolutional feature network for weakly supervised semantic segmentation,” *Neurocomputing*, vol. 421, p. 115–126, Jan. 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231220314533>
- K. Niu, Z. Guo, X. Peng, and S. Pei, “P-ResUnet: Segmentation of brain tissue with Purified Residual Unet,” *Computers in Biology and Medicine*, vol. 151, p. 106294, Dec. 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482522010022>
- S. Li, M. Dong, G. Du, and X. Mu, “Attention dense-u-net for automatic breast mass segmentation in digital mammogram,” *Ieee Access*, vol. 7, pp. 59 037–59 047, 2019.
- T. Shen, C. Gou, J. Wang, and F.-Y. Wang, “Simultaneous segmentation and classification of mass region from mammograms using a mixed-supervision guided deep model,” *IEEE Signal Processing Letters*, vol. 27, pp. 196–200, 2019.
- M. Z. Alom, C. Yakopcic, T. M. Taha, and V. K. Asari, “Nuclei segmentation with recurrent residual convolutional neural networks based u-net (r2u-net),” in *NAECON 2018-IEEE National Aerospace and Electronics Conference*. IEEE, 2018, pp. 228–233.
- A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, and A. S. Elmaghraby, “Connected-UNets: a deep learning architecture for breast mass segmentation,” *npj Breast Cancer*, vol. 7, no. 1, pp. 1–12, Dec. 2021, number: 1 Publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s41523-021-00358-x>
- W. Yue, H. Zhang, J. Zhou, G. Li, Z. Tang, Z. Sun, J. Cai, N. Tian, S. Gao, J. Dong, Y. Liu, X. Bai, and F. Sheng, “Deep learning-based automatic segmentation for size and volumetric measurement of breast cancer on magnetic resonance imaging,” *Frontiers in Oncology*, vol. 12, 2022. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fonc.2022.984626>

- M. S. Hossain, “Microcalcification segmentation using modified u-net segmentation network from mammogram images,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 2, pp. 86–94, 2022.
- Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.
- T. Alam, W.-C. Shia, F.-R. Hsu, and T. Hassan, “Improving breast cancer detection and diagnosis through semantic segmentation using the unet3+ deep learning framework,” *Biomedicines*, vol. 11, no. 6, p. 1536, 2023.
- I. Kiran, B. Raza, A. Ijaz, and M. A. Khan, “Denseres-unet: Segmentation of overlapped/clustered nuclei from multi organ histopathology images,” *Computers in Biology and Medicine*, vol. 143, p. 105267, 2022.
- D. Song, C. Xu, X. Jia, Y. Chen, C. Xu, and Y. Wang, “Efficient residual dense block search for image super-resolution,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 12 007–12 014, Apr. 2020. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/6877>
- F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, “ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, Apr. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271620300149>
- D. Müller, I. Soto-Rey, and F. Kramer, “Towards a guideline for evaluation metrics in medical image segmentation,” *BMC Research Notes*, vol. 15, no. 1, p. 210, 2022.
- A. F. Osman and N. M. Tamam, “Contrast-enhanced mri synthesis using dense-dilated residual convolutions based 3d network toward elimination of gadolinium in neuro-oncology,” *Journal of Applied Clinical Medical Physics*, vol. 24, no. 12, p. e14120, 2023.
- E. S. Kim and K.-S. Lee, “Artificial intelligence in colonoscopy: from detection to diagnosis,” *The Korean Journal of Internal Medicine*, 2024.

- T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, “Solving the multiple instance problem with axis-parallel rectangles,” *Artificial intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997.
- F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.
- E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, “Color transfer between images,” *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.
- J.-W. Lin, “Artificial neural network related to biological neuron network: a review,” *Advanced Studies in Medical Sciences*, vol. 5, no. 1, pp. 55–62, 2017.
- G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.
- G. V. Vlăsceanu, N. Tarbă, M. L. Voncilă, and C. A. Boiangiu, “Selecting the right metric: A detailed study on image segmentation evaluation,” *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, vol. 15, no. 4, pp. 295–318, 2024.
- Anon., “Senate regulation 25: Physical format, binding and retention of theses,” **url**: [www.city.ac.uk/about/city-information/governance/constitution/senate-regulations](http://www.city.ac.uk/about/city-information/governance/constitution/senate-regulations), 2015.
- L. Lamport, *Latex, A document preparation system*, 2nd ed. Addison-Wesley, Reading, MA, 1994.
- P. van Oostrum, “Phancyhdr package documentation,” **url**: <http://texdoc.net/texmf-dist/doc/latex/fancyhdr/fancyhdr.pdf>, 2004.
- M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, “Histopathological image analysis: A review,” *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.
- J. E. P. Reyes, “Cáncer: trascendencia de un término,” *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, vol. 41, no. 9, Sep. 2016, number: 9. [Online]. Available: <http://revzoilomarinellosld.cu/index.php/zmv/article/view/852>

- A. B. Pardee, G. S. Stein, and E. A. Bronstein, *What Goes Wrong in Cancer*. John Wiley & Sons, Ltd, 2008, ch. 1, pp. 1–19. [Online]. Available: <https://0-onlinelibrary-wiley-com.wam.city.ac.uk/doi/abs/10.1002/9780470374252.ch1>
- A. Mandloi, U. Daripa, M. Sharma, and M. Bhattacharya, “An automatic cell nuclei segmentation based on deep learning strategies,” in *2019 IEEE Conference on Information and Communication Technology*, 2019, pp. 1–6.
- Y. Hu and Y. Xia, “3D Deep Neural Network-Based Brain Tumor Segmentation Using Multimodality Magnetic Resonance Sequences,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, B. Menze, and M. Reyes, Eds. Cham: Springer International Publishing, 2018, pp. 423–434.
- M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, “Brain tumor segmentation with Deep Neural Networks,” *Medical Image Analysis*, vol. 35, pp. 18–31, Jan. 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841516300330>
- S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain tumor segmentation using convolutional neural networks in mri images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
- K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, and B. Glocker, “Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, B. Menze, and M. Reyes, Eds. Cham: Springer International Publishing, 2018, pp. 450–462.
- G. Kaur and H. Kaur, “An automated method of segmentation for tumor detection by neutrosophic sets and morphological operations using mr images,” in *2016 Conference on Emerging Devices and Smart Systems (ICEDSS)*, 2016, pp. 163–168.
- E. Krivov, M. Pisov, and M. Belyaev, “MRI Augmentation via Elastic Registration for Brain Lesions Segmentation,” in *Brainlesion: Glioma, Multiple Sclerosis,*

- Stroke and Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, B. Menze, and M. Reyes, Eds. Cham: Springer International Publishing, 2018, pp. 369–380.
- F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, “Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, B. Menze, and M. Reyes, Eds. Cham: Springer International Publishing, 2018, pp. 287–297.
- A. Kermi, I. Mahmoudi, and M. T. Khadir, “Deep Convolutional Neural Networks Using U-Net for Automatic Brain Tumor Segmentation in Multimodal MRI Volumes,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. van Walsum, Eds. Cham: Springer International Publishing, 2019, pp. 37–48.
- J. Stawiaski, “A Pretrained DenseNet Encoder for Brain Tumor Segmentation,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. van Walsum, Eds. Cham: Springer International Publishing, 2019, pp. 105–115.
- X. Feng, N. J. Tustison, S. H. Patel, and C. H. Meyer, “Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features,” *Frontiers in Computational Neuroscience*, vol. 14, 2020. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fncom.2020.00025>
- F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, “No new-net,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. van Walsum, Eds. Cham: Springer International Publishing, 2019, pp. 234–244.
- A. F. Kazerooni, N. Khalili, X. Liu, D. Haldar, Z. Jiang, S. M. Anwar, J. Albrecht, M. Adewole, U. Anazodo, H. Anderson, S. Bagheri, U. Baid, T. Bergquist, A. J. Borja, E. Calabrese, V. Chung, G.-M. Conte, F. Dako, J. Eddy, I. Ezhov, A. Familiar, K. Farahani, S. Haldar, J. E. Iglesias, A. Janas, E. Johansen, B. V. Jones, F. Kofler, D. LaBella, H. A. Lai, K. Van Leemput, H. B. Li, N. Maleki, A. S. McAllister, Z. Meier, B. Menze, A. W. Moawad, K. K. Nandolia, J. Pavaine, M. Piraud, T. Poussaint, S. P.

- Prabhu, Z. Reitman, A. Rodriguez, J. D. Rudie, M. Sanchez-Montano, I. S. Shaikh, L. M. Shah, N. Sheth, R. T. Shinohara, W. Tu, K. Viswanathan, C. Wang, J. B. Ware, B. Wiestler, W. Wiggins, A. Zapaishchykova, M. Aboian, M. Bornhorst, P. de Blank, M. Deutsch, M. Fouladi, L. Hoffman, B. Kann, M. Lazow, L. Mikael, A. Nabavizadeh, R. Packer, A. Resnick, B. Rood, A. Vossough, S. Bakas, and M. G. Linguraru, “The Brain Tumor Segmentation (BraTS) Challenge 2023: Focus on Pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs),” *ArXiv*, p. arXiv:2305.17033v6, Mar. 2024. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10246083/>
- “A Fast and Refined Cancer Regions Segmentation Framework in Whole-slide Breast Pathological Images | Scientific Reports.” [Online]. Available: <https://www.nature.com/articles/s41598-018-37492-9>
- Y.-G. Kim, S. Kim, C. E. Cho, I. H. Song, H. J. Lee, S. Ahn, S. Y. Park, G. Gong, and N. Kim, “Effectiveness of transfer learning for enhancing tumor classification with a convolutional neural network on frozen sections,” *Scientific Reports*, vol. 10, no. 1, p. 21899, Dec. 2020, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s41598-020-78129-0>
- B. Ehteshami Bejnordi, M. Balkenhol, G. Litjens, R. Holland, P. Bult, N. Karssemeijer, and J. A. W. M. van der Laak, “Automated detection of dcis in whole-slide h&e stained breast histopathology images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 9, pp. 2141–2150, 2016.
- “Automated and Manual Quantification of Tumour Cellularity in Digital Slides for Tumour Burden Assessment | Scientific Reports.” [Online]. Available: <https://www.nature.com/articles/s41598-019-50568-4>
- M.-A. Khalil, Y.-C. Lee, H.-C. Lien, Y.-M. Jeng, and C.-W. Wang, “Fast Segmentation of Metastatic Foci in H&E Whole-Slide Images for Breast Cancer Diagnosis,” *Diagnostics*, vol. 12, no. 4, p. 990, Apr. 2022, number: 4 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/2075-4418/12/4/990>
- H. Jiang, Z. Diao, and Y.-D. Yao, “Deep learning techniques for tumor segmentation: a review,” *The Journal of Supercomputing*, vol. 78, no. 2, pp. 1807–1851, Feb. 2022. [Online]. Available: <https://doi.org/10.1007/s11227-021-03901-6>

- D. Arthur and S. Vassilvitskii, “K-means++: the advantages of careful seeding,” in *In Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007.
- S. Di Cataldo, E. Ficarra, and E. Macii, “Computer-aided techniques for chromogenic immunohistochemistry: status and directions,” *Computers in biology and medicine*, vol. 42, no. 10, pp. 1012–1025, 2012.
- X. Yang, H. Li, and X. Zhou, “Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy,” *IEEE Transactions on Circuits and Systems I*, vol. 53, no. 11, p. 2405–2414, Nov 2006.
- “Cognitron: A self-organizing multilayered neural network | Biological Cybernetics.” [Online]. Available: <https://link.springer.com/article/10.1007/BF00342633>
- E. H. Adelson, P. J. Burt, C. H. Anderson, J. M. Ogden, and J. R. Bergen, “Pyramid methods in image processing,” *RCA engineer*, vol. 29, no. 6, pp. 33–41, 1984.
- P. Burt and E. Adelson, “The laplacian pyramid as a compact image code,” *IEEE Transactions on Communications*, vol. 31, no. 4, p. 532–540, Apr. 1983. [Online]. Available: <https://ieeexplore.ieee.org/document/1095851>
- S. Guan, N. Kamona, and M. Loew, “Segmentation of Thermal Breast Images Using Convolutional and Deconvolutional Neural Networks,” in *2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, Oct. 2018, pp. 1–7, iSSN: 2332-5615. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8707379>
- X. Xiao, Y. Qiang, J. Zhao, X. Yang, and X. Yang, “Segmentation of Liver Lesions Without Contrast Agents With Radiomics-Guided Densely UNet-Nested GAN,” *IEEE Access*, vol. 9, pp. 2864–2878, 2021, conference Name: IEEE Access. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9309035>
- G. Lian, “Rotation invariant color texture classification using multiple sub-dlbp,” *Journal of Visual Communication and Image Representation*, vol. 31, pp. 1–13, 2015.
- L.-C. Chen, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017.

- G. Challenge, “Breast cancer semantic segmentation challenge (bcss),” **url:** <https://bcsegmentation.grand-challenge.org/breast-cancer-semantic-segmentation-bcss>, 2022.
- Ananda, “Wrist fractures analysis as observed with x-ray imagin,” Ph.D. dissertation, City University of London, 2023.
- Cefa, “Volumetric analysis of hela cancer cells imaged with serial block face scanning electron microscopy,” Ph.D. dissertation, City University of London, 2020.
- B. Dev and L. D. Joseph, *Holistic Approach to Breast Disease*. Springer Nature, Aug. 2023, google-Books-ID: d27REAAAQBAJ.
- H.-B. Pan, “The Role of Breast Ultrasound in Early Cancer Detection,” *Journal of Medical Ultrasound*, vol. 24, no. 4, pp. 138–141, Dec. 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0929644116300686>
- M. Van Goethem, W. Tjalma, K. Schelfout, I. Verslegers, I. Biltjes, and P. Parizel, “Magnetic resonance imaging in breast cancer,” *European Journal of Surgical Oncology (EJSO)*, vol. 32, no. 9, pp. 901–910, 2006.
- A. C. Wolff, M. E. H. Hammond, D. G. Hicks, M. Dowsett, L. M. McShane, K. H. Allison, D. C. Allred, J. M. Bartlett, M. Bilous, P. Fitzgibbons *et al.*, “Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of american pathologists clinical practice guideline update,” *Journal of clinical oncology*, vol. 31, no. 31, pp. 3997–4013, 2013.
- P. Dey *et al.*, *Basic and advanced laboratory techniques in histopathology and cytology*. Springer, 2018.
- R. H. Grobstein, *The Breast Cancer Book: What You Need to Know to Make Informed Decisions*. Yale University Press, 2005. [Online]. Available: <http://www.jstor.org/stable/j.ctt1nq6sz>
- C. X. Ma and M. J. Ellis, “The cancer genome atlas: clinical applications for breast cancer,” *Oncology*, vol. 27, no. 12, pp. 1263–1263, 2013.

G. L. Menezes, F. M. Knuttel, B. L. Stehouwer, R. M. Pijnappel, and M. A. van den Bosch, “Magnetic resonance imaging in breast cancer: a literature review and future perspectives,” *World journal of clinical oncology*, vol. 5, no. 2, p. 61, 2014.

M. Yousif, P. J. van Diest, A. Laurinavicius, D. Rimm, J. van der Laak, A. Madabhushi, S. Schnitt, and L. Pantanowitz, “Artificial intelligence applied to breast pathology,” *Virchows Archiv*, pp. 1–19, 2022.