Machine Learning and Digital Histopathology Analysis for Tissue Characterization and Treatment Response Prediction in Breast Cancer

Khadijeh (Shirin) Saednia

A Dissertation Submitted to
the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy in
Electrical Engineering and Computer Science

Graduate Program in
Electrical Engineering and Computer Science
York University
Toronto, Ontario

December 2022

Abstract

Breast cancer is the most common type of diagnosed cancer and the leading cause of cancer-related death in women. Early diagnosis and prognosis in breast cancer patients can permit more therapeutic options for them and possibly improve their survival and quality of life. The gold standard approach for breast cancer diagnosis and characterization is histopathology assessment on biopsy specimens, that is time and resource demanding. In this dissertation project, state-of-the-art machine learning (ML) methods have been developed and investigated for breast tissue characterization, nuclei segmentation, and chemotherapy response prediction in breast cancer patients using pre-treatment digitized histopathology images. First, a novel multi-scale attention-guided deep learning model is introduced to characterize breast tissue on digital pathology images according to four histological types. Evaluation results on the test set show the effectiveness of the proposed approach in accurate histopathology image classification with an accuracy of 97.5%. In the next step, a cascaded deep-learning-based model is proposed to delineate tumor nuclei in digital pathology images accurately, that is an essential step for extracting hand-crafted quantitative features for analysis with conventional ML models. The proposed model could achieve an F1 score of 0.83 on an independent test set. At the end, two novel ML frameworks are introduced and investigated for chemotherapy response prediction. In the first approach, a digital histopathology image analysis framework has been developed to extract various subsets of quantitative features, from the segmented digitized slides for conventional ML model development. Several ML experiments have been conducted with different feature sets to develop prediction models of therapy response using a gradient boosting machine with decision trees. The proposed model with the optimal feature set could achieve an accuracy of 84%, sensitivity of 85% and specificity of 82% on an independent test set. In the second

approach, a hierarchical self-attention-guided deep learning framework is introduced to predict breast cancer response to chemotherapy using digital histopathology images of pre-treatment tumor biopsies. The whole slide images (WSIs) are processed automatically through the proposed hierarchical framework consisting of patch-level and tumor-level processing modules followed by a patient-level response prediction component. A combination of convolutional and transformer modules are utilized in each processing level. The proposed framework could outperform the conventional ML models with a test accuracy, sensitivity, and specificity of 86%, 87%, 83%, respectively. The proposed methods and the reported results in this dissertation are steps forward toward streamlining the histopathology workflow and implementing response-guided precision oncology for breast cancer patients.

Originality Statement

This dissertation has been written by Khadijeh (Shirin) Saednia under the supervision of Dr. Ali Sadeghi-Naini and Dr. William T. Tran. Parts of the material presented in this dissertation have been published or submitted for publication in peer-reviewed journals or conference proceedings, as listed below. The research presented in each publication has been performed by the principal author, supervised/supported by the senior authors, and in collaboration with other co-authors who participated in data acquisition/preprocessing.

- The material in Chapter 2 has been presented in:
    - K. Saednia, W. T. Tran, A. Sadeghi-Naini. *Automatic characterization of breast lesions using multi-scale attention-guided deep learning of digital histology images*. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. In Press, 2022. DOI: 10.1080/21681163.2022.2058415.

- The material in Chapter 3 has been presented in:
    - K. Saednia, W. T. Tran, A. Sadeghi-Naini. *A cascaded deep learning framework for segmentation of nuclei in digital histology images*. 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Glasgow, Scotland, UK. pp. 4764-4767, 2022.

- The material in Chapter 4 has been presented in:
    - K. Saednia, A. Lagree, M. A. Alera, L. Fleshner, A. Shiner, E. Law, B. Law, D. W. Dodington, F. I. Lu, W. T. Tran, A. Sadeghi-Naini. *Quantitative digital histopathology and machine learning to predict pathological complete response to chemotherapy in breast cancer patients using pre-treatment tumor biopsies*. Scientific Reports. 12: 9690, 2022.

- The material in Chapter 5 has been presented in:
    - K. Saednia, W. T. Tran, A. Sadeghi-Naini. *A hierarchical self-attention-guided deep learning framework to predict breast cancer response to chemotherapy using pre-treatment tumor biopsies*. Revision Submitted, Medical Physics. 2022.

Dedication

To the victims of flight #PS752 and all courageous Iranians who have been fighting for liberty and sacrificing their lives for a better world.

Acknowledgments

During my graduate studies, I had the opportunity to work with several people, and I would like to take this chance to thank all those who helped make this achievement happen.

I would like to express my gratitude to my supervisors, Dr. Ali Sadeghi-Naini and Dr. William T. Tran, for their forward and continuous encouragement during my Ph.D. studies. I would like to especially thank Dr. Sadeghi-Naini for his kind guidance, precise advice, constructive comments, and effortless contributions throughout these years. I was so lucky to have the opportunity of being co-supervised by Dr. Tran, one of the best teachers and mentors I have ever known, and I would like to thank him for being extremely caring and supportive. The best part about being supervised by my two supervisors was that they are both great, honest, hardworking people and working with them was a source of inspiration for me.

I also appreciate the time and efforts spent by members of my supervisory committee, Dr. Aijun An and Terry Sachlos, for serving and providing beneficial suggestions.

My colleagues at York University and Sunnybrook Research Institute made my Ph.D. studies a wonderful experience for me with their frequent constructive comments and unlimited support.

Lastly, words cannot express my gratitude to my kind family, especially my mom. I have the most loving mom, and nothing was harder than living thousands of kilometers far from her. My highest grace also belongs to my spouse, Farhad, whose belief in me has kept my spirits and motivation high during this journey, with all ups and downs. I'm also blessed with sweet and kind friends, whom I've never seen anything but support and love.

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

AG ……. Attention Gate
AI ……..... Artificial Intelligence
AJCC ……..…. American Joint committee on Cancer
AJI ……..……. Aggregated Jaccard Index
AUC ……..... Area Under Curve
BACH ……..…… Breast Cancer Histology
BE ……..…….. Barrett's esophagus
BI-RADS ……... Breast Imaging Reporting And Data System
CAM ……..……. Class Activation Maps
CBAM ……..…….. Convolutional Block Attention Module
CNB ……..…….. Core Needle Biopsy
CNN ……..……..… Convolutional Neural Network
CNS ……..…….. Central Nervous System
CT ……..…..…… Computed Tomography
DCIS ……..……..…… Ductal Carcinoma in Situ
DCNN ……..…….……. Deep Convolutional Neural Network
DL ……..……..……. Deep Learning
ER ……..……..……... Estrogen
FFPE ……..... Formalin-fixed paraffin embedded
FN ……...……. False Negative
FP …….. False Positive
GAP ……..… Global Average Pooling
GBM ……..…. Gradient Boosting Machine
GLCM ……..…. Gray-Level Co-occurrence Matrix
GLRLM …… Gray-Level Run Length Matrix
HER2 ……..……..…. Human Epidermal Growth Factor2 Receptor
H&E ……...……..……. Hematoxylin and Eosin
IBSN …….....… International Breast Cancer Screening Network
IDC ……..…….. Invasive Ductal Carcinoma
IHC ……...……..…. Immunohistochemistry
ILC ……..……..…….. Invasive Lobular Carcinoma
IRB ……..…… Institutional ethics Review Board
ISBI ……...……... International Symposium on Biomedical Imaging
LABC ……..…… Locally Advanced Breast Cancer
LCIS ……..…… Lobular Carcinoma in Situ

LDA …… Linear Discriminant Analysis
LSTM …… Long short-term memory
LUAD …… Lung adenocarcinoma
LUSC …… Lung squamous cell carcinoma
MCC …… Matthews Correlation Coefficient
ML …… Machine Learning
MLP …… Multi-layer Perceptron
MoNuSeg …… Multi-organ Nuclei Segmentation
MRF …… Markov Random Field Model
MRI …… Magnetic Resonance Imaging
mRMR …… maximum Relevance Minimum Redundancy
NAC …… Neoadjuvant Chemotherapy
NAT …… Neoadjuvant Therapy
NGTDM …… Neighborhood Gray Tone Difference Matrix
NLP …… Natural Language Processing
NSCLC …… Non-small Cell Lung Cancer
NST …… No Special Type invasive breast cancer
OBSP …… Ontario Breast Screening Program
OS …… Overall Survival
pCR …… Pathologic Complete Response
PET …… Positron Emission Tomography
PR …… Progesterone
QDA …… Quadratic Discriminant Analysis
Quantimb …… Quantitative imaging and Biomarkers Laboratory
RCBI …… Residual Cancer Burden Index
RFS …… Recurrence-Free Survival
ROC …… Receiver Operating Characteristic
ROI …… Region of Interest
SGD …… Stochastic Gradient Descent
SVM …… Support Vector Machine
TCIA …… The Cancer Imaging Archive
TCPS …… Canadian Tri-Council Policy Statement
TDLU …… Terminal Duct Lobular Unit
TIFF …… Tag Image File Format
TMA …… Tissue microarray
TN …… True Negative
TP …… True Positive
US …… Ultrasound
VAB …… Vacuum-assisted biopsy
ViT …… Vision Transformer

WHO …….…..... World Health Organization
WSI …….…….…….. Whole slide image

Chapter 1

Introduction

1.1 Background and Motivation

1.1.1 Breast Cancer

Cancer is one of the major causes of mortality worldwide and is responsible for over 28% of all deaths in Canada (Figure 1.1) [1]. Breast cancer has the greatest incidence among all cancer types in women and account for 24.5% of newly diagnosed malignancies in 2020, according to the world health organization (WHO) report [2], [3]. About 685,000 breast cancer patients die, and approximately 2.3 million new cases are diagnosed each year in the world [3], [4]. Among these patients, mortality rate is estimated at 30%, mainly due to breast cancer metastasis. Breast cancer predominantly affects women; however, approximately 1% of cases are found in men [5], [6]. An effective treatment plan can increase the survival rate; however, offering the best treatment option at the proper time is still challenging for clinicians. The complications and monetary costs of treatment and also the high prevalence of this disease, make it one of the most challenging health problems worldwide [7].

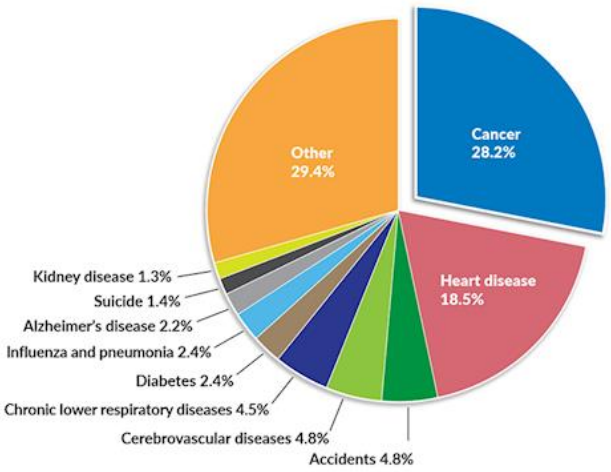


Figure 1.1 - Proportion of deaths due to cancer compared to other causes in Canada [1].

1.1.2 Breast Cancer Types and Stages

When the cells in the body become damaged or aged, new cells take their place. Normally, the cells growth and division are strictly regulated by hundreds of genes in body, and require a balance between the activity of those genes that promote cell proliferation and those that suppress it. On the other hand, a cancer cell is a natural but immature cell in the body that has developed mutations in its DNA. These mutations lead to cancer cells repeatedly and uncontrollably dividing, which creates daughter cells that are both mutated and unspecialized and continue to divide (Figure 1.2). The tumors can be benign (non-cancerous) or malignant (cancerous) [8]. Cancer development may start in almost any part of the body, including different organ tissues, including visceral, skeletal, and central nervous system (CNS) [9]. In breast cancer, cells from any part of the breast (described below) may grow uncontrollably, and the histological type of cancer is defined based on the part in which the cancer cells have initiated, cancer cell type and the growth pattern of them. Breast tissue consists of three main parts: lobules, ducts, and connective tissue. More than 80% of breast cancer tumors originate from transformations of the epithelial cells that line the ducts or lobules of the breast [10]. Carcinomas refer to the cancers that originated from the epithelial tissue and can be categorized

into two main types: non-invasive carcinoma, in which the abnormal cells would not spread to other parts of the breast, and invasive carcinoma, which has a high potential to spread from its origin to other parts of the body [11], [12]. The most common histological types of non-invasive carcinoma include ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS). DCIS is also named intraductal cancer as the cancer cells grow and remain in the milk ducts. DCIS is considered as early-stage breast cancer. LCIS refers to an area of abnormal carcinoma cells remaining within the lobules. This malignant type is not often considered as breast cancer, but it indicates that a person has a high risk of developing invasive breast cancer [13]. Invasive carcinoma can be classified to several histological types including invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) [12], [14]. IDC is the most common type of invasive breast cancer in both women and men, also called no special type of invasive breast cancer (NST). ILC originates in the lobules and have the potential to spread into the surrounding breast tissue. This type of cancer accounts for about 15% of all breast cancer and is considered the second most common type of breast cancer [14].

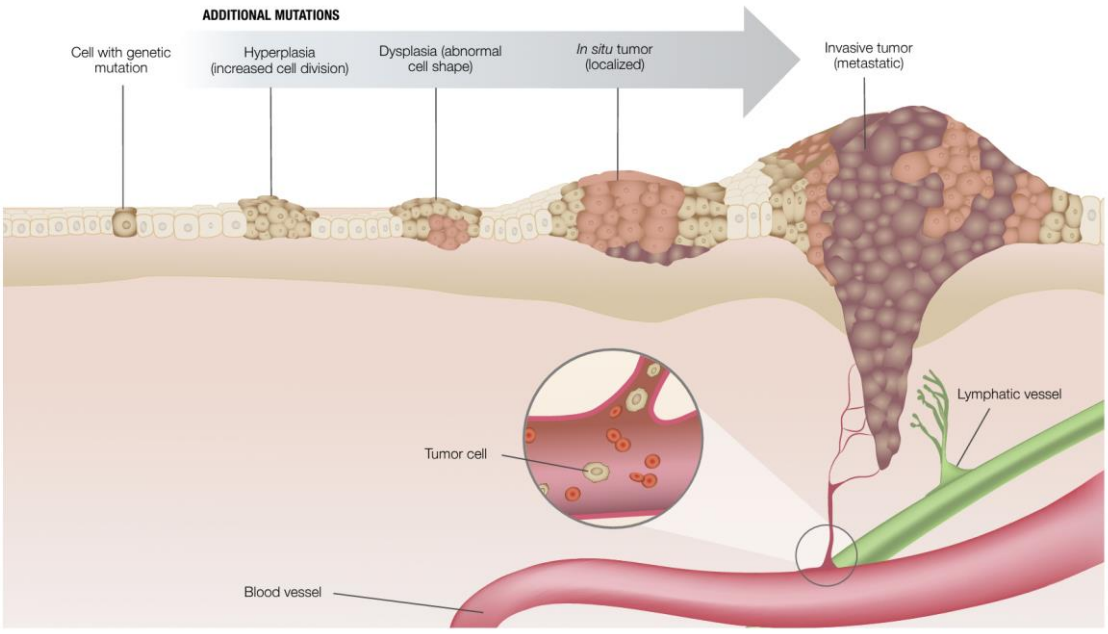


Figure 1.2 – Tumors develop stages, normal division to invasive tumor [9].

Receptors are special proteins that can be found within and on the surface of cells in the body, including the breast cells. The presence of estrogen (ER) and progesterone (PR) hormone receptors and human epidermal growth factors receptor 2 (HER2) on cancer cells categorizes carcinoma into four identifiable molecular subtypes and can affect prognosis and selection of pharmaceutical treatments [15]. These subtypes include luminal A (ER+/PR+/HER2-), luminal B (ER+/HER2-/[high Ki-67 level or PR-]), HER2-enriched (ER-/PR-/HER2+), and triple negative or basal-like (ER-/PR-/HER2-) breast cancer. Breast cancers with detectable hormone receptors, called hormone receptor-positive, use the estrogen and/or progesterone hormones to grow and spread [16]. The receptor status could be detected using different biological markers [15]. Based on the study by the American Cancer Society, about two-thirds of all breast cancers are hormone-receptor positive.

Clinicians use the staging guidelines to describe the characteristic of breast cancer in two principal groups including anatomic, which is based on extent of cancer, and prognostic, which includes anatomic staging information plus tumor grade and the receptor status (HER2, ER, and PR). The improvement of the staging process in breast cancer has led to increases in local control rates [17]. The most common criteria to identify the breast cancer anatomic stage is the TNM system introduced by the American joint committee on cancer (AJCC) [18]. The TNM system defines the breast cancer stage based on three factors: the primary size and the location of the tumor (T), the status of the regional lymph nodes (N), and metastasis (M) status that shows whether cancer has spread to a different part of the body. The assessment results for these factors will be used together to determine the cancer stage for each patient [19]. Stage 0 refers to the non-invasive DCIS, and stages one to four are used for the lowest to the highest stage of invasive carcinoma, respectively.

Locally Advanced Breast Cancer (LABC) is a class of breast cancer based on clinical presentation, with tumors that are frequently larger than 5 cm in size or spread to multiple local lymph nodes,

possibly with chest wall, nipple, and/or skin involvement [20], [21]. Because of the high risk for early metastatic spread and disease progression, patients with LABC usually have poor long-term survival rates compared to early-stage breast cancer [22].

1.1.3 Origin of Breast Cancer and Risk Factors

Most breast carcinomas appear to originate in normal breast glandular epithelium, which covers the lumen of terminal duct lobular units (TDLUs). The process of converting the normal TDLU to invasive carcinoma takes a long evolution period in which epithelial cells undergo different stages of progression to cancer. When the normal epithelium starts to transform into malignant, it goes through the first stage of extreme proliferation called hyperplasia, followed by the appearance of cells showing abnormal characteristics. At a later phase known as DCIS, the cells obtain a complete malignant phenotype with different morphological and biochemical characteristics, except for the ability to infiltrate the surrounding parenchyma through the basal membrane of the ducts or lumens. In the final stage, the carcinoma cells break down the basal membrane and turn into invasive carcinoma. The process of changing the normal TDLU epithelium to cancer cell has been characterized based on the coincidence of the cancer cells and the surrounding DCIS in the biopsies of patients at the cell-biology level [23], [24]. However, the biological evolution of carcinoma is more complex than described [25]. It is not feasible to determine the exact percentage of cases in which the DCIS may progress to invasive cancer, but it has been estimated to be less than 60% [26].

Although the exact cause of breast cancer is still unclear, there are several risk factors that have been identified in different studies [27]. At 20 years (age), the chance of developing breast cancer is about 0.06%, while this number can go up to 3.84% by the age of 70 years. Women with mutations in certain genes (BRCA1 and BRCA2) have a higher chance of developing breast cancer [28]. Also, previous development of some types of benign tumors can increase the risk of developing breast

cancer [29]. Another risk factor is estrogen exposure which shows a correlation to increasing the chance of developing breast cancer [17]. Women with dense breasts have a higher chance of breast cancer development [30]. Also, elevated body habitus and obesity after menopause can increase the likelihood of developing breast cancer, possibly because of increasing estrogen levels [31]–[33].

1.1.4 Breast Cancer Screening and Diagnosis

According to the international breast cancer screening network (IBSN) recommendation, breast cancer screening should be done every two years for women between the ages of 50 and 74. The screening is performed to detect signs of disease before a person has symptoms. The aim of screening tests is to diagnose cancer at an early stage when the tumors are small, less likely to have spread and more likely to be cured successfully [34], [35]. The most common screening tests for breast cancer include mammography, ultrasound (US) and magnetic resonance imaging (MRI). Mammography is x-ray radiography of the breast that uses low doses of radiation for 2D projection imaging, and potentially detects both cancerous and non-cancerous tumors in the breast. The procedure involves compressing each breast with two plates, and acquiring two to four projection images from different angles. X-rays are passed through the breast, and because of higher x-ray attenuation in dense tissues such as lesions, these tissues look brighter on the final image. Mammography images can also show deposited calcium within the breast, called microcalcifications. The features of the microcalcifications around lesions, such as distribution, density and size, can be used to classify the lesions as malignant or benign [36], [37]. Detection of tumors within dense breast tissue using mammography is challenging. Dense breast tissue appears white or bright gray on a mammogram, similar to breast abnormalities such as calcifications and tumors. This can make a mammogram harder to read, and may make it more difficult to detect breast cancer in women with dense breasts [38].

US is used as a standard method for breast imaging following the detection of suspicious lesions, *e.g.*, via clinical breast examination or mammography. This modality is not usually used alone as a diagnostic screening test; rather, it would be used as an adjunct to mammography tests [39]. It is a fast, inexpensive, and portable solution to follow up on suspicious lesions. US imaging is performed by sending high-frequency sound pulses through the breast, receiving the echoes, and processing the received echoes to generate an image. If any abnormal tissue is seen on mammography, the US is usually used to determine whether the tissue is a solid mass (such as a benign fibroadenoma or malignant tumor) or a fluid-filled lump (such as a benign cyst). The US cannot detect calcifications or ascertain the malignancy of solid lumps [40].

MRI is currently suggested for identifying occult lesions in the breast, or imaging high risk or locally advanced disease. It may also be used for screening in woman at high risk of breast cancer along with mammography. As a non-invasive imaging modality, MRI uses powerful magnets to create a strong magnetic field that compels protons in the body to align with that field and, by measuring the tissue magnetization properties, generate a three-dimensional detailed image of the breast [41]. In breast MRI, a gadolinium-based contrast agent is often used through intravenous injection to increase sensitivity in detecting and characterizing tumors. One superiority of MRI in comparison to mammography is that it can generate 3D volumetric images, so the malignancies within the dense tissue can be detected and visualized with higher sensitivity [37]. However, MRI is relatively expensive, not always available, and has relatively low specificity, so it is not routinely used for breast cancer screening [41].

The breast imaging reporting and data system (BI-RADS) is the standard system for reporting breast cancer imaging findings that was initially defined for describing mammography and later adapted for US and MRI image modalities. The first component of the reporting system contains the

patient history, and the next part includes the breast composition findings on imaging with four categories of entirely fatty, scattered areas of fibro glandular density, heterogeneously dense and extremely dense tissues. The BI-RADS score is in the range of 0 to 6 and estimates the risk of breast cancer based on imaging findings higher scores (≥ 4) show a higher probability of existing malignancies (score 6 is reserved for biopsy-proven malignancy), and smaller values (< 4) are associated with existing benign tumors or no masses on imaging (BIRADS $= 1$). A score of 0 indicates an incomplete imaging test.

While clinical assessments based on breast examinations and standard imaging are required for breast cancer screening, a biopsy followed by histopathology analysis on breast tissue specimens is crucial for a definitive diagnosis of breast cancer [42]. Different types of breast biopsy procedures include fine-needle aspiration biopsy (a quick way to distinguish between a fluid-filled cyst and a solid mass), open (surgical) biopsy and core needle biopsy. Core needle biopsy (CNB) is an invasive test that is performed using a hollow needle to take out pieces of breast tissue from the area of concern which is detected via imaging. The tissue samples come out as long narrow pieces, and their micro-structure remains intact. The consecutive decision regarding the follow-up and possible treatment plans is mainly based on the histopathology analysis of biopsy specimens; therefore, diagnostic accuracy is essential [43]. The pathologist uses biopsy specimens to determine the cellular, biochemical, and physiological characteristics of the tissue specimen. Histopathological assessment using microscopy is the gold standard approach to confirm the malignancy of suspicious lesions detected using the medical imaging modalities such as mammography, US and MRI [44].

1.1.5 Breast Cancer Histopathology Images

The breast tissue biopsy or surgical specimens are analyzed in pathology laboratories to diagnose and characterize malignancies or evaluate neoadjuvant treatment responses (described in Section

1.1.7). The first step of tissue preparation is formalin fixation and embedding in paraffin. The paraffin blocks are cut to the sections of $3-5\ \mu m$ thickness and mounted on the glass slides. But the regions of interest, such as nuclei and cytoplasm are not easily visible on the slides. Therefore, staining methods are routinely used to dye different tissue and cell structures. The most popular staining technique uses hematoxylin and eosin (H&E) to highlight the structures of interest on histology slides. Hematoxylin binds to DNA and dyes the nuclei purple, while eosin binds to proteins and dyes cytoplasm and extracellular matrix pink [45], [46].

Another staining technique, immunohistochemistry (IHC), entails chromogens attached to antigens in the tissues with the aid of antibodies. Immunohistochemistry is used to detect the presence of estrogen (ER), progesterone (PR), and human epidermal growth factor2 receptors (HER2). Tissue microarrays (TMAs) constructed by punching core biopsies from the region of interest are used by researchers to gain high-throughput analysis, but this method is not a standard routine in the clinic yet. The TMAs provide the opportunity to stain tissues from different patients under the same condition, resulting in less staining variability [45].

Manual assessment of the histopathology slides is time-consuming, resource-intensive and requires substantial expertise. The advance of slide digitization systems helps pathologists to use digitized slides (images) for more accurate analysis. Most whole slide image (WSI) scanners can perform slide scanning at 20× or 40× magnification with a spatial resolution of $0.5\ \mu m/pixel$ and $0.25\ \mu m/pixel$, respectively. The WSIs are mostly compressed using JPEG2000 and can be stored in a pyramid structure for multiscale image analysis [45]. The so-called SVS files are used by most medical scanners like Aperio (Leica Biosystems Inc., IL, USA), utilizes the tiled image capabilities based on the Tag Image File Format (TIFF) format. The SVS files can be opened using tiff readers, but if the JPEG2000 compression is used, the tiff readers provide a low-resolution preview of

images. Some software such as Sedeen viewer [47] and QuPath application [48] is introduced to view and edit the tiff-based files with high resolution. Figure 1.3 shows an example of a WSI acquired from breast biopsy specimens.

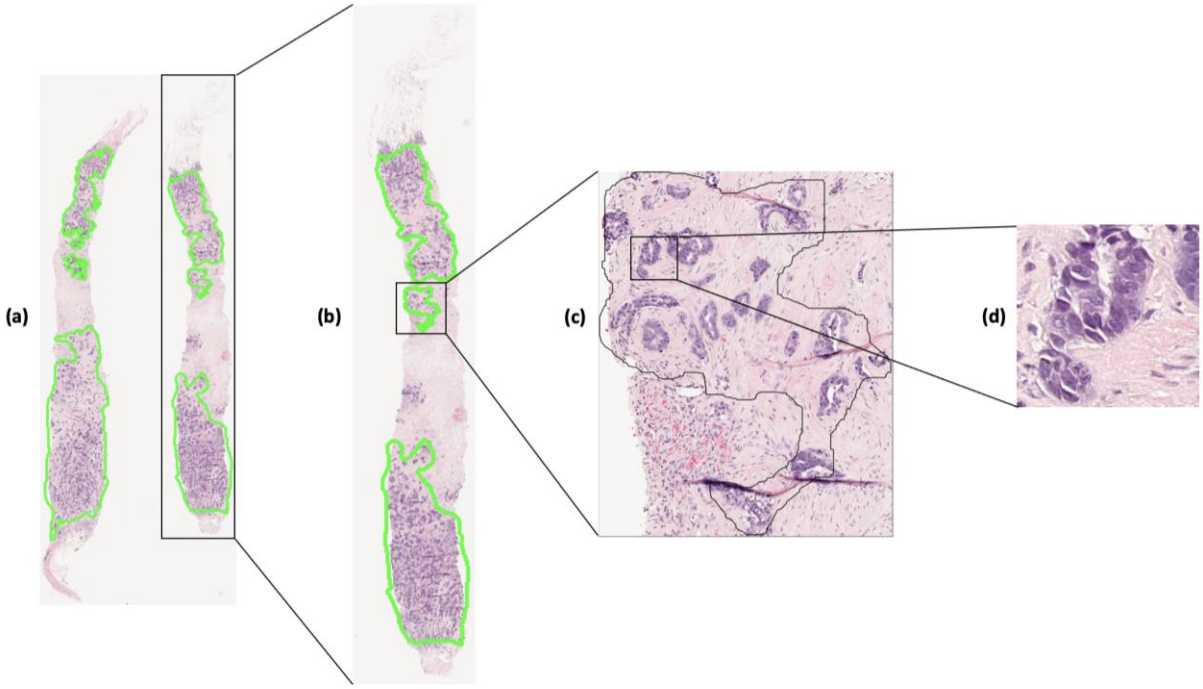


Figure 1.3 – (a) A representative WSI acquired from breast biopsy specimens (two cores) and scanned at 40× (39314×80528 pixels), (b) a biopsy core with annotated tumor region (17483×79764 pixels), (c) a tumor region (3624×4472 pixels), (d) a patch extracted from the tumor region (512×512 pixels)

1.1.6 Breast Cancer Treatment

A number of clinical factors including the patient's age, tumor size, cancer stage, receptor status, and proliferative index, are considered to prescribe an appropriate treatment plan for breast cancer patients [29]. The main treatment options for breast cancer include a combination of systemic therapies (chemotherapy, hormonal therapy, targeted therapy, immunotherapy), radiation therapy, and/or surgery [49]. Chemotherapy is a treatment that employs medications to induce cell death in, and/or block or decrease the growth of breast cancer cells. Chemotherapy is regarded as a systemic therapy since it circulates throughout the body via the bloodstream [50]. Hormonal therapies are

used to either lower the estrogen/progesterone levels or stop them from acting on the cancer cells. Tamoxifen and Fulvestrant are examples of drugs that block the estrogen receptors [51]. There are two types of surgery to remove breast cancer. In the mastectomy, the surgeon will remove the entire breast, while the breast conserving surgery (lumpectomy) refers to the surgery in which only the tumor and a small rim of normal tissue around it will be removed. During surgery, various types of lymph node biopsy, including sentinel lymph node biopsy and complete axillary lymph node dissection can be done for further diagnosis and treatment planning. Neoadjuvant therapies encompass all treatments that are administered before the main treatment that is usually surgery, to help reduce the size of a tumor or kill cancer cells that have spread, whereas adjuvant therapies are delivered after the primary treatment to destroy remaining cancer cells [52], [53].

1.1.7 Evaluation of Treatment Response and Outcome

Following breast surgery (*i.e.*, lumpectomy or mastectomy), histopathological assessment of the breast surgical specimens is the standard method for evaluating tumor response to neoadjuvant treatments. This is carried out systematically either through synoptic pathology reporting or by using commonly accepted methods such as the residual cancer burden index (RCBI) [54]. Complete eradication of cancer cells following neoadjuvant chemotherapy is defined as a pathologic complete response (pCR). Here, pathologists assess the specimen under the microscope and evaluate the presence of residual tumor cells within the breast and regional lymph nodes. There is strong evidence to suggest that a pCR is correlated to better prognostic outcomes [55]. However, even for patients with pCR, cancer may recur after initial treatment. Recurrent carcinoma may occur months or years after the treatment, and it may recur at the same location as original cancer (local recurrence) or spread to other areas of the body (distant recurrence or metastatic recurrence). On average, 7% to 11% of women with early-stage breast cancer will experience a local recurrence during the first five

years after treatment. Overall, patients with a BRCA1 or BRCA2 gene mutation or with a family history of cancer have a higher recurrence rate [56]. Various clinical and pathological factors have been shown to be correlated with different therapy outcomes in breast cancer. For example, according to [57], the number of positive nodes (nodal status) is associated with recurrence-free survival (RFS) and overall survival (OS) after chemotherapy, and the tumor size and HER2 status are associated with RFS and OS after radiotherapy.

1.1.8 Challenges in Personalized Management of Breast Cancer

As discussed in Section 1.1.4, histopathological assessment using microscopy is the gold standard approach for breast cancer diagnosis, while tumor characterization through histopathology is a necessary step for treatment planning. However, manual characterization of breast tissues on histology slides/images by pathologists is time-consuming, may associate with human errors, and lacks scalability. Developing an intelligent system for automated classification of breast histology images can streamline the histopathology assessments workflow in clinic and offer scalability by sharing the findings, growing datasets, and system improvements between different cancer institutions around the world.

As described in Section 1.1.7, there are different factors associated with treatment outcome in breast cancer, while response to neoadjuvant chemotherapy has shown highly correlated to improved survival outcomes [58]. The standard approach for evaluating the tumor response to neoadjuvant chemotherapy is based on post-treatment anatomical imaging and histopathology on surgical specimens. However, the options for treatment modification or adaptive therapy, if any, is very limited at that time. Also, it may take several months for changes in the physical characteristics of a tumor to be detectable on anatomical imaging, and in some situations, there may be no change in the tumor's size despite a pathological response to treatment (*e.g.*, the composition of the residual tumor

comprises percentages of epithelium, stroma, adipose and lymphocytic infiltrate, indicating a pathological response). Predicting the tumor response to standard treatment options accurately at early stages of diagnosis is highly beneficial for effective treatment planning and optimizing the treatment regimen for individual patients. Such predictions can help to avoid weeks to months of ineffective treatments where time is an important factor in overcoming the disease. Predicting the cancer response to up-front therapy would open opportunities for research toward personalized medicine for cancer patients. A precision cancer therapeutics paradigm can lead to improved treatment response and outcomes for patients, such as increased rates of pCR, reduced rates of disease relapse, and improved survival and quality of life.

Histopathology on pre-treatment biopsy specimens is performed for breast cancer patients as part of the standard of care. Although pathologists can characterize tumors on the histology slides/images, analyzing these slides/images at pre-treatment to predict the cancer response to treatment is outside the scope of their standard clinical practice. The advancement in histology slide digitization systems that can generate whole slide images (WSIs) at different magnifications with high resolution, open new pathways to develop and investigate high-throughput computational frameworks for analyzing pre-treatment histopathology images and explore the possibility of deriving quantitative imaging features that are associated with patterns of tumor response to treatment. Integrating such image analysis frameworks within an intelligent automated system for therapy response prediction can provide a valuable clinical decision support tool for treatment planning and optimization for breast cancer patients. A potential challenge in developing automated frameworks for histopathology image analysis is the very large size of WSIs that, in contrast to typical input image size for image analysis computational models, can be as large as 150000 × 150000 pixels.

1. 2 Literature Review

1.2.1 Radiomic and Pathomic Biomarkers

Biomarkers are metrics, structures or processes in the body that can be measured using different modalities, including imaging and used objectively for diagnosis, characterization, and prognosis of a medical state [59]. Biomarkers can be derived from the analysis of biological samples such as blood and tissue specimens or using imaging modalities such as ultrasound [60], MRI [61], computed tomography (CT) [62], positron emission tomography (PET) [63] , and digital pathology images [64]. Several studies have demonstrated that imaging biomarkers could be adapted to characterize cancer or predict the outcome of anti-cancer therapies since the information contained within the images are associated with biological processes [65]–[67].

Due to the development of digital imaging scanners as the successors of the analog systems, the utilization of electronic signal detectors to implement more advanced systems, such as CT, MRI, and digital X-ray, has increased in previous decades. The widespread utilization of digital radiology has allowed greater opportunities for quantitative imaging and radiomics. Radiomics refers to the high-throughput extraction and analysis of high-dimensional features from medical images, possibly acquired with different modalities, to identify quantitative imaging biomarkers that are indicative or predictive of a medical state or outcome for specific clinical applications [68]. In oncology applications, radiomic descriptors may include morphological features of lesions, as well as first-, second-, or higher-order intensity features (*e.g.*, texture parameters) that can be linked to tissue microstructure and heterogeneity [65]. Texture features can be extracted from original or filtered medical images using statistical approaches, including gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), and neighborhood gray tone difference matrix (NGTDM) [69]. Multiple studies have shown the potential of radiomic features for stratification of tumor

histology, molecular subtypes and clinical stages and predicting various clinical outcomes, *e.g.*, tumor response, recurrence, and survival, after treatment [64], [70].

Unlike the radiomic features that are extracted from medical imaging modalities, the pathomic features are derived from digital microscopy images of tissue samples, cells, and subcellular structures. The pathomic features can be categorized into different categories, including the morphological, intensity and gradient-based, textural, graph-based, and filter-based features [71]–[74]. Utilizing these features for biopsy specimens requires the preparation of fixed tissue samples on histology slides, applying microscopy imaging to digitize the histology slides at high magnification, segmenting the regions of interest (ROIs) on digital histopathology images, feature extraction, reduction, and selection (biomarker discovery), and developing machine learning models to utilize the derived biomarkers in diagnostic, theragnostic and prognostic applications [75]. Figure 1.4 shows a general scheme of pathomic modeling frameworks. As conventional pathology reporting plays an important role in clinical decision-making, pathomic features can potentially be adapted for different clinical applications to help with diagnosis, tumor characterization, and therapy response prediction [76].

Previous pathomic studies have developed computational methods to automate the detection of conventional attributes of histomorphology (*i.e.*, nuclear shape, size, and color) and tissue-spatial characteristics such as nuclei distribution [77]. Pathomics aims to extract hand-crafted features, *i.e.*, features derived using well-define algorithms and mathematical equations to describe specific information present in an image, from histopathology images systematically to develop data-driven biomarkers for disease diagnosis, theragnosis and prognosis. In other words, the pathomic data refers to the immense array of data that is yielded from image analyses to achieve quantitative features to describe diverse phenotypic features of tissue samples in histopathology images. Analysis of digital

histopathology images using image processing techniques is included to detect, segment, label and classify regions in terms of cells, nuclei, and any other histological attributes [78]. Pathomic features can potentially be complemented with clinical, radiologic, and genomic data to improve their efficacy in diagnosing the diseases and predicting the treatment response and clinical outcomes [78].

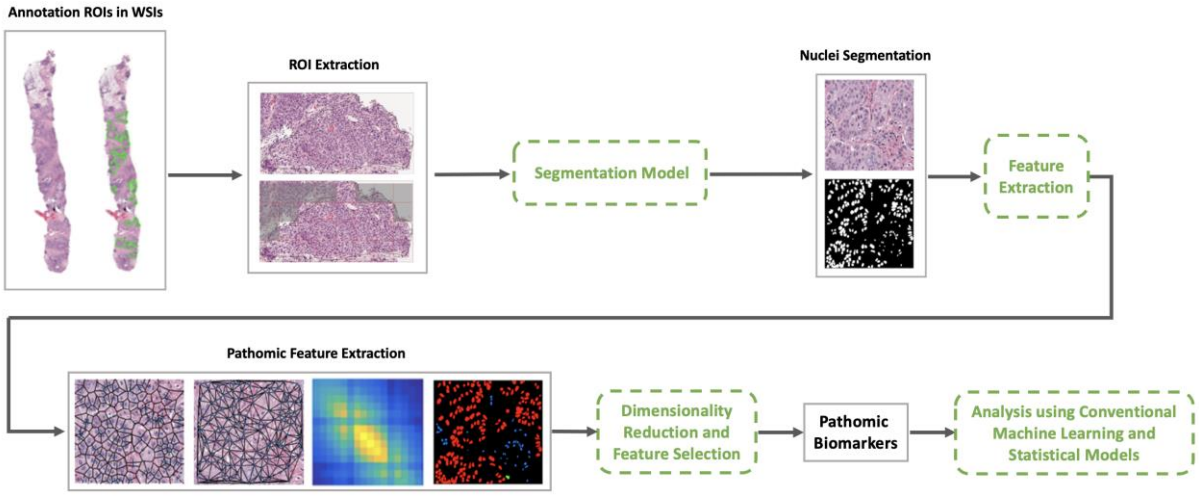


Figure 1.4 – General scheme of pathomic modeling framework. The regions of interest are extracted from the whole slide images, and nuclei are segmented within the ROIs. The pathomic features are usually derived from the segmented images can be categorized in different groups such as graph-based features, texture features, and morphological and intensity-based features. These features are processed through dimensionality reduction and feature selection procedures to develop pathomic biomarkers. The biomarkers are analyzed using machine learning and statistical modeling for different diagnostic and prognostic applications.

Whole-slide digital histopathology images contain thousands to millions of objects, including different classes of structures and various types of cells with different morphologies, such as epithelial breast cells, vasculature, and adipose cells. Several research projects are ongoing to quantify these histological features to possibly identify or describe the intrinsic biological behavior of different phenotypes [78], [79]. The pathomic features, similar to the radiomic features, require identification of appropriate ROIs (*e.g.*, tumor components versus stromal area, or nuclei within a tumor region) based on the specific application for feature extraction. Therefore, accurate segmentation of ROIs is crucial to extract relevant, robust, and reproducible features.

1.2.2 Segmentation Techniques

In histopathological analysis and pathomics, accurate detection and segmentation of different structures at microscopic scale (*e.g.*, lymphocytes and cell nuclei), and larger scale (*e.g.*, glands) are essential for achieving accurate and reproducible morphological statistics and, subsequently discriminative features [80]. Figure 1.5 and Figure 1.6 show examples of nuclei and glands segmented manually by pathologists.

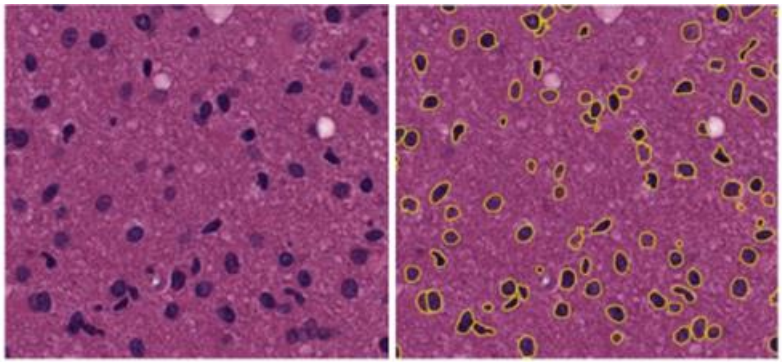


Figure 1.5 – Nuclei segmentation [64].

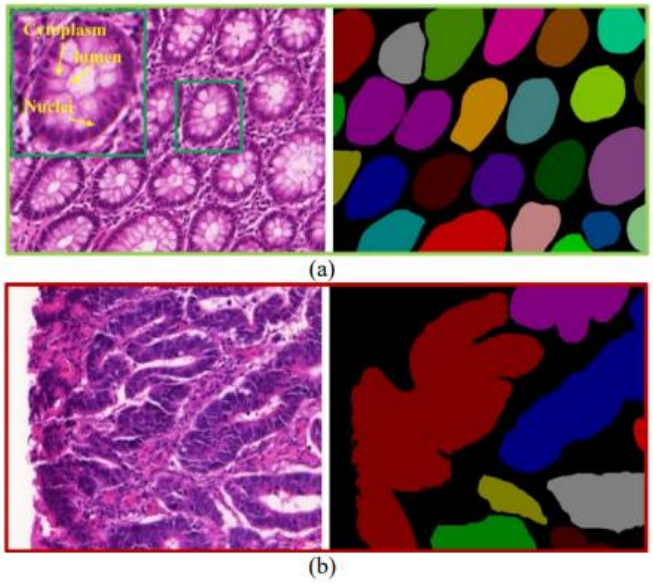


Figure 1.6 – Gland segmentation (a) benign versus (b) malignant [64].

The presence, extent, size, shape, and other morphological appearance of the structures are taken into account by pathologists as important factors to determine the presence or characteristics of an abnormality. Another motivation and importance for accurate segmentation of histological structures is the need for counting nuclei or whole cells, which facilitate diagnosis and characterization of the abnormalities of the specimen [81]. In pathomics, segmentation of relevant structures such as cell nuclei is important to derive morphological features and to quantify other characteristics of these structures on histopathology images, including those associated with the intensity, gradient, texture, and distribution [64], [82].

With the advent of deep learning methods and their applications in different fields of research from engineering to medical sciences, many deep networks, often consist of convolutional layers, have proposed for image segmentation [83]. The U-Net [84] [89] (Figure 1.7) was specifically introduced for biomedical image segmentation and could outperform the prior best models in the International Symposium on Biomedical Imaging (ISBI) challenge (2015) [85]. The architecture of U-Net contains two paths; the first one is the contracting path (encoder) that captures the context of the image. The second path, which is called the symmetric expanding path (decoder), is used to precisely localize with upsampling layers followed by deconvolutional layers [86]. The network contains 19-layers in total, which are all fully convolutional layers. Skip connections between different levels of the network overcome the trade-off between context and localization [87]. A limitation of the U-Net model in segmenting small objects such as nuclei in high-resolution images is the lack of accuracy in detecting the precise border of the close or touching objects.

The emergence of U-Net has opened pathways for variants of its architecture to enhance the performance of segmentation tasks in biomedical images. The attention U-Net was proposed to improve the segmentation performance by integrating grid-based attention gates (AGs) on top of U-

Net architecture [88]. Models trained with the AGs learn to suppress unnecessary regions in an input image while highlighting important features for a particular task. Each skip connection's gating signal incorporates image features from multiple imaging scales to remove the requirement of having explicit tissue localization modules. The attention U-Net model was originally proposed for segmenting the pancreas in computed tomography (CT) abdominal images [88], but then adapted in other applications, including nuclei segmentation [89].

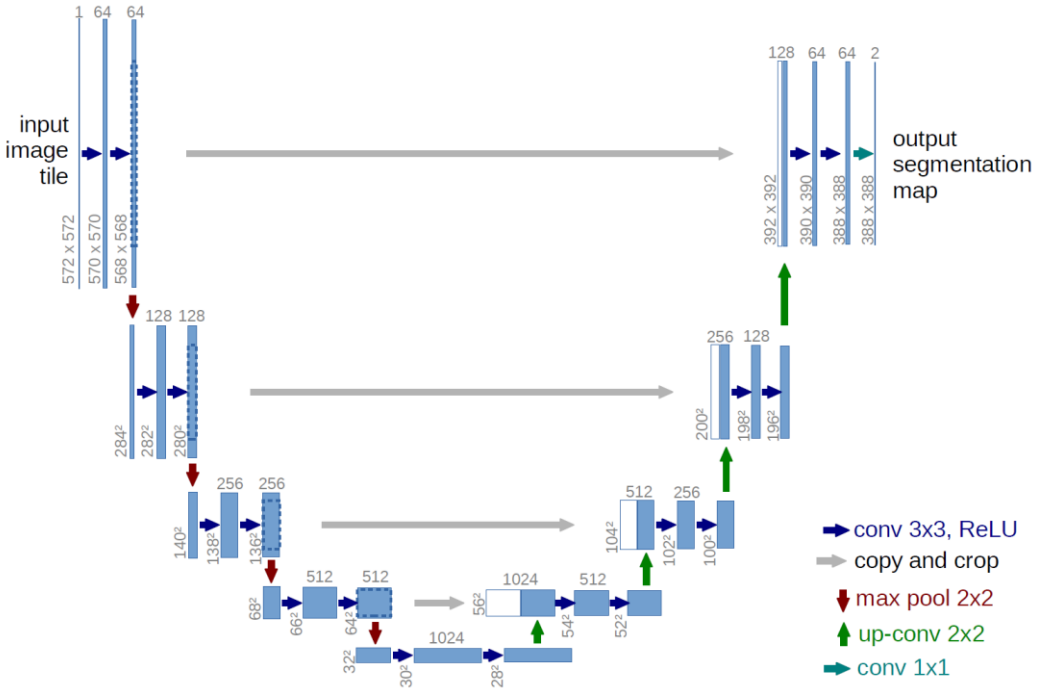


Figure 1.7 – Overview of U-Net architecture [84].

1.2.3 Automated Tissue Classification Models

One of the challenging tasks in cancer-related research is rapid characterization of different tissues, *e.g.*, normal, benign, and malignant tissues, with high accuracy. The gold standard method for tissue characterization in pathology is histopathology analysis on the tissue specimens. Recently, the

advent of high-throughput digital histology imaging systems has provided the opportunity for automatic tissue characterization with the aid of machine learning models [90], [91].

A number of models based on convolutional neural networks (CNNs) have previously been introduced for the classification of digital histopathology images to characterize the tissue samples. In [92], histopathology images at different magnifications were classified into binary labels (*i.e.*, benign versus malignant) using a CNN model inspired by the ImageNet model [93]. Several patches with different sizes were extracted from each image and used to train the CNN model. The patch extraction was used to handle the limitation of the small dataset. The classification was performed by concatenating the class probabilities associated with patches of each image using maximum, product or sum rules. A limitation of that study is that the classification model was trained on small patches and local features only; consequently, this omits the global information that would have been yielded from the whole-image analysis. In [94], a CNN architecture was designed to capture information of histopathology images at multiple scales to classify them into four different classes of normal tissue, benign lesion, *in situ* carcinoma and invasive carcinoma. The multi-scale strategy was adapted to retrieve information from nuclei and whole tissue organization simultaneously. The features extracted from the CNN model were fed into a support vector machine (SVM) for final classification. In [95], for classifying histopathology images into the four classes mentioned above, a patch-wise approach was used with an auto-encoder network to extract the features of tissue micro-structure at high magnification from each image. The auto-encoder network was then followed by an image-wise CNN to classify the whole image. The models developed in [94] and [95] utilized both local and global features for classification; however, all features were assigned an equal weight for classification. Consequently, the irrelevant features of an image may reduce the accuracy of the model in classifying the whole slide.

1.2.4 Machine Learning in Pathomics

The advancement of machine learning opens new avenues of research to help pathologists and clinicians with accurate diagnosis, prognosis and eventually personalized treatment planning. Multiple studies have demonstrated the potential of machine learning methods in finding ROIs within images of different modalities such as MRI, CT and histopathology images [96]–[98]. A number of other studies have investigated various radiomic and pathomic features extracted from such ROIs for different diagnostic and prognostic applications [64], [99], [100].

Recent studies have shown promising results for the prediction of treatment outcome and cancer recurrence using the combination of pathomic and radiomic features [101], [102]. In [101], for the prediction of recurrence in early-stage non-small-cell lung cancer (NSCLC) patients, 124 radiomic features were extracted from CT images within the ROIs delineated manually by an expert person. The extracted features included Gabor features, Haralick features and Laws features. For extracting the pathomic features, first, the whole-slide H&E-stained histopathology images were analyzed using a CNN-based architecture to segment the nuclei automatically. A total of 242 pathomic features, including Gabor graph features, local nuclear cluster graph, nuclear shape features, nuclear orientation entropy, and nuclear texture features, were extracted from the segmented nuclei. A maximum relevance minimum redundancy (mRMR) feature selection method, which tends to choose a subset of features that are least correlated among themselves and most correlated with the class, was used to extract the most informative features. Three different classifiers (*i.e.*, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and support vector machine (SVM)) were adapted to classify each patient in terms of recurrence outcome. This model could achieve promising results for recurrence prediction of an independent test set and outperformed a number of previous models, including the model proposed by Wang et al. that uses nuclei shape,

architecture and texture features as a predictor of recurrence in early-stage NSCLC [103]. However, in the proposed framework, the ROIs were selected manually in CT images, which is time-consuming. In addition, the feature selection method was naïve in comparison to deep learning methods. With the aid of deep neural network methodologies, the outcome prediction can be performed by developing fully automatic models as end-to-end architectures with potentially higher performance.

1.2.5 Deep Learning Approaches in Histopathology Image Analysis

Recent research has explored the use of deep learning methods in various medical image analysis applications [104]. One superiority of these methods compared to conventional ML models is that the process of extracting hand-crafted features can be eliminated. Instead, during the iterative training phase, the deep learning frameworks optimize their data-driven filter weights to extract optimum feature maps from the input images [105]. A number of studies have investigated deep learning frameworks for analysis of histopathology images in terms of grading and subtyping of tumor tissue and more advanced tasks such as survival prognosis [106]–[109]. In [106], a deep convolutional neural network architecture called DeepSurvNet was developed to predict the survival rate in brain cancer patients using histopathology images. Five well-known architectures were utilized as the backbone of the proposed model for comparison, including: VGG19, GoogleNet, ResNet50, InceptionV3 and MobileNetV2. In order to relax the problem and utilize the classifiers' abilities, the survival rates of patients were categorized to four different classes. The patches of various sizes were extracted from ROIs including 256×256, 512×512, and 1024×1024 pixels to compare the effect of input size on the framework. This study has shown that the GoogleNet model with 256×256 pixel input image size, can outperform other models in terms of highest average precision for classifying survival rates. In [108], InceptionV3 model was trained and

evaluated to classify the histology lung tissues into three classes of adenocarcinoma (LUAD), squamous cell carcinoma (LUSC) or normal lung tissue by extracting small patches from ROIs. The proposed model was utilized to predict the ten most mutated genes in LUAD, and the results have shown that six of them can be predicted using histopathology images, including: STK11, EGFR, FAT1, SETBP1, KRAS and TP53. Although these studies have demonstrated the potential of deep learning models and specifically the deep convolutional networks in histopathology image analysis, the size of WSIs remains a constraint in optimizing such models, and developing predictive models of therapy outcome using WSIs still remains challenging due to their complex natures.

1.3 Objectives

The overarching goal of this dissertation project is to facilitate a personalized medicine paradigm for breast cancer by streamlining the histopathology assessments workflow and predicting cancer response to prescribed treatments for individual patients using the data available at pre-treatment. The focus of this research is to develop, optimize, and evaluate innovative machine learning approaches and quantitative image analysis methods for ROI segmenting, tissue characterization, and therapy response prediction to identify patients who are at risk of suboptimal treatment response using digital histopathology images of pre-treatment biopsy specimens. As described earlier, conventional machine learning and deep learning methods are two main approaches that can be utilized for intelligent analysis of digital histopathology images. Both these approaches are investigated in this dissertation project to pursue its goal through four objectives.

As the first objective, a novel breast tissue characterization model is introduced and investigated for classifying digital histopathology images using CNN-based deep learning architectures. This

model is developed as an automated system for cancer diagnosis and characterization, as it is an important step for breast cancer management.

As the second objective, a novel U-Net-based framework is proposed and investigated for nuclei segmentation on digital histopathology images and extraction of ROIs for further analysis. Nuclei segmentation is one of the pivotal components of an automated system aiming to mine quantitative features of tissue on digital histopathology images.

As the third objective, a machine learning framework is introduced to predict breast cancer response to neoadjuvant chemotherapy using hand-crafted pathomic features extracted from histopathology images at pre-treatment. Various models developed based on this framework are investigated to explore the potential of histopathology images of pre-treatment biopsy specimens with and without standard clinical attributes in therapy response prediction.

As the fourth objective, an innovative deep-learning architecture is introduced with convolutional and transformer blocks to analyze the whole-slide digital histopathology images automatically for chemotherapy response prediction at pre-treatment. The model is investigated through several ablation experiments on its component and compared with similar models with different architectures in terms of performance on an independent test set.

1.4 Thesis Outline

The thesis objectives explained above have been presented in four chapters of this dissertation, followed by a concluding chapter in which a summary and future directions are explored. The content of each chapter is summarized below.

1.4.1 Chapter 2

In Chapter 2, a novel multi-scale attention-guided deep learning model is proposed to characterize breast tissue on digital histopathology images according to four different histological types, including normal, benign, *in situ* carcinoma and invasive carcinoma. The framework includes two parallel convolutional neural networks with modified VGG16 architecture. The first network analyzes the whole-sample images at low magnification. The second network focuses on the patches extracted from the whole-sample images at high magnification using an attention mechanism. At the end, the output probability vectors of both networks are fused using a multilayer perceptron network to generate a tissue classification label. The results demonstrate that a multi-scale strategy coupled with an attention mechanism can improve the accuracy of deep learning models in classifying digital histopathology images.

1.4.2 Chapter 3

In Chapter 3 a novel segmentation model is introduced to delineate nuclei in digital histopathology images accurately, that is an essential step for extracting hand-crafted pathomic features. A cascaded deep learning U-Net based model is proposed to overcome the issue of delineating the border of touching and closing nuclei in histopathology images. The obtained results demonstrate the effectiveness of the proposed U-Net based model with customized weighted loss function followed by a U-Net based model with VGG16 backbone and a soft Dice loss function in segmenting the nuclei accurately.

1.4.3 Chapter 4

Chapter 4 investigates quantitative digital histopathology analyses coupled with machine learning to predict tumor response to neoadjuvant chemotherapy *a priori*. A histopathology image analysis

framework has been developed to extract various subsets of pathomic features, including the morphological, intensity-based, texture, graph-based and wavelet features, from segmented digitized slides. Several machine learning experiments have been conducted with different feature sets to develop prediction models of therapy response using a gradient boosting machine with decision trees. The results demonstrate the potential of quantitative digital histopathology features integrated with machine learning methods in predicting breast cancer response to neoadjuvant chemotherapy.

1.4.4 Chapter 5

In Chapter 5, a hierarchical self-attention-guided deep learning framework is introduced to predict pathological complete response to chemotherapy in breast cancer patients using digital histopathology images of pre-treatment tumor biopsies. The WSIs are processed automatically through the proposed hierarchical framework consisting of patch-level and tumor-level processing modules followed by a patient-level response prediction component. A combination of convolutional layers and transformer self-attention blocks are utilized in the patch-level processing architecture to generate optimized feature maps. The feature map sequences are defined based on the patch positions within the tumor beds and the bed positions within the biopsy slide for analysis through two adapted vision transformer architectures for tumor-level processing and patient-level response prediction components, respectively. The results demonstrate the high potential of the proposed hierarchical deep-learning methodology in the analysis of very large-size digital histopathology images to predict pathological response to NAC at pre-treatment.

1.4.5 Chapter 6

The material presented in chapters 2 to 5 is summarized in this chapter. It also brings this dissertation to a conclusion by outlining potential future directions for the presented project.

1.5 Contributions

This dissertation proposes and investigates novel ML methods for breast tissue characterization, nuclei segmentation, and chemotherapy response prediction in breast cancer patients using pre-treatment digital pathology images. The principal contributions of this dissertation are summarized below.

As the first contribution, an innovative multi-scale attention-guided deep learning model is proposed to characterize breast tissue on digital histopathology images according to four different histological types, including normal, benign, *in situ* carcinoma and invasive carcinoma. The proposed model incorporates local features and global relations between them by fusing the low-magnification and high-magnification networks. Another contribution of proposed architecture is integrating an attention mechanism in the low-magnification network, that refines the feature weights to improve the network's decision-making process.

The second contribution of this dissertation is proposing a cascaded U-Net-based deep-learning model to overcome the issue of delineating the border of touching and closing nuclei in histopathology images. The suggested architecture consists of a pixel-wised weighted U-Net model, followed by a U-Net model with VGG16 backbone and soft dice loss function. The weighted U-Net model is used to distinguish between the border of touching and closed nuclei accurately, while the second model is utilized to penalize low-confidence predictions for maximizing the correctly detected pixels.

The main contribution of this dissertation is proposing and investigating, for the first time, two automated ML-based platforms for NAC response prediction in breast cancer patients using pre-treatment histopathology images. In the first approach, the hand-crafted pathomic features extracted

from the pre-treatment histopathology images are used in conjunction with conventional ML models to predict pCR versus non-pCR. Different experiments have been conducted to assess various initial feature sets. A gradient boosting machine (GBM) has been optimized as the predictive model using the selected features (optimal biomarker) in each experiment. In the second approach, a deep-learning-based hierarchical framework is proposed for predicting the NAC response using the pre-treatment histopathology images. In this approach, three levels of processing are cascaded including a patch-level processing to extract the local features, a tumor-level processing to aggregate the intra-tumor information, followed by a patient-level processing which exploits the global dependencies in the whole slide images for response prediction. The proposed hierarchical framework consists of a CoAtNet architecture with convolutional and self-attention modules as the patch-level processing component, followed by two vision transformer (ViT) architectures as the tumor-level and patient-level processing components. The proposed predictive models have shown a good potential for predicting the NAC response at pre-treatment that can pave the way toward personalized medicine for breast cancer patients.

Chapter 2

Automatic Characterization of Breast Lesions using Multi-scale Attention-Guided Deep Learning of Digital Histopathology Images*

2.1 Introduction

Breast cancer is the most commonly diagnosed malignancy and the leading cause of cancer-related death in women [110]. The lifetime probability of developing invasive breast cancer is approximately 12% (1 in 8 women). Fortunately, breast cancer mortality has decreased by 40% over the past 30 years due to improved treatment strategies, earlier detection, and enhanced diagnostic technologies [111].

Clinical imaging using digital mammography, tomosynthesis, ultrasound, and magnetic resonance imaging (MRI) are well-established methods to screen for breast cancer [44], [112]. However, the gold standard for diagnosis is histopathological examination [42]. To carry this out, tissue biopsies of the suspicious breast lesion are extracted and prepared onto histological slides. The specimens are stained with hematoxylin and eosin (H&E), and subsequently analyzed under microscopy to assess the cellular, morphological, and other biomarker-based attributes. Pathologists draw on their

*A version of the material presented in this chapter has been published in Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. In Press, 2022. DOI: 10.1080/21681163.2022.2058415.

cognition, perception, experience, and judgement to distinguish between normal, versus benign, or malignant tissues [113]. However, manual assessment is time consuming, resource-intensive and requires substantial expertise. To address these clinical workflow challenges, recent efforts have focused on developing computational methods to carry out automatic analysis and classification tasks to assist pathologists. The introduction of digital pathology imaging platforms has created new opportunities to develop these computational frameworks; specifically exploiting machine learning methods for identifying tissue abnormalities and resolving tissue classification problems. The potential benefits of computationally driven approaches include better access to pathology services and increased efficiency in the pathology workflow [114]. In this context, computer-aided systems can be adapted to streamline the histopathology workflow, facilitate fast and accurate decision-making, and enable high-throughput reporting to manage high-volume caseloads.

Machine learning (ML) frameworks, and specifically deep learning (DL) techniques have made substantial progress to handle complex pattern recognition tasks such as image classification [115]. In particular, DL has provided superior achievement in different applications of medical image analysis including detection, segmentation, and characterization of various abnormalities, *e.g.*, cancer [116],[117]. The advent of fast digital slide scanners which can automatically generate high-resolution microscopy images of multiple slides in a batch has resulted in increasing interest and potential in developing data-driven models for analysis of digital histopathology images [118]. However, automated tissue characterization in digital histopathology images using deep learning approaches still encounter various technical challenges [44]. Some of these challenges stem from the intrinsic properties of histopathology images. Firstly, the images processed by convolutional neural networks (CNNs) have typically a small size, but digital histopathology images tend to have notably large sizes, specially at higher magnifications. As a result, deep convolutional neural

networks (DCNNs) cannot be trained directly on large whole-slide images. Another challenge is the size of the dataset. In order to train a DCNN, a large set of relevant images are typically required. However, large datasets of digital pathology images are usually not available in many applications [119].

Previous studies have developed and evaluated various methods for automatic segmentation and classification of breast tissue in digital histopathology images [120], [121]. Among these studies, CNNs have been proposed for classification of breast histopathology images. In [114], histopathology images at different magnifications were classified into binary labels (*i.e.*, benign versus malignant) using a CNN model inspired by the ImageNet model [122]. Several patches with different sizes were extracted from each image and used to train the CNN model. The patch extraction was used to handle the limitation of a small dataset. The classification was performed by concatenating the class probabilities associated with patches of each image using maximum, product or sum rules. A limitation of that study is the biased classification model that was trained on small patches and local features only; consequently, this omits the global information that would have been yielded from whole-image analysis. In [123], a CNN architecture was designed to capture information of histopathology images at multiple scales to classify them into four different classes of normal tissue, benign lesion, *in situ* carcinoma and invasive carcinoma. The multi-scale strategy was adapted to retrieve information from nuclei and whole tissue organization simultaneously. The features extracted from the CNN model were fed into a support vector machine (SVM) for final classification. In [124], for classifying histopathology images into the four classes mentioned above, a patch-wise approach was used with an auto-encoder network to extract the features of tissue micro-structure at high magnification from each image. The auto-encoder network was then followed by an image-wise CNN to classify the whole image. The model developed in that study utilized both

local and global features for classification; however, all features were assigned an equal weight for classification. Consequently, the irrelevant features of an image may reduce the accuracy of the model in classifying the whole slide. In [125], multiple compact CNNs were utilized within a hybrid architecture that contains a global and a local branch to classify histopathology images into the four aforementioned classes. A local voting was used to combine the results of two branches. In another study to classify the histopathology images in the BACH dataset [126], the fine-tuned Inception-v3 CNN model was fed with the patches with high nuclear density [127]. The patches from uninformative regions of each image such as non-epithelial areas were automatically excluded at an initial step. The final class of each image was determined using a majority vote over the corresponding nucleus-rich patches. In a different study, an ML pipeline was developed to combine an SVM model (shallow learner) with an Inception-v3 model (deep learner) for classification of breast histopathology images, where the integrated model outperformed both the individual models [128].

An attention mechanism in deep learning refers to a training strategy through which higher weights are assigned systematically to certain factors of interest when analyzing the data [129]. The attention mechanism was initially introduced in natural language processing (NLP) applications, and it has recently been adapted in other applications of deep learning such as image analysis and computer vision [130]. The attention could be considered as a component of a deep neural network that quantifies the interdependence between the input and output of the model (general attention), or just within the input of the model (self-attention) [131]. The self-attention concept has recently been adapted in medical image analysis applications to train the network to pay less attention to the irrelevant parts of an input image while highlighting more important features for specific tasks such as classification [132]. Attention can be incorporated into a network using a number of different

approaches, including a recurrent mechanism and a channel/spatial attention [133][134]. In recurrent attention mechanism, the network iteratively uses the information obtained in the previous step to find a next proper region of focus instead of processing an entire image at once. The network gradually merges information at each step using its hidden state, to build up a dynamic representation of each image [135]. In channel/spatial attention mechanisms, the channel attention module is responsible to identify what is relevant for the task in an input image as each channel of a feature map is considered as a feature detector, whereas the spatial attention module focuses on finding the informative regions of the input image [136][137]. The attention mechanisms have widely been used for image classification, and recently adapted for digital histopathology analysis in tissue characterization applications. In [138], a CNN with attention mechanism was designed for classification of esophageal tissue to Barrett's esophagus (BE) and esophageal adenocarcinoma. A grid-based feature extraction approach was used to generate feature maps from high-resolution histopathology images. The feature maps were fed to a soft attention mechanism to produce weighted maps that were used for the prediction of final class labels. Although the applied attention mechanism can potentially assign relevant weights to each feature in high-resolution histopathology images, the dependencies between different grids were not considered in this study. A recently proposed attention mechanism that can be utilized to infer separate channel and spatial attention maps is the convolutional block attention module (CBAM) [133]. An advantage of the CBAM is that it can be integrated with any feed-forward CNN with minimal overhead as it is a lightweight module. The CBAM consists of cascaded channel and spatial attention modules. In the channel attention module, the max-pooled and average-pooled features are calculated from the input feature maps, and passed through a multi-layer perceptron (MLP) network with shared weights. To compute the spatial attention, channel-wise average-pooling and max-pooling operations are applied on the

channel-refined input features and the resulting maps are concatenated to generate an efficient feature descriptor. A convolution layer is then applied to generate a spatial attention map. At the end, element-wise multiplication will be performed between the attentions map and input features to obtained refined adaptive feature maps. The CBAM attention module has been implemented with many CNN architectures where it has demonstrated improvement in object detection tasks compared to the original CNN without attention[133][139].

In this chapter, a multi-scale attention-guided deep learning model was investigated for classifying histopathology images of breast tissue specimens according to four types: 1) normal tissue, 2) benign lesion, 3) *in situ* carcinoma and 4) invasive carcinoma. The proposed model consists of two streams of deep learning on histopathology images at high- and low-magnification integrated with a recurrent attention mechanism. The hierarchical structure of the model was proposed to incorporate the global and local features of histopathology images in tissue characterization, simultaneously. The two parallel networks are followed by MLP network to classify the histopathology images. The VGG16 CNN architecture was adapted for both the high and low-magnification networks [140]. In the low-magnification network, class activation maps were extracted by adding a global average pooling layer at the end of the network. These maps were fed into the recurrent attention mechanism to increase the contribution of the salient features to the model's decision. In a separate experiment, the CBAM attention was integrated with the low-magnification network to compare its performance with the recurrent attention mechanism. At the last layer of the high-magnification network the probability vectors associated with the four histological classes were calculated for each image by averaging over all patches extracted from that image. The model was trained on the histopathology images acquired from 320 samples and subsequently evaluated using a test set with 80 samples. The proposed model outperformed the separate high- and low-magnification networks, and the multi-

scale model without an attention mechanism and with the CBAM attention mechanism in characterizing the histopathology images of the test set. The results demonstrated that a multi-scale strategy integrated with a recurrent attention mechanism can improve the accuracy of automatic tissue characterization models using digital histopathology images.

2.2 Materials and Methods

2.2.1 Dataset

The data applied in this study for development, optimization and evaluation of models were acquired from the ICIAR 2018 grand challenge on Breast Cancer Histology (BACH) [141]. The ICIAR 2018 BACH dataset includes 400 digital microscopy images of H&E-stained breast tissue specimens with four different histological types, acquired at 40× magnification. The dataset includes region of interest (ROI) images (2048 × 1536 pixels) that are extracted from the whole-slide images and annotated according to the histological type by experts. The ROI images are uncompressed in a tiff format. The histological types include normal, benign, *in situ* carcinoma, and invasive carcinoma. The dataset includes an equal number of samples for each histological type (n = 4 × 100). The dataset was split randomly into training (80%) and test (20%) sets for development and evaluation of the models, respectively.

The values in the RGB channels of all images were normalized into the interval [0 1] by dividing by the maximum value of each channel (255). The proposed model consists of two parallel CNNs working on histopathology images at high and low magnification, respectively (described in Section 2.2.2). For the high-magnification network, 54 non-overlapped patches with a size of 224 × 224 pixels were extracted from each ROI image with the size of 2048 × 1536 pixels. For the low-magnification network, the ROI images were resized to 512 × 384 pixels using a bilinear

interpolation [142]. In order to augment the training data, and also make the networks rotation invariant, the low- and high-magnification images of the training set were randomly rotated by 0 to 360 degrees and added as extra training samples.

2.2.2 Model Architecture

The proposed model consists of two streams of CNNs processing the histopathology images at high- and low-magnification, followed by a recurrent attention mechanism. The high-magnification network in the first stream performs a patch-wise classification, whereas the low-magnification network in the second stream carries out an image-wise classification. The attention model explores the correlation between the class labels and their contributing regions in each image to assign higher weights to the more relevant regions for characterization. Subsequently, a multi-layer perceptron (MLP) model fuses the output of the two streams and generate the final label for each sample. Figure 2.1 shows an overview of the proposed model. More details about the model architecture are provided in the following sections.

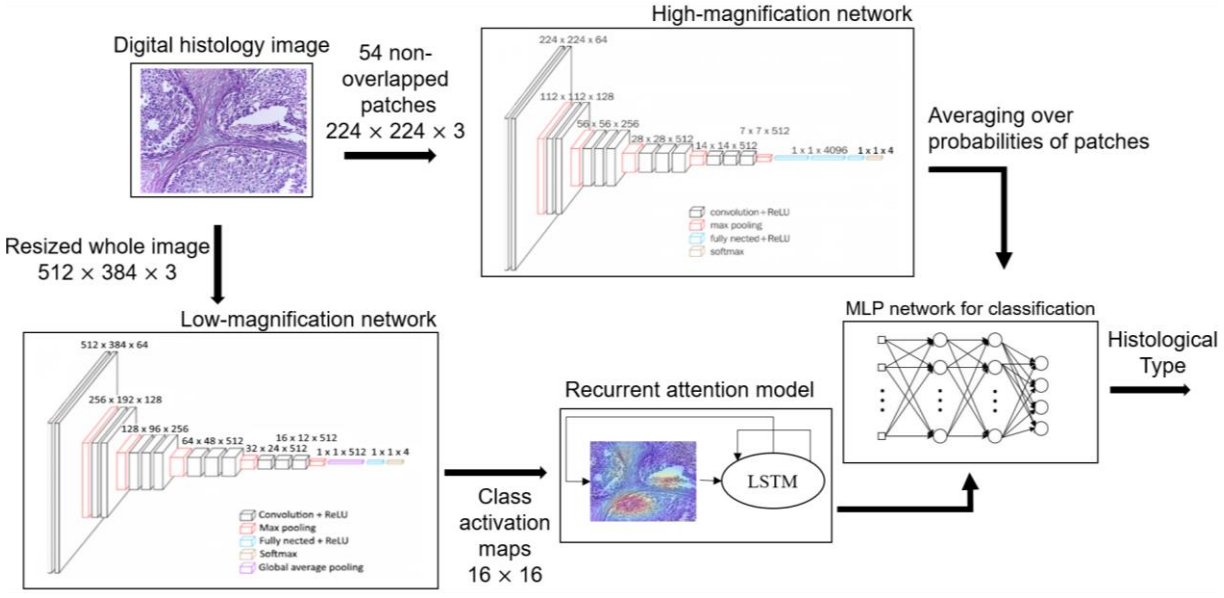


Figure 2.1. An overview of the proposed model. The VGG16 networks were trained independently on patches (high magnification) and whole images (low magnification). The recurrent attention methods were fed with the CAMs from the low-magnification network. The final class label for each sample was generated using the MLP network.

2.2.2.1 High and Low-Magnification Networks

The VGG16 CNN architecture was adapted for both the high- and low-magnification networks [140]. The weights of a pre-trained model on ImageNet were used as the initial weights in the high-magnification network [122]. The low-magnification network was trained from scratch with randomly generated initial weights. A global average pooling (GAP) layer [143] was added at the end of the low-magnification network to generate class activation maps (CAM) [144]. All connection weights between the GAP layer (after a SoftMax activation) and the fully connected layer for each class were extracted. The feature maps of the last convolutional layer of the network were also obtained. These feature maps were serially multiplied by their corresponding weights and summed up to generate the CAMs. The CAM for each class was used to visualize the network spatial attention associated with that class in form of a heat map overlaid on the corresponding sample [145]. The CAMs from the low-magnification network were fed into the attention model. The class probabilities generated in the last layer of the high-magnification network were averaged over all patches of each sample to obtain the sample-wise probabilities that were input the fusing MLP network. The high- and low-magnification networks were trained independently with a batch size of 8, a learning rate of 0.001, and 300 epochs using the Adam optimizer [146], and on an NVIDIA GeForce 2080 RTX GPU.

2.2.2.2 Recurrent Attention Mechanism

The CAMs generated by the low-magnification network was fed into the recurrent attention model. The CAMs (16×16) were vectorized to a size of 256×512 to be compatible with the number of hidden states (512) in the LSTM model. The attention model was incorporated to assign higher weights in calculating the class probabilities to the feature maps obtained for the regions of each image (sample) with a higher relevance to the corresponding class. The attention mechanism

consisted of a single layer long short-term memory (LSTM) with a hidden state of size 512 followed by a sigmoid activation function, and an output unit size of 16 before the fully connected layer [147], [148]. The LSTM network iteratively estimated the probability vector of class labels at each time step using the previous hidden state, a context vector and the class probabilities estimated in the preceding steps. At each time step, the estimated context vector represented the relevance score for different regions of each image in terms of conditional probability. The attention network was trained with 49 time-steps and 100 epochs.

2.2.2.3 Multi-Scale Decision Fusion

An MLP network was used to fuse the class probabilities estimated by the high-magnification and attention-guided low-magnification networks for each sample. The MLP network consisted of an input layer of size eight, three hidden layers of size eight, and an output layer of size four. A stochastic gradient descent (SGD) optimization with momentum was used for training the network with a batch size of 8, and 100 epochs. For each sample, the MLP network generated a class probability vector of size four including the estimated probability for each histological type. The histological type corresponding to the highest estimated probability was determined as the final predicted class for each sample.

2.2.3 Evaluation

The performance of the models was evaluated using the accuracy, loss, F1-score and Matthews correlation coefficient (MCC) [149], in addition to the confusion matrix, in classification of the histopathology images of the test set. Each model was evaluated five times using different validation sets selected randomly from the training set (10%) before the start of training process. In each experiment, the model was trained from the scratch with completely disjoint training and validation sets. The test set was kept unchanged in all experiments. The evaluation was performed on the low-

and high-magnification networks separately, and in combination, with and without the attention mechanism. Separate models were also developed with the recurrent and CBAM attention mechanisms for comparison. Specifically, the models were evaluated with the CBAM attention combined with the low-magnification network as an alternative to the recurrent attention mechanism.

2. 3 Results

Figure 2.2 demonstrates representative samples of digital histopathology images applied in this study for normal breast tissue, benign lesion, *in situ* carcinoma, and invasive carcinoma. The images demonstrate that various local and global information spots are potentially complementary in characterizing the samples and can be captured at high- and low-magnification, respectively.

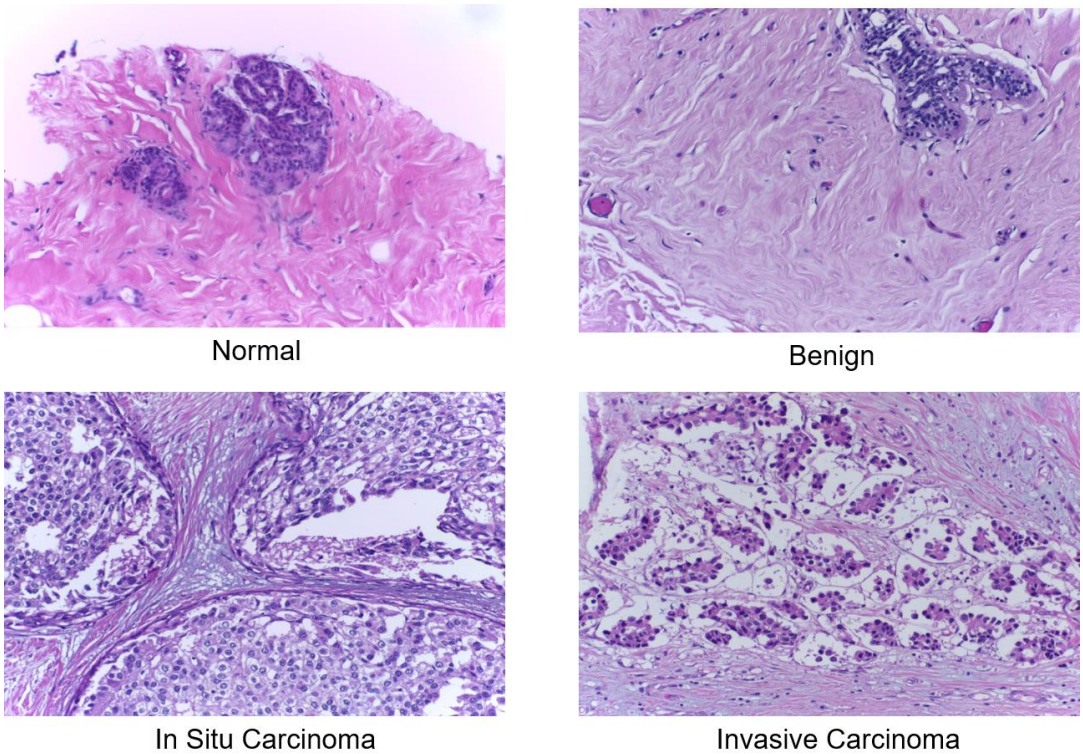


Figure 2.2. Representative microscopy images (at 40×) of H&E-stained breast tissue samples with different histological types from the dataset.

Figure 2.3 demonstrates representative images of different histological types overlaid with the attention heat maps. The heat maps visualize the relative weights of the feature maps associated with different regions of each image in classifying the corresponding sample. In other words, the attention maps show the relative contribution of different regions to the model's decision. The attention maps indicate that the network put more focus on regions with histological structures containing nuclei as well as the surrounding such structures.

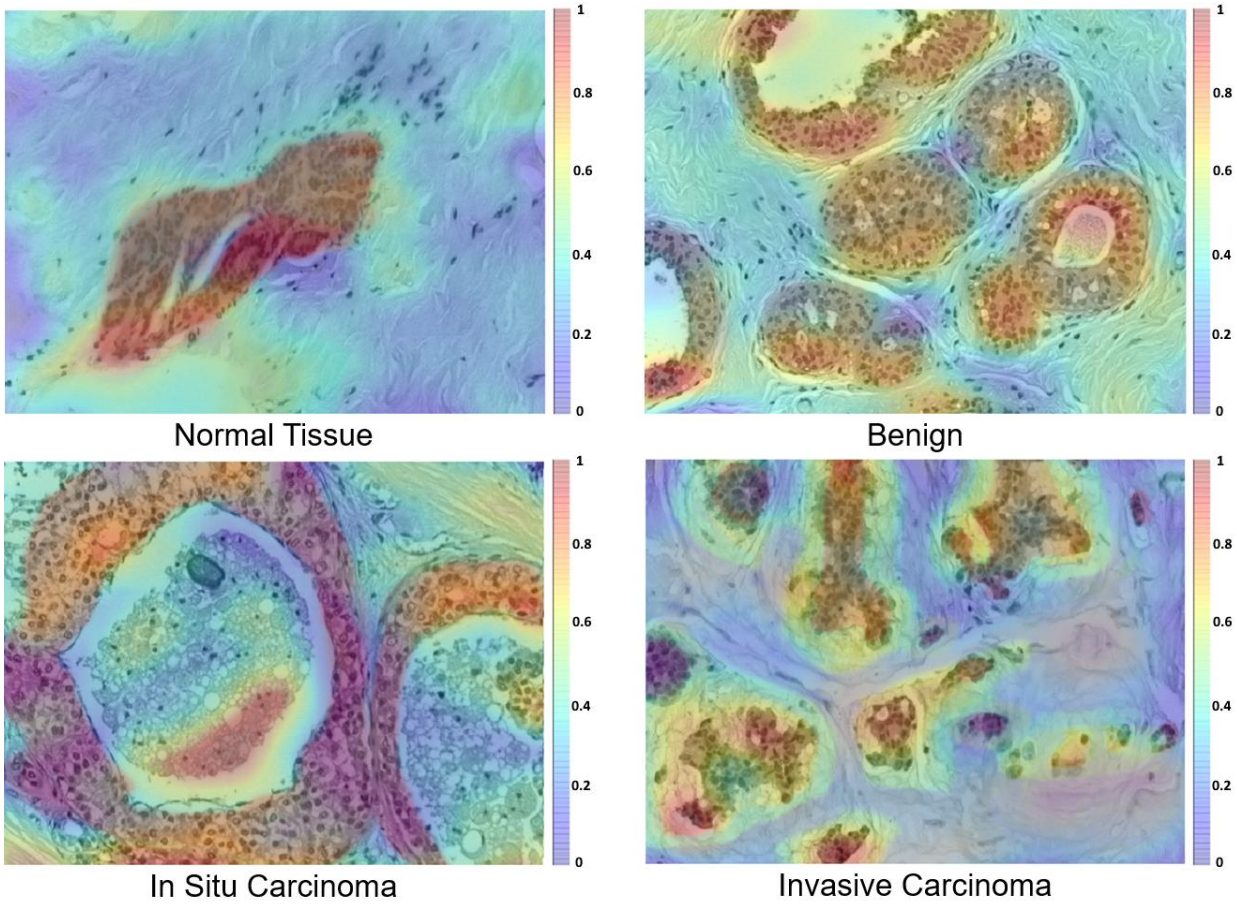


Figure 2.3. Attention heat maps for representative samples (imaged at 40×) with different histological types. The recurrent attention mechanism assigned higher weights to the regions with nuclei-containing structures and their surroundings.

Performance of different models on representative images of four histology types of breast tissue has been demonstrated in Figure 2.4. The individual high- and low-magnification networks

demonstrates lower performance compared to the multi-scale model in differentiating images of different histology types that include similar features, including normal versus benign tissues. Incorporating the attention mechanisms improve the performance of both low-magnification and multi-scale models, in most cases. For example, for the histopathology image associated with invasive carcinoma that is misclassified by the models without attention as *in situ* carcinoma or normal tissue, the models with attention demonstrate an accurate classification.

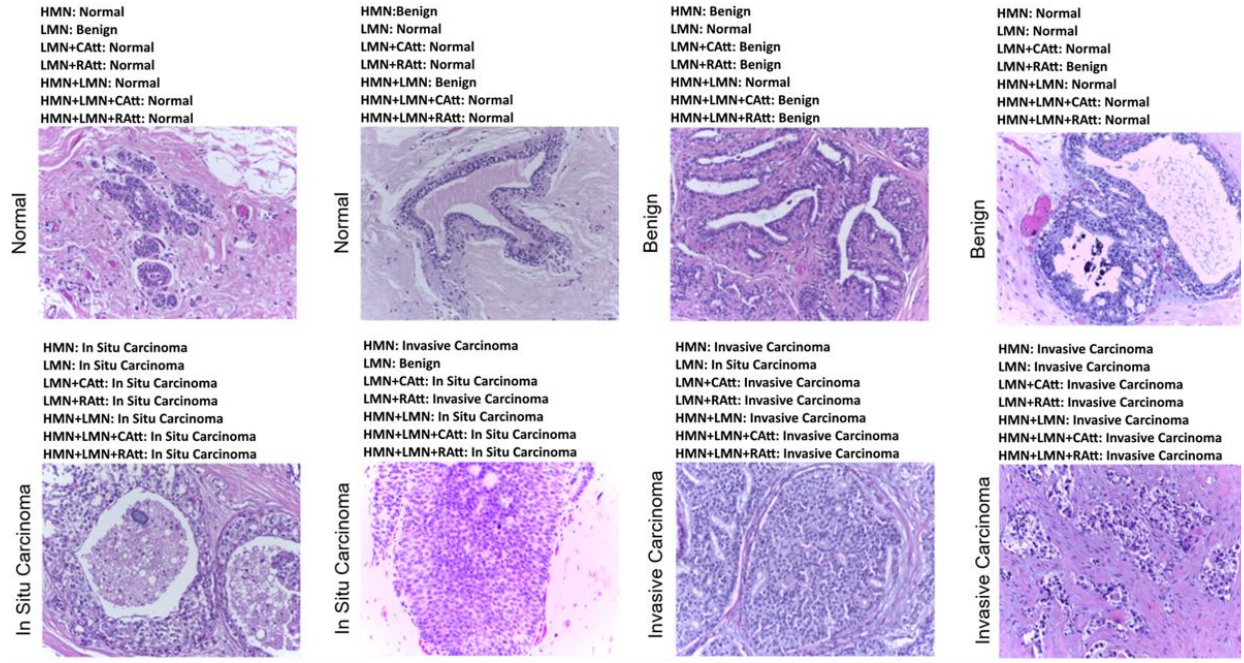


Figure 2.4. Representative histopathology images classified by different models. The ground truth labels have been shown on the left side of each image. The annotations made by different models are listed on the top of each image. HMN: High-magnification network, LMN: Low-magnification network, CAtt: CBAM Attention mechanism, RAtt: Recurrent Attention mechanism.

The performance of different models evaluated in this study for characterizing breast tissue samples in digital histopathology images have been summarized in Figure 2.5 and Table 2.1. The average confusion matrices obtained over five experiments for the high- and low-magnification networks separately, and in combination, with and without the CBAM and recurrent attention

mechanisms have been presented in Figure 2.5. Table 2.1 presents the mean accuracy and loss obtained for each model over the five experiments. The same training, validation and test sets were used for all models in each experiment. The results indicate that the multi-scale model outperformed the individual high- and low-magnification networks with a mean accuracy of 96.3% ± 1.2% versus 94.5% ± 0.6% and 93.5% ± 1.0%, respectively, on the test set. Further, the multi-scale model demonstrated an improved mean loss of 0.6 ± 0.1 versus 0.7 ± 0.3 and 0.8 ± 0.2 obtained, respectively, for the high- and low-magnification networks on the test set. The results also indicate that incorporating the attention mechanisms improved the performance of the models in terms of loss, accuracy, F1-score and MCC. The low-magnification network with recurrent attention classified the images with more confidence demonstrating a mean loss and accuracy of 0.5 ± 0.1 and 95.5% ± 0.8%, and a mean F1-score and MCC of 95.6% ± 0.7% and 0.94 ± 0.01, respectively, on the test set. The same model without attention resulted in a mean loss and accuracy of 0.8 ± 0.2 and 93.5% ± 1.0%, respectively. The low-magnification network with CBAM attention could slightly outperform the low-magnification network with recurrent attention in terms of mean accuracy (95.8% ± 0.4%) and F1-score (95.7 ± 0.4), but not loss and MCC. The best model in terms of both loss and accuracy was the proposed multi-scale attention-guided model. This model outperformed the multi-scale model without attention and the multi-scale model with CBAM attention with a mean loss and accuracy of 0.4 ± 0.1 and 97.5% ± 1.0% versus 0.6 ± 0.1 and 96.3% ± 1.2%, and 0.5 ± 0.1 and 96.8% ± 0.8%, respectively, on the test set. The multi-scale recurrent attention-guided model also demonstrated an improvement in terms of F1-score and MCC (97.5 %± 0.6% and 0.97 ± 0.01) compared to the multi-scaled model without attention (96.4% ± 0.9% and 0.95 ± 0.01) and with CBAM attention (96.7% ± 0.6% and 0.96 ± 0.01). The confusion matrix results obtained for different

models (Figure 2.5) are also in agreement with the accuracy, loss, F1-score and MCC results and demonstrate a better performance of the proposed model for all classes.

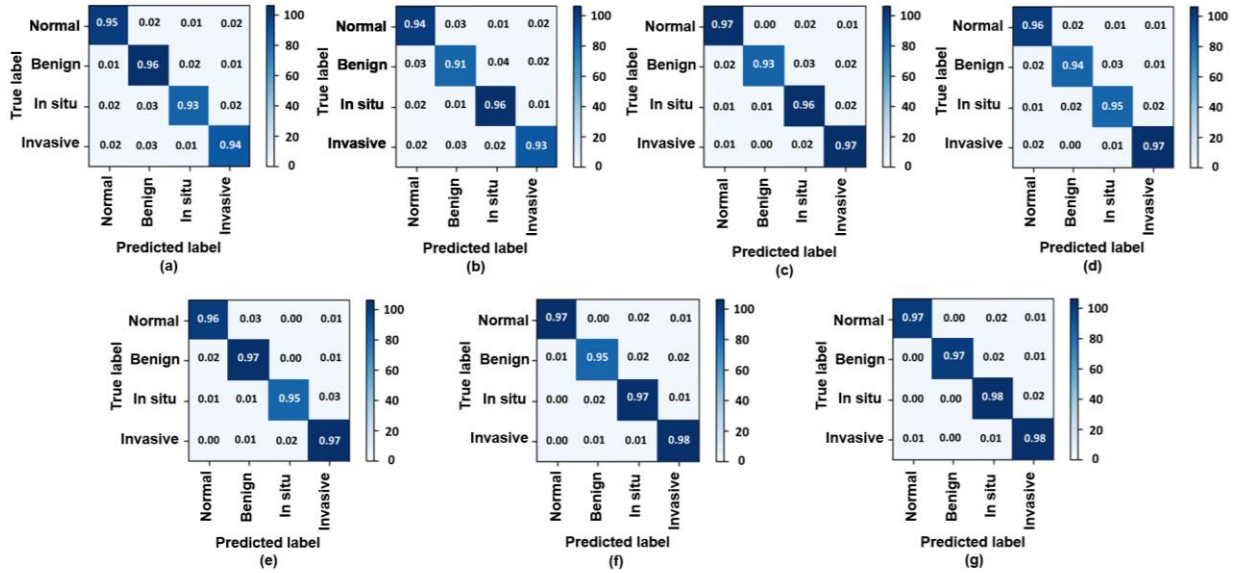


Figure 2.5. Confusion matrices summarizing the classification performance of (a) High-magnification network, (b) Low-magnification network, (c) Low-magnification network + CBAM attention, (d) Low-magnification network + recurrent attention, (e) High-magnification + Low-magnification networks, (f) High-magnification + Low-magnification networks + CBAM attention, (g) High-magnification + Low-magnification networks + recurrent attention, on the test set.

Table 2.2 compares the performance of the proposed model with those of the previous models developed and evaluated using the BACH dataset. The results demonstrate that the proposed model could outperform the previous models in classifying the histopathology images in term of validation and test accuracy. Similar classification accuracies obtained by the proposed model on the validation and test sets implies a high generalizability of the attention-guided multi-scale model on test sets.

Table 2.1. Performance of different models in classifying digital histopathology images of breast tissue specimens. Mean and standard deviation are given for accuracy, loss, F1-score and Matthews correlation coefficient (MCC) of each model in five experiments with different training and validation sets. Tr: Train, Val: Validation, Te: Test, Acc: Accuracy.

	Tr Acc (%)	Val Acc (%)	Te Acc (%)	Tr Loss	Val Loss	Te Loss
High-magnification Network	95.8 ± 1.1	94.4 ± 0.6	94.5 ± 0.6	$(1.2 \pm 0.4) \times 10^{-3}$	0.5 ± 0.2	0.7 ± 0.3
Low-magnification Network	95.2 ± 1.3	95.3 ± 0.9	93.5 ± 1.0	$(1.0 \pm 0.2) \times 10^{-3}$	0.6 ± 0.2	0.8 ± 0.2
Low-magnification Network + CBAM Attention	97.1 ± 1.2	96.7 ± 0.9	95.8 ± 0.4	$(4.2 \pm 1.1) \times 10^{-4}$	0.5 ± 0.1	0.5 ± 0.2
Low-magnification Network + Recurrent Attention	96.8 ± 1.1	96.5 ± 1.0	95.5 ± 0.8	$(4.8 \pm 1.6) \times 10^{-4}$	0.4 ± 0.1	0.5 ± 0.1
High-magnification + Low-magnification Networks	97.8 ± 1.1	96.6 ± 1.1	96.3 ± 1.2	$(1.2 \pm 0.3) \times 10^{-3}$	0.5 ± 0.1	0.6 ± 0.1
High-magnification + Low-magnification Networks + CBAM Attention	97.9 ± 0.9	97.3 ± 1.2	96.8 ± 0.8	$(2.3 \pm 1.4) \times 10^{-4}$	0.3 ± 0.1	0.5 ± 0.1
High-magnification + Low-magnification Networks + Recurrent Attention	98.2 ± 0.8	97.8 ± 0.7	97.5 ± 1.0	$(7.7 \pm 1.3) \times 10^{-5}$	0.1 ± 0.0	0.4 ± 0.1

	Tr F_1-Score	Val F_1-Score	Te F_1-Score	Tr MCC	Val MCC	Te MCC
High-magnification Network	95.5 ± 0.9	94.2 ± 0.8	94.4 ± 0.7	0.94 ± 0.03	0.93 ± 0.04	0.93 ± 0.02
Low-magnification Network	94.9 ± 1.0	95.0 ± 0.7	93.5 ± 0.9	0.94 ± 0.02	0.93 ± 0.02	0.91 ± 0.04
Low-magnification Network + CBAM Attention	96.9 ± 0.9	96.6 ± 0.6	95.7 ± 0.4	0.95 ± 0.02	0.94 ± 0.03	0.94 ± 0.02
Low-magnification Network + Recurrent Attention	96.7 ± 0.8	96.4 ± 0.8	95.6 ± 0.7	0.96 ± 0.01	0.94 ± 0.03	0.94 ± 0.01
High-magnification + Low-magnification Networks	97.5 ± 1.2	96.4 ± 0.9	96.4 ± 0.9	0.96 ± 0.02	0.95 ± 0.02	0.95 ± 0.01
High-magnification + Low-magnification Networks + CBAM Attention	97.6 ± 0.8	97.3 ± 0.8	96.7 ± 0.6	0.97 ± 0.01	0.96 ± 0.01	0.96 ± 0.01
High-magnification + Low-magnification Networks + Recurrent Attention	98.0 ± 1.1	97.7 ± 0.6	97.5 ± 0.8	0.97 ± 0.02	0.97 ± 0.02	0.97 ± 0.01

Table 2.2. Comparison of the results of the proposed model with the previous models evaluated on the BACH dataset. Val: validation, Te: test, Acc: accuracy.

Models	Val Acc	Te Acc
Nazari *et al*. [124]	95%	-
Golatkar *et al.* [127]	-	85%
Zhu *et al.* [125]	-	86.6%
Vizcarra *et al.* [128]	-	92%
Proposed Model	97.8%	97.5%

2.4 Discussion and Conclusion

In this chapter, a multi-scale attention-guided model was proposed and investigated for classification of breast cancer tissue using histopathology images. In this model, two parallel CCNs process the images at high and low magnification with a recurrent attention mechanism, and an MLP network fuses the network decisions for a final characterization. The high-magnification network processes the image patches to extract local features for tissue characterization. The low-magnification network, on the other hand, explores the entire image to derive global features and is followed by a recurrent attention model to put more focus on the relevant regions of images in characterizing samples. The fusion of the class probabilities estimated by these two networks enriches the model with decisions based on both the local and global features and permits a tissue characterization with high confidence.

The results demonstrate that the proposed multi-scale strategy improves the performance of the model for tissue characterization. Specifically, the multi-scale model demonstrated a higher accuracy and a lower loss in classifying the images. Another contribution of this chapter is to incorporate a recurrent attention mechanism into the histopathology image analysis framework. The attention mechanism provides a dynamic representation of the relevant regions of an image for classification in terms of a context vector in which the higher values represent a higher relevance.

The recurrent attention model iteratively learns to give higher weights to the features corresponding to the more informative regions of the image for characterization. The results obtained in this study demonstrate that incorporating the attention mechanism into the models improves their performance in classifying the histopathology images. Specifically, both the low- and multi-scale models with attention outperformed their counterparts with no attention mechanism in terms of loss and accuracy in tissue characterization. The results demonstrate that a combination of high- and low-magnification networks empowered by attention mechanism makes the model more confident in classifying the histopathology images of different types that may include similar features, e.g., normal versus benign and *in situ* versus invasive carcinoma. Specifically, the attention-guided multi-scale model could correctly annotate the histopathology images that were misclassified by individual high- and low-magnification networks.

The performance of the proposed model was compared to the previous studies in which the same or a similar dataset has been applied. The best validation accuracy obtained in the studies by Araujo et al. [123] and Nazari et al. [124] is 77.8% and 95%, respectively. The test accuracy of the models proposed by Zhu et al. [125] and Golatkar et al. [127] on the BACH dataset is 86.6% and 85%, respectively. Vizcarra et al. could achieve a test accuracy of 92% by integrating the shallow and deep models [128]. The results obtained in this study for different models show improvement in the classification accuracy compared to the previous works. Specifically, the CBAM attention-guided low-magnification network, the multi-scale network with CBAM attention, and the recurrent attention-guided multi-scale network demonstrated a mean accuracy of 95.8%, 96.8%, and 97.5%, respectively, on a test set.

In Summary, this study demonstrate that an attention-guided multi-scale strategy can improve the accuracy of deep learning models for automatic tissue characterization using digital histopathology

images. The multi-scale strategy applied in this study is based on the fusion of class probabilities obtained from two independent CNN. Other strategies such as cascading CNNs in an end-to-end framework in which a low-magnification network processes the stitched features maps extracted form image patches at high magnification can also be considered for comparison. Such strategies have been planned to be investigated as a future work on larger datasets. In this study, the patches associated with specific histology types were formerly extracted from the whole slide images (WSIs). Automatic annotation of WSIs to different histology types can be investigated as an extension of this study. A multi-scale analysis along with an attention mechanism can be utilized to identify the regions associated with different histology types within the WSIs.

Chapter 3

A Cascaded Deep Learning Framework for Segmentation of Nuclei in Digital Histopathology Images*

3.1 Introduction

Early-stage diagnosis of breast cancer is effective in preserving treatment options, reducing costs, and improving the survival and quality of life for patients [4]. While standard clinical imaging such as mammography and ultrasound are widely used for breast cancer screening, a biopsy followed by histopathology analysis on the acquired specimens is the gold standard for definitive diagnosis of breast cancer [5]. Recent studies have been shown that quantitative features describing the morphology, distribution, and texture of the tumor nuclei on digital histopathology images can be used in machine learning frameworks for various diagnostic, tissue characterization and prognostic applications [6], [7]. Accurate nuclei segmentation is an essential step of extracting such features from the histopathology images [8], [9].

Recent advances in deep convolutional networks have led several researchers to propose data-driven frameworks for automated image segmentation [10]. The proposed techniques mostly comprise an encoder-decoder architecture [11]. The requirement for accurate segmentation in addition to the lack of labeled training data raises the importance of developing a task-specific

*A version of the material presented in this chapter has been published in Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Glasgow, UK. pp. 4764-4767, 2022.

framework for segmenting medical and biomedical images [12]. The U-Net architecture [13] that consists of two symmetric contracting and expanding paths, was specifically introduced for biomedical image segmentation. The proposed model could be trained with a relatively small training set, and it could outperform the other frameworks in the ISBI cell tracking challenge 2015 [14]. However, U-Net struggles with finding the exact border of nuclei, and in practice, the model cannot accurately segment the touching or very close nuclei.

The emergence of U-Net has opened pathways for variants of its architecture to enhance the performance of segmentation tasks in biomedical images. The attention U-Net was proposed to improve the segmentation performance by integrating grid-based attention gates (AGs) on top of U-Net architecture [15]. Models trained with the AGs learn to suppress unnecessary regions in an input image while highlighting important features for a particular task. Each skip connection's gating signal incorporates image features from multiple imaging scales to remove the requirement of having explicit tissue localization modules. The attention U-Net model has been originally proposed for segmenting pancreas in computed tomography (CT) abdominal images [15], but then adapted in other application including nuclei segmentation [16]. The U-Net++ which has a deeply supervised encoder-decoder architecture, is another proposed variant of U-Net to enhance the performance of biomedical image segmentation frameworks [17]. In the proposed model, the encoder and decoder sub-networks are linked through several nested dense skip connections. The redesigned skip paths aim to minimize the semantic gap between the feature maps of the encoder and decoder blocks in the network.

Due to the importance of nuclei segmentation in digital histopathology images, a dataset was introduced for the multi-organ nuclei segmentation (MoNuSeg) challenge in MICCAI 2018 [18]. Several studies have investigated various approaches to achieve the best result in this challenge. In [8], a contour-aware informative aggregation network (CIA-Net) was presented as an innovative deep

neural network with a hierarchical information aggregation module between two task-specific decoders. The proposed architecture could enhance the generalization ability of the framework in segmenting the unseen organ nuclei. In [150], a pretrained EfficientNet was applied as the backbone of U-Net++ architecture for the breast tumor nuclei segmentation. The adapted backbone resulted in having fewer parameters and outperformed previous convolutional neural networks (CNNs) in terms of efficiency and accuracy. In [151], a novel self-supervised learning framework was introduced that fully exploits the capacity of extensively employed CNNs. The proposed method entails two sub-tasks of scale-wise triplet learning and count ranking, which allow the networks to exploit prior knowledge about nucleus size and number to extract instance-aware feature representations from the raw information. Most of the previous studies have demonstrated suboptimal performance in finding the border of the nuclei accurately, and in practice, touching or very close nuclei are always challenging to be segmented.

In this chapter, a novel cascaded framework is investigated to overcome the limitations of previous models and improve the accuracy of nuclei segmentation in histopathology images. The framework consists of a weighted U-Net model followed by a U-Net architecture with a VGG16 backbone and a soft Dice loss function. The weighted pixel-wise masks were generated for the training data and utilized along with the binary masks in calculating the loss function for the weighted U-Net in order to train the model to classify the touching and very close nuclei more accurately. The obtained results demonstrate a considerably better performance of the proposed model compared to each of the single networks and the previously proposed frameworks.

3.2 Materials and Methods

3.2.1 Dataset

The proposed framework was pretrained using the Post-NAT-BRCA dataset collected from the cancer imaging archive (TCIA) [152], and trained and evaluated using the multi-organ nuclei segmentation (MoNuSeg) dataset [153]. All the histopathology images were H&E stained and scanned at 40× magnification. The Post-NAT-BRCA dataset includes histopathology images of surgical tissue specimens acquired from breast cancer patients following neoadjuvant therapy (NAT). The nuclei in these images were annotated manually using the Sedeen software (Pathcore, Toronto, Canada). The dataset contains 92 histopathology images of various size, with a total of 25,675 annotated nuclei. The MoNuSeg dataset consists of thirty training histopathology images and fourteen test images with a size of 1000 × 1000 pixels, collected from seven organs (*i.e.*, breast, liver, kidney, prostate, bladder, colon, stomach). The MoNuSeg training and test sets contain 21,623 and 7,223 annotated nuclei, respectively. In this study, the MoNuSeg training set was randomly split into the training (80%) and validation (20%) sets at patient level.

3.2.2 Data Preprocessing

The histopathology images (three-channel RGB) in the MoNuSeg dataset were zero padded to a size of 1024 × 1024 pixels (12 padded pixels from each side), and then patched to 16 non-overlapped patches of size 256 × 256 pixels. A total of 664 patches with the same size (256 × 256 pixels) were extracted from the Post-NAT-BRCA dataset for model pretraining.

The binary masks were generated for each image patch. The weighted masks were only generated for the patches of the training set. The weight of each pixel (x) in the weighted mask was calculated using the Equation (3.1) [154].

$$w(x) = w_c(x) + w_0 \times exp\left(-\frac{\left(d_1(x) + d_2(x)\right)^2}{2\sigma^2}\right) \quad (3.1)$$

In Equation (3.1), w_c is the binary mask, d_1 is the distance to the border of the nearest nucleus, and d_2 is the distance to the border of the second nearest nucleus. w_0 and σ are constants that were selected empirically. Using this equation for generating the weighted masks, the pixels between closely adjacent nuclei are assigned higher weights in order to force the model to learn the separation in these regions with higher priority during the training process when a weighted loss function is applied.

3.2.3 Framework

Figure 3.1 shows the overview of the proposed framework. The cascaded framework consists of a U-Net model with a weighted pixel-wised loss function followed by a vanilla U-Net model with a VGG16 backbone and a soft Dice loss function [155], [156]. An effective nuclei segmentation model should learn the separation between adjacent nuclei while pixel-level accuracy is insufficient to evaluate the model on large histopathology images with many nuclei. This is because the number of pixels between the touching or very close nuclei are typically very few compared to the entire image and misclassifying them does not affect the pixel-level accuracy considerably during the model training. Generally, such separation can be accomplished by performing morphological operations on the images, but these operations are difficult to be incorporated into the model's learning process. As an alternative approach, the network could be forced to learn zone separations exclusively from the data. Following this approach, the pre-calculated weight maps were incorporated into the loss function to penalizes the loss in border areas between the touching or very close nuclei more than anywhere else (Equation (3.1)) [154]. Considering the constant parameters in Equation (3.1) (w_0, σ), there is a trade-off between the level of loss penalty for the border pixels between touching or very

close nuclei, and the priority of not missing the entire (small) nuclei. In this equation, a higher value of w_0 results in increasing the weight difference between the nuclei (foreground) and background regions within the border areas, while increasing the value of σ increases the difference in pixel weights based on the distance to the adjacent nuclei. On this basis, if the weighted U-Net is utilized with a high penalization weight for the areas near the touching or very close nuclei, the model would fail to detect many small nuclei or the center regions of larger objects, and in case a low penalization weight is applied, the model would fail in detecting the border of very close nuclei accurately. The cascaded segmentation framework in this study was proposed to address this issue by balancing the trade-off between detecting the small nuclei and segmenting the border of touching or very close nuclei accurately.

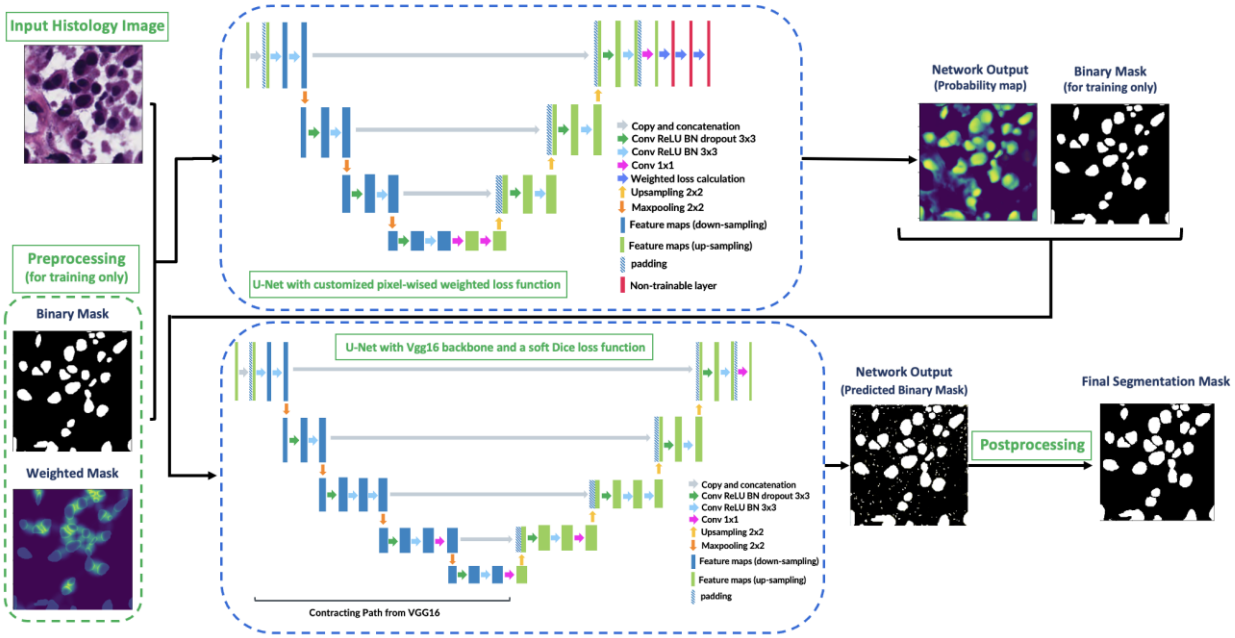


Figure 3.1. An overview of the proposed model consisting of a U-Net model with a weighted pixel-wised loss function followed by a vanilla U-Net model with a VGG16 backbone and a soft Dice loss function.

For training the weighted U-Net the loss function was modified from the regular cross entropy to a weighted cross entropy by multiplying $w(x)$ to it. In computing the custom loss function, the weights were applied to the log of the activation before calculating the pixel-wise sum. In the implementation step, as the model needs an ordinary loss function for training optimization, three non-trainable layers were added at the end of the weighted U-Net model for the training phase to embed the weighting of the log of the activation into the model and only leaving the summation step to the custom loss function. In this phase, the weighted U-Net was fed by the training images along with the corresponding binary and weighted masks and the segmentation probability maps were generated as the output of the model. The probability maps along with the binary masks were passed to the second component of the framework, *i.e.*, the vanilla U-Net model with a VGG166 backbone and a soft Dice loss function. The reason for adding the VGG16 backbone is to reduce the number of network parameters for a better generalization. The choice of a smaller convolution kernel (*i.e.*, 3×3) is indeed a key to VGG's remarkable accomplishment. Compared to other networks such as ResNet [157] or AlexNet [158], when acquiring the same receptive field, a smaller convolution kernel can not only use less computation and provide more non-linearity, but also make the model more powerful in fitting ability [155]. The choice of the soft Dice loss function potentially results in penalizing low-confidence predictions of the nuclei, since the ground truth is binary in this model and only the pixels belong to the nuclei are considered in the loss calculation.

The proposed approach needs calculation of weighted masks for each input image, while in the test phase this information is not available. Considering that the last layers of the weighted U-Net model are non-trainable and only used for loss calculation, we address this issue by initializing two different models for the training and test phases. Specifically, the training model applied both the weighted and binary masks as the ground truth for the weighted loss calculation, whereas the test

model only applied the binary masks for model evaluation on the unseen samples using the output of the layer before the non-trainable layers.

3.2.4 Postprocessing

Two morphological operations of erosion and dilation with the kernel size of five pixels were applied on the predicted segmentation masks to remove very small pixel clusters associated with noise., Considering that the histology slides were scanned at 40× magnification, the nuclei could not be apparent with such a small diameter on the images. It should be noted that since this post processing step only removes very small pixel clusters it would mainly improve the F1-score, with minimal effect on the Aggregated Jaccard Index (AJI) metric. At the end, the segmentation masks generated for the patches of each image were concatenated to reconstruct the whole image masks (1000 ×1000 pixels) for evaluation.

3.2.5 Evaluation

Performance of the proposed model was evaluated on the MuNoSeg test set and compared with other state-of-the-art segmentation frameworks. The evaluation was performed at both the object-level (nuclei detection) and pixel-level (accuracy of nuclei segmentation contours), using the F1-score and AJI metrics, respectively. In calculating the F1-score, the true positive (TP), false positive (FP) and false negative (FN) were determined as the number of correctly detected, wrongly detected, and undetected nuclei in each image, respectively.

3.3 Results

Figure 3.2 demonstrates weighted masks generated for a representative patch with different constant values. In this study, the values of $w_0 = 3$ and $\sigma = 10$ were selected empirically for these parameters

to force the weighed model to differentiate between the nuclei and the background regions within the border area of very close nuclei with high confidence. The generated weighted mask resulted in distinguishing the touching nuclei effectively, while missing some small nuclei and the center region of the larger nuclei. The probability maps generated by the first network were passed to the second network to enhance the final prediction mask.

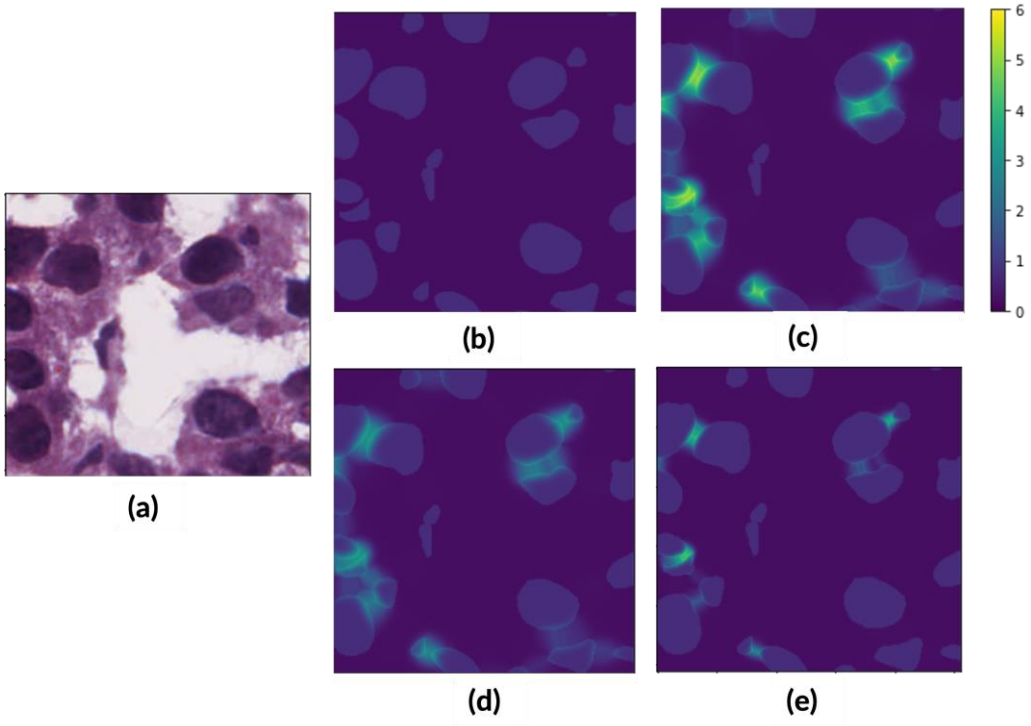


Figure 3.2. The binary and normalized weighted masks generated for a representative histopathology image patch (40× magnification) (a) with different parameters: (b) binary ground truth ($w_0 = 0$), (c) $w_0 = 5, \sigma = 10$, (d) $w_0 = 3, \sigma = 10$, and (e) $w_0 = 10, \sigma = 5$.

Figure 3.3 compares the output of different trained model on MoNuSeg test set with the ground truth. The segmentation mask generated by the attention U-Net (Figure 3.3(c, d)) and vanilla U-Net (Figure 3.3(e, f)) lacks accuracy in finding the nuclei boarders, particularly for the very close nuclei. The generated mask in Figure 3.3(g, h) shows the ability of the weighted U-Net model in separating the touching nuclei, however, the model missed to detect a few small nuclei and the center of few

larger nuclei. The cascaded model (Figure 3.3(i, j)) took advantage of both networks in finding the accurate border of very close nuclei, detecting small nuclei, and completely annotating the larger nuclei.

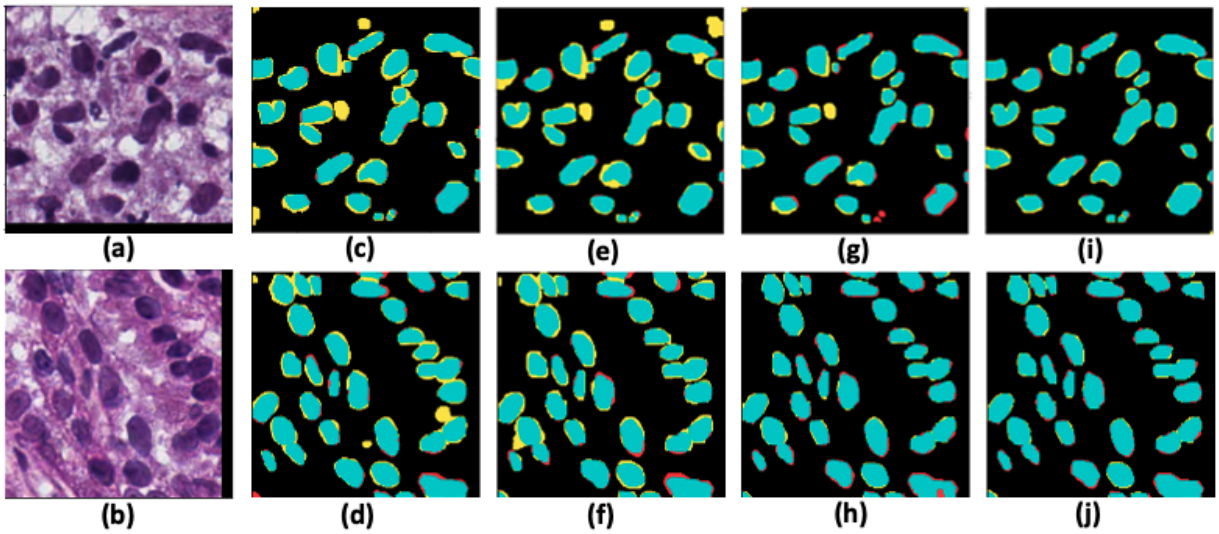


Figure 3.3. Comparison of the output segmentation masks of different networks with ground truth for two representative patches (top and bottom; 256×256 pixel): (a, b) histopathology images, (c, d) attention U-Net, (e, f) vanilla U-Net with the VGG16 backbone, (g, h) weighted U-Net, (i, j) cascaded framework. The pixel colors indicate true positive (green), true negative (black), false positive (yellow), and false negative (red).

Table 3.1 shows the performance of the proposed model on the MoNuSeg dataset compared to the attention U-net, vanilla U-Net with the VGG16 backbone and the weighted U-Net model in terms of AJI and F1-score metrics for the training, validation, and unseen test sets. The cascaded model could outperform each single model in all metrics.

Table 3.1. Performance of the proposed model on the MoNuSeg dataset

Model	Training		Validation		Test	
	AJI	**F1**	**AJI**	**F1**	**AJI**	**F1**
Attention U-Net	0.70	0.79	0.67	0.76	0.67	0.74
Vanilla U-Net with VGG16 backbone	0.70	0.80	0.68	0.77	0.68	0.76
Weighted U-Net	0.72	0.82	0.72	0.81	0.70	0.79
Cascaded Model	**0.74**	**0.84**	**0.73**	**0.84**	**0.72**	**0.83**

Table 3.2 demonstrates the AJI of the top five state-of-the-art models on the same dataset based on the MoNuSeg test set leaderboard, where the superior performance of the cascaded model in nuclei segmentation task is indicated.

Table 3.2. Comparison of the proposed model performance with other state-of-the-art models

Segmentation Model	AJI
Yunzhi [248]	0.68
Pku.hzq [248]	0.69
BUPT.J.LI [248]	0.69
CIA-Net [249]	0.70
U-Net++ [250]	0.70
SSL [251]	0.71
Proposed Cascaded Model	**0.72**

3.4 Discussion and Conclusion

In this chapter, a cascaded U-Net based framework was proposed for segmenting the nuclei in digital histopathology images accurately. A U-Net model with a pixel-wised weighted loss function followed by a vanilla U-Net with a VGG16 backbone have been adapted in the framework to annotate the touching nuclei accurately while keeping the number of correctly detected nuclei as high as possible. Three non-trainable layers have been added to the weighted U-Net model to calculate the pixel-wised weighted loss function during the training phase. The probability maps generated by the weighted U-Net are passed to the U-Net model with the VGG16 backbone and a

soft Dice loss function to generate the final masks. The weighted U-Net is responsible to remove the deceptive texture in the background which could be incorrectly considered as object (nuclei) and detect the border of the touching nuclei with high precision while the second U-Net is in charge of giving the same priority to all pixels to detect small nuclei and the missed (center) regions in the larger nuclei. The proposed framework could generate the binary mask on MoNuSeg test set with an AJI of 0.72, which shows a considerable improvement compared to the previous models. Whereas the cascaded model could outperform the former models, further investigation is still required with the state-of-the-art methods to evaluate and potentially improve the performance of the nuclei segmentation framework on larger datasets.

Chapter 4

Quantitative Digital Histopathology and Machine Learning to Predict Pathological Complete Response to Chemotherapy in Breast Cancer Patients using Pre-treatment Tumor Biopsies*

4.1 Introduction

The annual rate of breast cancer occurrence has increased by 0.3% in the United States [159], and associated with a risk of one in eight women who will develop breast cancer [160]. Approximately 5% to 20% of diagnoses include locally advanced breast cancer (LABC) [161], defined as stage III and a subgroup of stage IIB [162]. It typically comprises tumors larger than 5 cm, may involve skin or chest wall invasion, or with extensive axillary lymph node metastases [22], [163]. Due to the high risk of cancer progression, metastatic spread, and loco-regional recurrence, LABC is associated with a poorer prognosis compared to early-stage breast cancer [161], [163], [164]. The 10-year survival is approximately 44%, which is dependent on breast cancer subtype and response to therapies [165]. Definitive treatment for LABC includes neoadjuvant chemotherapy (NAC) followed by surgery [166]. However, only 10-30% of LABC patients demonstrate pathological complete response (pCR) to NAC, defined as a complete clearance of invasive carcinoma in the breast and regional lymph nodes [167]–[170]. Previous studies have shown a correlation between pCR and improved 5-year survival of up to 70% [171]–[174]. However, pathologic assessment is carried out after surgery,

*A version of the material presented in this chapter has been published in Scientific Reports. 12: 9690, 2022.

which limits the opportunity to adapt NAC treatments according to tumor response. Accordingly, early prediction of treatment response is needed to guide cancer therapy decisions based on individualized patient factors.

Several studies have investigated imaging biomarkers for early diagnosis, prognosis and the prediction of treatment responses in breast cancer [161], [163], [164]. However, histological examination remains the standard for cancer diagnosis, while genetic and immunohistochemical assessments may be used for prognosis and treatment outcome prediction [175], [176]. Machine learning (ML) algorithms, along with the development of whole slide imaging, has opened new research directions for early assessment of therapy response using quantitative digital histopathology. Digital pathology has the potential to yield large datasets from microscopic imaging and for automated analysis. Availability of such data permits development of computational tools and data-driven ML algorithms to process and interpret different types of tissue in high-resolution histopathology images and derive quantitative features for various diagnostic and prognostic applications [177]. Recent studies have demonstrated promising results for predicting cancer treatment outcome and recurrence using a combination of quantitative imaging (radiomic) and digital histopathology (pathomic) features [178], [179].

The objective of the study presented in this chapter is to investigate ML methods coupled with quantitative digital histopathology to predict pCR to neoadjuvant chemotherapy in breast cancer patients using pre-treatment biopsy specimens.

4.2 Materials and Methods

4.2.1 Study Protocol and Data Acquisition

This investigation was a single institution, retrospective study. Ethical approval was obtained from the institutional ethics review board (IRB) at Sunnybrook Health Sciences Centre and York University, Toronto, Canada, prior to data collection and analysis; research was conducted in accordance with the Declaration of Helsinki. As this was a retrospective non-interventional study, a consent waiver was obtained from the IRB under the provision of the Canadian Tri-Council Policy Statement 2 (TCPS; 2018) Articles 3.1-3.5 and 3.7A (*i.e.*, Ethical Conduct for Research Involving Humans). All study data were anonymized; specifically, patient identifiers were removed from each sample prior to analysis. Patients were included in the study based on the following inclusion criteria: confirmed diagnosis of invasive breast cancer, age (18+), and undergoing Anthracycline and/or Taxane based neoadjuvant chemotherapy followed by surgery. There were 149 patients included in the study. All patients had a breast core needle biopsy before NAC with a pathological review as part of their standard of care. Clinicopathological and imaging information were collected for all patients. Clinical data included patient age, menopausal status (pre/post), clinical tumor size (largest radiologically reported dimension from either mammogram ultrasound or magnetic resonance imaging; mm), histological type (ductal versus lobular carcinoma), Nottingham grade (G1/G2/G3), and the presence or absence of inflammatory cancer (defined as breast carcinoma with dermal lymphatic invasion). Estrogen receptor (ER) status (+/-), progesterone receptor (PR) status (+/-), human epidermal growth factor receptor-2 (HER2) status (+/-) were also obtained for all patients.

Treatment response endpoints were evaluated after surgery and classified into pathological complete response (pCR) versus pathological non-complete response (non-pCR), as ground truth labels for subsequent modelling. A standard assessment method using the residual cancer burden

index (RCBI) [180] was employed for ground truth labeling of response. An RCBI score of 0 (*i.e.,* pCR) was defined as the absence of residual invasive and *nodal* disease [181]. Patients who demonstrated residual disease were classified as non-pCR (*i.e.,* RCBI>0). All pathology reviews (pre-treatment and post-surgery histopathology) were evaluated by a board-certified breast pathologist as part of the patient's standard of care. Similarly, radiological reporting was carried out at the time of diagnosis by board-certified breast radiologists. The patients were randomly partitioned into a training (75%; n=111 patients) and an unseen test set (25%; n=38 patients). The training set was used for feature reduction/selection and development of predictive models (described below). The test set was used to evaluate the performance of the predictive models independently.

4.2.2 Core Biopsy Sample Preparation

Formalin-fixed paraffin embedded (FFPE) blocks containing core biopsy specimens obtained from each patient at pre-treatment were microtomed into 4 μm sections. Specimens were prepared onto glass slides and stained with hematoxylin & eosin (H&E). The slides were digitized into whole slide images (WSI) using a TissueScope LE digital pathology image scanner (Huron Digital Pathology Inc, St. Jacobs, Canada) at 40× magnification. All WSI were reviewed to ensure image integrity before image processing and analysis. If any image was distorted, blurry, or contained occlusions, the associated slide was re-imaged.

4.2.3 Preprocessing of Histopathology Images and Pathomic Feature Extraction

Figure 4.1 shows the overview of proposed pathomic framework for NAC response prediction. The tumor regions were annotated on the WSIs by an expert pathologist using the Sedeen software package[182]. The pre-processing steps were performed on three-channel RGB images. The tumor region annotations were preprocessed to extract non-overlapping tiles with a size of 768×768 pixels

by including tumor margins, when required. From the extracted tiles only the ones with more than 50% tumor tissue and less than 10% white background were retained for analysis (Figure 4.2(a) and (b)). In total, 647,328 image tiles were included in this study (484,071 and 163,257 tiles in the training and test sets, respectively). A pre-trained weighted U-Net based model was utilized to segment the nuclei in each tile accurately [183]. The histopathology images from the Cancer Imaging Archive (TCIA) and the Multi-Organ Nucleus Segmentation (MoNuSeg) datasets were used to train the model [184][185]. Each tile was patched to 256×256 pixel patches with 128 pixels overlap between the adjacent patches. After segmenting the nuclei, the patches were merged by averaging over the output probability map of the segmentation model within the overlapped regions. The binary nuclei mask for each tile was generated by thresholding the associated averaged probability map with a threshold level of 0.5. The detected nuclei with less than 50 pixels were eliminated in the generated masks based on the empirical observation that the actual nuclei cannot include less than 50 pixels on the histopathology images acquired at 40× magnification. Figure 4.2(c) and (d) shows a tile extracted from the tumor region of a representative WSI and the binary nuclei mask generated for it.

Using the HistomicsTK [186] and PyRadiomics [187] open-source packages, 549 pathomic features were extracted from each image tile for analysis. The features were representative of five categories: nuclear morphology and Fourier shape descriptors (16 features) [71], nuclear intensity and gradient features (20 features) [188], first- and second-order texture features (93 features) [189], graph-based features (49 features) [74] and wavelet features consisting of intensity, gradient and texture features extracted from wavelet filtered images (371 features) [190]. The morphological features and Fourier shape descriptors as well as the graph-based features were derived using the binary nuclei masks. The binary masks and the gray-scale image tiles were used

to calculate the nuclear intensity, gradient, texture, and wavelet features. The extracted features were averaged over all image tiles associated with each WSI to obtain the overall features for each patient. The number of nuclei in each image tile was applied to derive a weighting factor for the tile in calculating the averaged features.

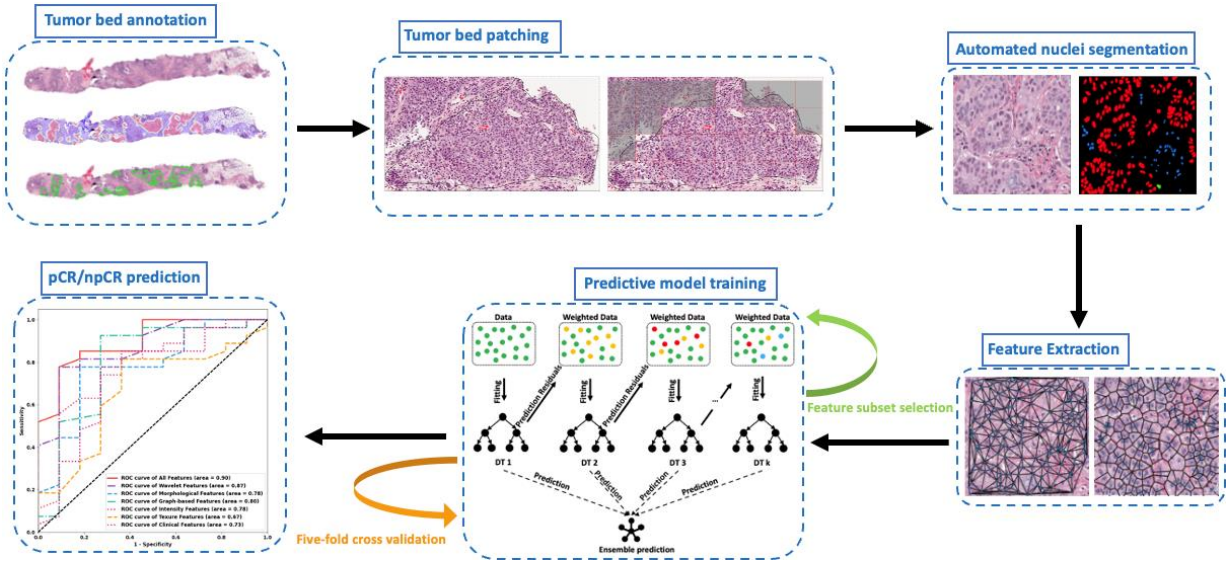


Figure 4.1. Overview of proposed pathomic framework for NAC response prediction. Annotated tumor beds were patched to 768×768 pixel tiles, and the pathomic features were extracted from the automatically segmented nuclei. A GBM model was adopted to select the optimal feature set, and subsequently build a predictive model that was evaluated on an independent test set.

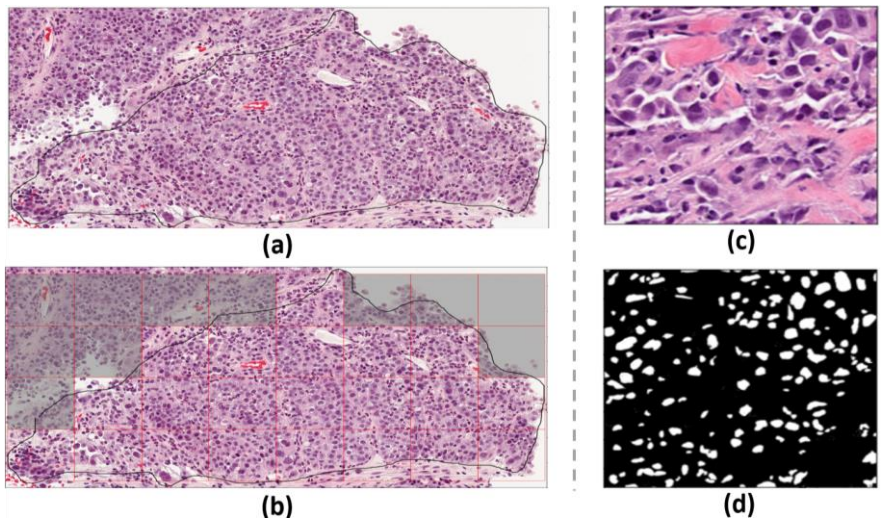


Figure 4.2. (a) An example of tumor bed annotation on a representative segment of the core on WSI at 40× (2188 × 4124 pixels), (b) the non-overlapping 768 × 768 pixel tiles extracted from the tumor region with the excluded tiles shaded (less than 50% tumor or more than 10% white background), (c) an extracted tile with 100% tumor tissue, and (d) the generated binary mask of the nuclei in the tile.

4.2.4 Feature Reduction/Selection and Tumor Response Prediction

The clinical and pathomic features were analyzed through a feature reduction and selection process on the training set to develop optimal biomarkers for NAC response prediction. Seven different experiments were conducted to analyze different feature subsets including the clinical, morphological, intensity-based, texture, graph-based and the wavelet feature subsets, in addition to a union of all feature subsets as the initial feature set. All the features were normalized to scale between zero and one before the analysis. A gradient boosting machine (GBM) with decision trees was trained as the classifier for response prediction in each experiment to calculate the contribution of each feature in the associated feature subset to the prediction model based on the importance gain score [191].The first few features with highest contribution to the model that demonstrated a meaningful difference in the importance gain score compared to rest of the features were selected in each experiment and included in the biomarker.

The GBM model was adapted to develop a predictive model of NAC response using the optimum biomarker in each experiment. A five-fold cross validation on the training set was used with area under the receiver operating characteristic (ROC) curve (AUC) as the criteria to optimize the model hyperparameters. To address the imbalance issue of the dataset, the training samples of the minority class (pCR) in each round were oversampled to a double number using the SMOTE method [192]. The final predictive model was developed using the entire training set with oversampled minority class, a learning rate 0f 0.1, a maximum depth of 10, and 2000 estimators. The performance of the predictive model with the optimal biomarker was subsequently evaluated on the independent test set using accuracy, sensitivity, specificity, and AUC. A threshold value of 0.5 was used as the cut-off to calculate the sensitivity and specificity.

4.3 Results

Table 4.1 shows clinical and pathological characteristics of the patients in the training and test sets. Among the 149 patients, 57.7%, 55.7%, and 44.3% had tumors with an ER+, PR+, and HER2+ receptor status, respectively. A majority of the patients (n = 123) were diagnosed with invasive ductal carcinoma, and a smaller proportion (n = 26) with invasive lobular carcinoma. The patients had an average initial tumor sizes of 46.4 ± 27.1 mm. Pathologic assessment after surgery demonstrated 34% (n = 50) of patients achieved a pCR; whereas 66% (n = 99) were non-pCR. The patients in the training and test set had similar statistics in terms of clinical and pathological characteristics, and similar proportions of patients with pCR and non-pCR were randomly included in both sets.

Table 4.1. Demographic and clinical information of the patients involved in the study. The distribution of each variable was compared between the training and test sets using the Pearson's chi-squared homogeneity test for categorical variables and the t-test for continuous variables; the p-values are reported in the last column.

Clinicopathologic Characteristics	**Count (%)**				**p-value**
	Train (n=111)		**Test (n=38)**		
Patient Demographics	**pCR (n=39)**	**non-pCR (n=72)**	**pCR (n=11)**	**non-pCR (n=27)**	p = 0.49
Median Age (Years)	48.4	52.2	50.6	52.1	p = 0.33
Menopausal Status					
Pre/Peri-menopausal	22 (56%)	36 (50%)	6 (55%)	15 (56%)	p = 0.75
Post-menopausal	17 (44%)	36 (50%)	5 (45%)	12 (44%)	
Receptor Status					
ER Positive	14 (36%)	53 (74%)	2 (18%)	18 (67%)	p = 0.27
PR Positive	13 (33%)	43 (60%)	1 (9%)	16 (59%)	p = 0.11
HER2 Positive	27 (69%)	22 (31%)	7 (64%)	10 (37%)	p = 0.32
Histology					
Invasive Ductal Carcinoma	39 (100%)	63 (88%)	11 (100%)	10 (37%)	p = 0.49
Invasive Lobular Carcinoma	0(0%)	9 (9%)	0 (0%)	17 (63%)	
Nottingham Grade					
1	1 (2%)	3 (4%)	0 (0%)	0 (0%)	p = 0.50
2	9 (23%)	36 (50%)	2 (18%)	18 (67%)	
3	29 (75%)	33 (46%)	9 (82%)	9 (33%)	
Tumor Size					
Mean Tumor Size (mm; ± SD)	37.2 ± 21.2	50.9 ± 28.9	44.1 ± 25.8	48.9 ± 27.9	p = 0.78
Other Clinical Information					
Inflammatory Breast Cancer	4 (10%)	10 (14%)	0 (0%)	1 (4%)	p = 0.53

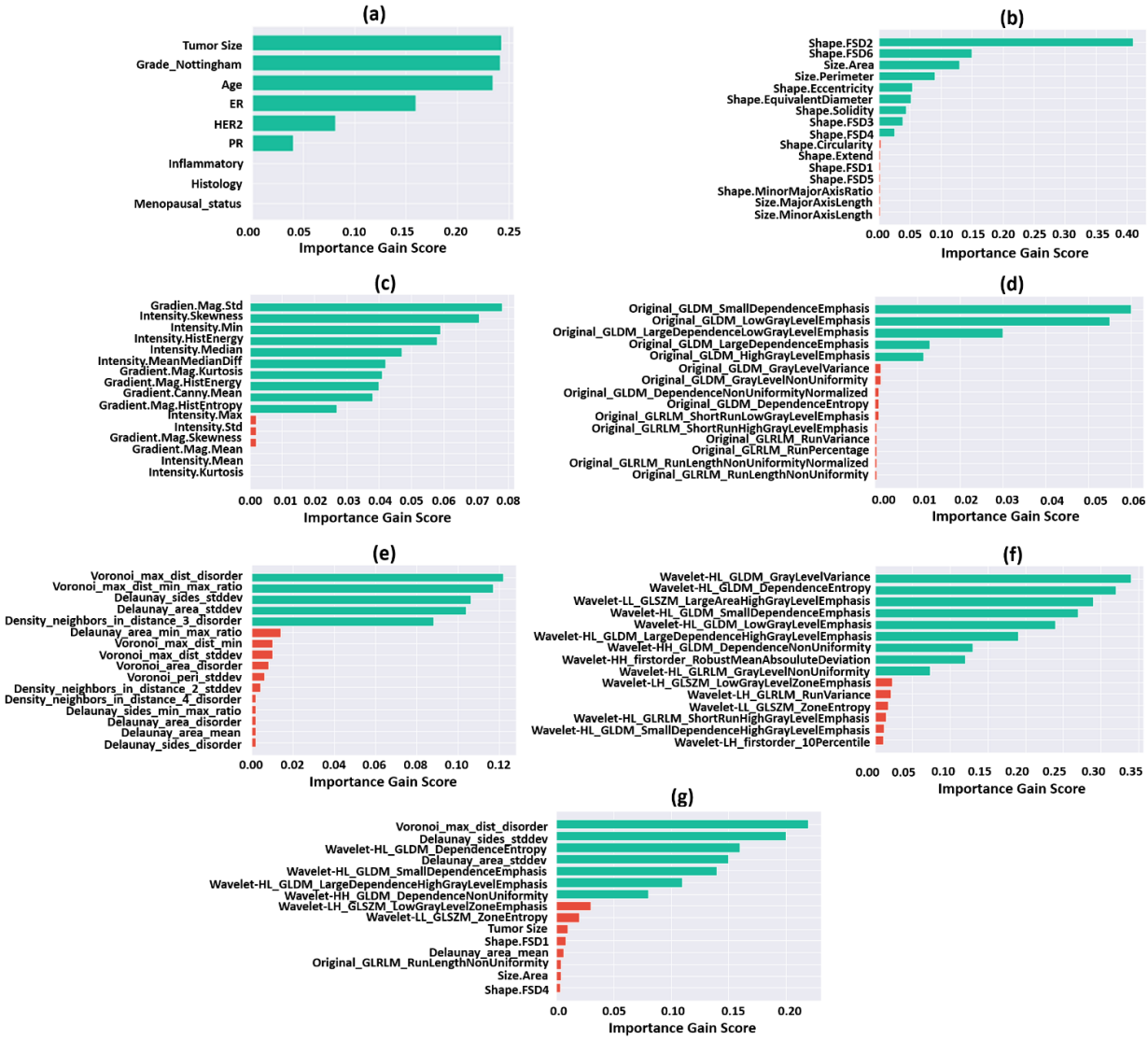


Figure 4.3. The importance gain score of the first 15 features with highest contribution to the predictive model for different feature subsets: (a) clinical, (b) morphological, (c) intensity-based, (d) texture, (e) graph-based, (f) wavelet, and (g) all features. The green bars are associated with the features included in the NAC response biomarker in each experiment.

Figure 4.3 shows the importance gain score of the first 15 features with the highest contribution to each predictive model. The best features were selected in each experiment based on the importance gain score as shown in the figure. In the first experiment, six clinical features including the tumor size, Nottingham grade, age, as well as the ER, HER2, and PR status demonstrated a non-zero importance gain score and were selected for model development. In the second to sixth experiments,

nine morphological features, ten intensity-based features, five texture features, five graph-based features, and nine wavelet features demonstrated a notable difference in their importance gain score compared to the rest of the features and were selected as the best features. Similarly, in the last (seventh) experiment that incorporated all the feature subsets (clinical and pathomic features), the first seven features were selected and included in the biomarker as their importance gain score demonstrated a considerable difference compared to rest of the features. The selected features in this biomarker only include the pathomic features, with four texture features derived from the wavelet-filtered images, and three graph-based features extracted from the tumor nuclei masks.

Figure 4.4 demonstrate the box plot of the selected features in different experiments for the pCR and non-pCR populations of the training set. The plots in Figure 4.4(a)-(f) show a relatively good separation in statistical distribution of the selected features between the two groups, particularly for those in the graph-based and wavelet feature subsets. The features selected among all feature subsets in the last experiment (Figure 4.4(g)) demonstrate a very good separation between the quartiles and median of feature values obtained for the two cohorts.

The evaluation results of the predictive models developed in different experiments have been presented in Table 4.2. The training and five-fold cross-validation accuracies of the developed models were very close together in each experiment and in the range of 72% to 85% and 71% to 84%, respectively. The perdition performance of the models on the independent test set has also been reported in the table. The test accuracy, sensitivity, and specificity of the models in different experiments were in the range of 71% to 84%, 70% to 85%, and 64% to 82%, respectively. The best results were obtained in the seventh experiment with the response biomarker consisting of the wavelet and graph-based features with a test accuracy of 84%, a sensitivity of 85%, and a specify of 82%. The ROC curves obtained on the independent test set is shown in Figure 4.5 for different

models. The AUC of the models ranged between 0.67 and 0.90 with the best result associated with the model developed using the wavelet and graph-based features.

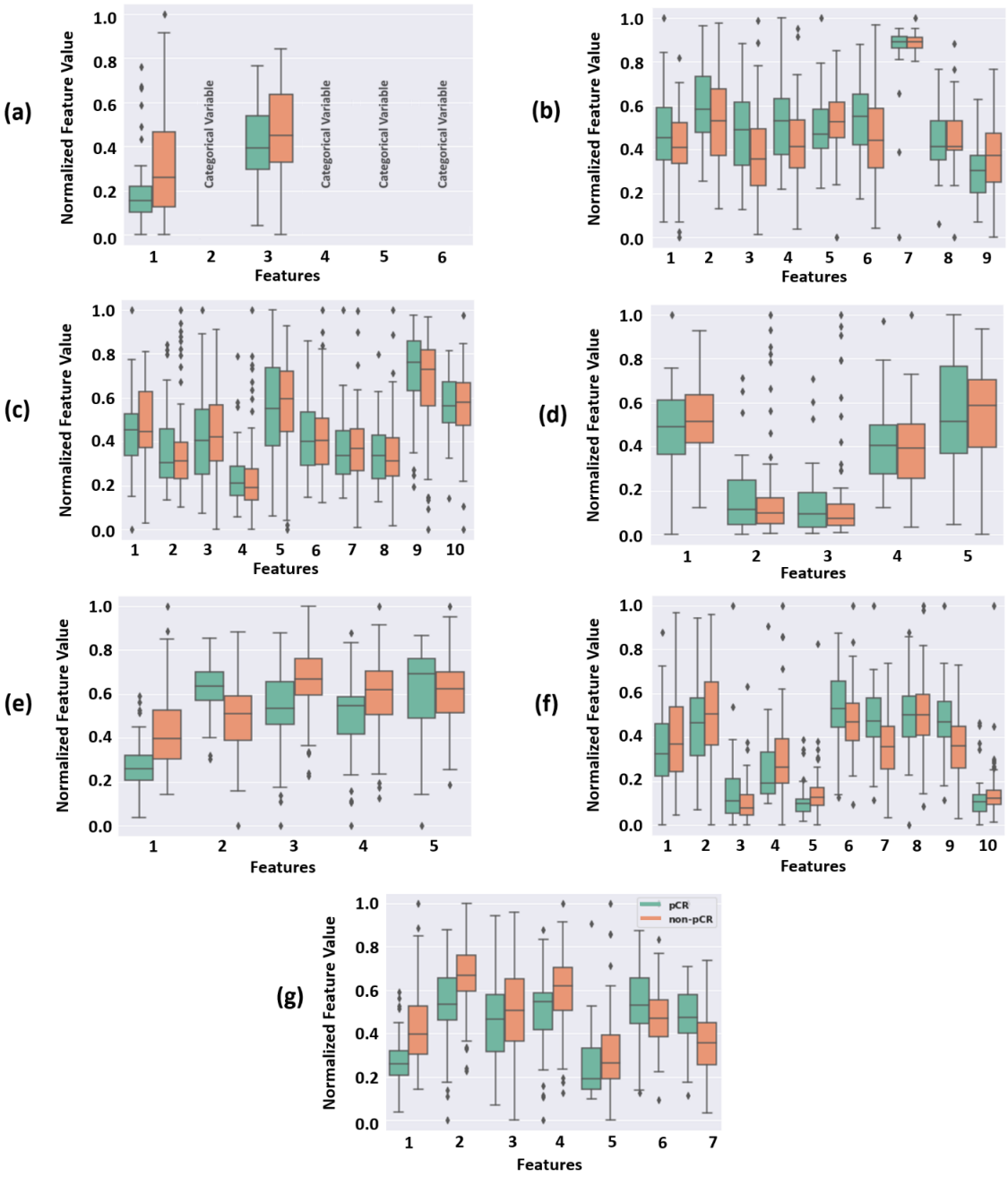


Figure 4.4. Box plots of the selected features for the pCR and non-pCR cohorts of the training set obtained in the seven experiments conducted using different feature subsets: (a) clinical, (b) morphological, (c) intensity-based, (d) texture, (e) graph-based, (f) wavelet, and (g) all features. The feature values are normalized in the range of 0 and 1. The order of features in each plot is the same as that of the associated plot in Figure 4.3.

Table 4.2. Results of NAC response prediction at pre-treatment using the clinicopathological and/or pathomic feature, on the training, validation and test sets. The features included in each optimal biomarker have been listed in Figure 4.3. For the validation set, the 95% confidence intervals are reported over the five folds of cross validation.

Features	Number of Features in Optimal Biomarker	Tr Acc	Val Acc	Val AUC	Val Sen	Val Spec	Te Acc	Te AUC	Te Sen	Te Spec
Clinical Features	6	0.72	0.71±0.02	0.74±0.04	0.70±0.04	0.73±0.03	0.71	0.73	0.70	0.73
Morphological Features	9	0.76	0.76±0.03	0.78±0.05	0.78±0.03	0.67±0.02	0.74	0.78	0.78	0.64
Intensity-based Features	10	0.78	0.75±0.02	0.77±0.03	0.77±0.02	0.69±0.02	0.74	0.78	0.78	0.64
Texture Features	5	0.76	0.75±0.01	0.70±0.03	0.78±0.02	0.68±0.03	0.74	0.67	0.78	0.64
Graph-based Features	5	0.75	0.76±0.01	0.79±0.02	0.79±0.03	0.72±0.03	0.76	0.80	0.78	0.73
Wavelet Features	9	0.82	0.83±0.02	0.84±0.03	0.80±0.04	**0.83±0.02**	0.82	0.87	0.81	**0.82**
All Features	7	**0.85**	**0.84±0.03**	**0.89±0.04**	**0.84±0.04**	0.82±0.02	**0.84**	**0.90**	**0.85**	**0.82**

Abbreviations: Acc = accuracy; AUC = area under the curve; Sen = sensitivity; Spec = specificity; Tr = train; Te = test; Val = validation

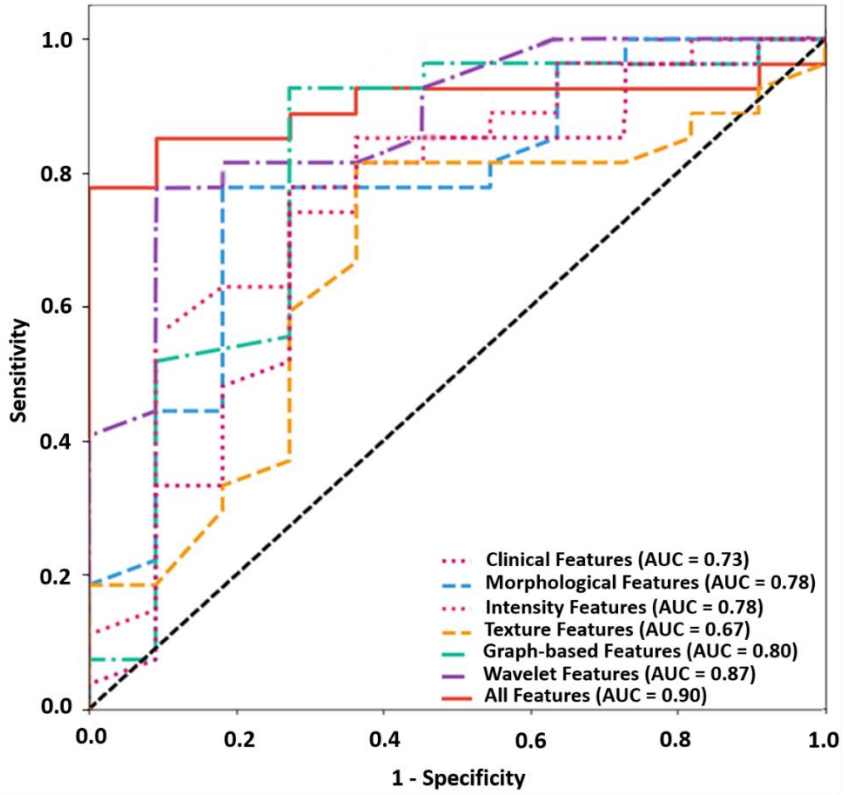


Figure 4.5. Receiver operating characteristic (ROC) curves on the independent test set for the predictive models developed with the selected features obtained in different experiments. In the last experiment and from all feature subsets, 7 wavelet and graph-based features were selected.

4.4 Discussion

In this study, a GBM multi-feature ML model was investigated with various sets of clinicopathological and quantitative pathomic features derived from pre-treatment core biopsy specimens to predict the therapy response of breast cancer patients undergoing NAC. Seven experiments were conducted to explore the efficacy of various feature subsets in predicting the therapy outcome. The results demonstrated superior performance of the wavelet and graph-based feature in the predictive modeling of NAC response at pre-treatment. The best results were obtained in the final experiment where all the clinical and pathomic features were included in the initial feature set. The response signature developed in this experiment consisted of seven features, including four wavelet and three graph-based features. Results of descriptive analysis demonstrated a promising separation among the quartiles and median of these features between the two response cohorts. The ML model developed using this biomarker predicted the NAC response of the patients in the independent set with a sensitivity, specificity, and AUC of 85%, 82% and 0.90, respectively. The multivariate GBM demonstrated that the non-linear combination of the selected pathomic features has a very good predictive ability for NAC response at pre-treatment

The results of experiments conducted in this study demonstrate that the pathomic features outperform the clinical variables in NAC response prediction. Whereas a few pathomic feature subsets including the morphological, intensity-based and texture features could differentiate the response cohorts with slightly better prediction accuracy compared to the clinical features, the wavelet and graph-based features demonstrated a considerably better efficacy. This observation was further confirmed in the last experiment where among all features, only the pathomic features from these two subsets were selected in the optimal biomarker with no feature from the clinicopathologic subset. Among the seven features included in this biomarker, the three graph-based features

characterize the variations in spatial distribution of the intra-tumor nuclei with different measures. Specifically, the Voronoi_Max_Distance_Disorder feature measures the variations in maximum distance within polygons in Voronoi diagram of the nuclei, while Delaunay_Sides_Stddev and Delaunay_Area_Stddev measure the variations in the side length and area of triangles in Delaunay triangulation graph generated using the Voronoi partitions [193]. The wavelet features selected in the biomarker, on the other hand, characterize the spatial heterogeneity within the tumor nuclei by quantifying the gray-level dependencies in the associated wavelet-filtered images. Specifically, the Wavelet_HL_GLDM_DE, Wavelet_HL_GLDM_SDE, Wavelet_HL_GLDM_LDHGLE, Wavelet_HH_GLDM_DNU features measure the entropy in gray-level intensity dependence, texture homogeneity, distribution of close similarities with higher intensity values, and the uniformity of intensity values within the nuclei [194].

Imaging features confer information about cell-cell interactions and activity within the tumor microenvironment [195]. Determining actionable biomarker signatures, derived by mapping tumor subcomponents and characterizing the biological heterogeneity has the potential to improve diagnosis and response-guided treatment strategies. Previous studies have investigated the efficacy of the pathomic features in conjunction with the genomic features for other applications of cancer diagnosis and prognosis [196], [197]. The findings of those studies are in agreement with the observations in this study. The genomic data, however, are not routinely acquired for LABC. Therefore, incorporating these parameters in predictive modeling of therapy outcome requires extra data acquisition and processing that may not be always feasible. A number of other studies have explored the performance of quantitative imaging data (radiomic features) acquired at early stage of diagnosis for NAC response prediction [198][199][200]. The observations of those studies confirm that the breast cancer characteristics such as intra-tumor heterogeneity quantified using pretreatment

imaging can reasonably be correlated to the NAC outcome. One limitation associated with predictive modeling using the imaging-based features is that the performance of such systems could possibly be affected by imaging acquisition protocols including variations in resolution, magnification, and gain parameters [201]. A number of previous studies have focused on post-treatment nonsurgical techniques including biopsy and imaging to detect residual cancer in the breast or axilla after NAC [202], [203]. Specifically, pre-surgical vacuum-assisted biopsy (VAB) coupled with machine learning methods have been investigated to identify patients with pCR to NAC who may not need to undergo surgery. The results demonstrate that combining the clinical, imaging and VAB variables integrated with machine learning models can improve the performance in pre-surgical NAC response identification. Whereas applying such methods at post-treatment may spare the patients with pCR from an unnecessary mastectomy or lumpectomy, they cannot facilitate treatment adjustments or switching to alternative treatments for non-responders.

The results of this study were obtained using a relatively small dataset (n = 149) acquired from a single institution. A test set was randomly selected and kept completely unseen during the model development and tuning to assess the performance of the models independently. Whereas similar performance of the models on the validation and independent test sets can imply a good generalizability of the models on unseen samples, no external test set was available in this study to minimize the chance of bias in model evaluations. As such, to evaluate the robustness of the methods and assess the performance and applicability of the developed models in clinic rigorously, further investigations are required on larger cohorts of patients with multi-institutional data.

In conclusion, this study demonstrates a very good potential of hand-crafted pathomic features integrated with ML techniques in predicting the pathological response of breast cancer patient to NAC. The promising results obtained in this study is a step forward toward *a priori* chemotherapy

response prediction in high-risk breast cancer patients using smart quantitative histopathology methodologies at pre-treatment. Early prediction of NAC response permits timely therapy adjustment by oncologists or switching to more effective treatment for individual patients. A personalized oncology paradigm for breast cancer patients is expected to improve their overall therapy outcome and quality of life. The promising results obtained in this study pave the way for further investigations and encourage future studies to integrate more advanced ML methodologies including the end-to-end deep learning architectures with digital histopathology for NAC response prediction.

Chapter 5

A Hierarchical Self-Attention-Guided Deep Learning Framework to Predict Pathological Complete Response to Chemotherapy in Breast Cancer Patients using Pre-Treatment Tumor Biopsies*

5.1 Introduction

About 20% of breast cancer patients are diagnosed with locally advanced breast cancer (LABC) [204], [205]. LABC includes stage III and a subdivision of stage IIB breast cancer, with a tumor size of usually greater than 5 cm that may extend to the skin and/or chest wall, or with axillary lymph node involvement [22], [199], [206]. LABC patients typically have poorer prognosis compared to early-stage breast cancer due to high risk of cancer progression, local recurrence and metastasis [207]–[209]. LABC is currently treated with neoadjuvant chemotherapy (NAC) followed by surgery and, if needed, hormone therapy, targeted therapy, and/or adjuvant radiotherapy [210], [211]. Despite multimodal treatment plans, LABC patients still have low overall survival, highly dependent on tumor response to NAC [212]. Pathological response to NAC has shown high correlations with patient survival [169]. However, only up to about 30% of LABC patients present pathological complete response (pCR) to standard NAC [167], [213]. The definitive method for pathological assessment of response to NAC is through histopathology on surgical specimens at post-treatment.

*A version of the material presented in this chapter has been submitted for publication in Medical Physics, 2022.

However, the options for modifying NAC based on tumor response are limited at late stages during the treatment course. Predicting response to NAC either before or early after the treatment initiation could potentially improve the therapy outcomes through facilitating personalized patient care [214], [215].

Prior research has explored various quantitative imaging approaches for assessing chemotherapy response in breast cancer patients at the diagnosis stage or early after starting the treatment [215]–[218]. Quantitative biomarkers derived using different medical imaging modalities, including ultrasound, diffuse optical imaging, magnetic resonance imaging (MRI), and positron emission tomography (PET), have shown promise in characterizing breast cancer in terms of responsiveness to chemotherapy, particularly when coupled with machine learning (ML) models [219]–[222]. However, histopathological assessment remains the gold standard approach for cancer diagnosis and characterization in clinical practice. In Chapter 4 of this dissertation, digital histopathology images of diagnostic tumor biopsies have been applied for predicting pCR to NAC in breast cancer patients using a pathomic framework. The results show the potential of quantitative information derived from the pre-treatment digital pathology images for predicting NAC outcome. However, the hand-crafted pathomic features that are defined based on closed-form mathematical equations may not capture all the salient information from the digital pathology images.

Recent research has explored the efficacy of deep learning (DL) methods in various medical image analysis applications [104]. One superiority of these methods compared to conventional ML models is that the process of extracting hand-crafted features can be eliminated [105]. Previous studies have investigated the performance of various DL frameworks in analyzing histopathology images for tumor grading and subtyping, and predicting the patient survival [106]–[109]. These studies have demonstrated the potential of the DL models and specifically the deep convolutional networks in digital pathology image analysis. The size of whole slide images (WSIs), however, remains as a

constraint in training the DL models, due to memory limitations. Specifically, developing adequate DL models for analyzing the WSIs at high magnification is a challenging task as the size of the high-resolution WSIs can be as large as 150000 × 150000 pixels [223], [224]. This issue can possibly be addressed by resizing (downscaling) the WSIs and/or extracting small patches from them for analysis by the DL models. Downscaling reduces the image resolution and potentially the efficacy of the information derived by the model. The patching approach requires development of efficient strategies for fusing the patch-level information, while considering the global dependencies, to make relevant conclusions on the WSI level.

Several recent studies have focused on developing efficient DL models for image analysis and classification in computer vision applications. The recent state-of-the-art models are categorized into three general approaches. The first category includes the convolutional neural network (CNN) models. The Xception model [225], a well-known model in this group, is an extended version of the Inception-V3 architecture [226] in which the Inception modules have been replaced by depth-wise separable convolutions. This model could outperform the Inception-V3 model in classifying the ImageNet dataset [227]. The second category comprises the transformers [228] that utilize an encoder-decoder architecture with self-attention mechanism [229]. The self-attention mechanism in the transformers differentially weights the significance of each part of the input data for the target analysis. The recently introduced vision transformer (ViT) architecture [230] has demonstrated a high performance in extracting the global relations in the input images [231]. However, compared to the CNN-based models, this architecture requires larger amount of training data to yield an adequate generalizability [232]. The last category includes the networks that combine the convolutional layers with the attention mechanisms. The convolutional block attention module (CBAM) [133], a recently proposed attention mechanism, can be used to infer independent channel and spatial attention maps. The CoAtNet [233] is the most recent state-of-the-art architecture that

stacks depth-wise convolutional layers and the self-attention mechanism to improve the generalizability of the model [233].

In this chapter, a hierarchical attention-guided deep learning framework is proposed to predict pCR to NAC in breast cancer patients using digital histopathology images of pre-treatment biopsy specimens. The proposed model consists of three modules, including patch-level processing, tumor-level processing, and patient-level response prediction. The hierarchical flow presented in this study could overcome the difficulty of deriving the global relations between different tumor areas in high-resolution WSIs while utilizing the local information within the tumor regions in the analysis.

5.2 Materials and Methods

5.2.1 Dataset

This retrospective study was conducted in accordance with the institutional ethics review board (IRB) approval from Sunnybrook Health Sciences Centre, Toronto, Canada. The inclusion criteria for the study included: biopsy-confirmed diagnosis of invasive breast cancer, age (18+), and treatment with Anthracycline and/or Taxane based neoadjuvant chemotherapy followed by surgery (any type). Following these criteria, 207 patients were included in the study. The patients aged between $28-79$ years with a mean and standard deviation of 51.1 ± 10.4 years. The patients had an initial tumor size in the range of $1.1-15.5\ cm$ with a mean and standard deviation of $5.01 \pm 2.9\ cm$. The clinical N stage was 0 (no positive lymph nodes) for 24.6% ($n = 51$); 1 (1-3 positive lymph nodes) for 66.2% ($n = 137$); 2 (4-9 positive lymph nodes) for 8.2% ($n = 17$); and 3 ($\geq$10 positive lymph nodes) for 1% ($n = 2$) of the patients. Among 207 patients, 62.3%, 54.1%, 62.3% and 41.5% had tumors with an ER+, PR+, and HER2+ receptor status, respectively. Most of the patients ($n =$

192; 93%) had invasive ductal carcinoma (IDC), with a minority ($n = 15; 7\%$) diagnosed with invasive lobular carcinoma (ILC).

All patients had a breast core needle biopsy before NAC with a pathological review as part of their standard of care. The formalin-fixed paraffin embedded (FFPE) blocks containing core biopsy specimens were microtomed into 4 μm sections. The specimen sections were prepared onto histopathology slides and stained with hematoxylin and eosin (H&E). Digital histopathology images of the H&E-stained slides were acquired using a digital pathology imaging system (Huron Digital Pathology, St. Jacob's Canada). The images were acquired at 40× (pixel size: 2 μm) for all patients. The digital WSIs were manually reviewed for artifacts or occlusions within the specimen before analysis and any slides associated with a distorted or blurry image was re-imaged.

The treatment response was assessed for each patient after surgery and categorized into pathological complete response (pCR; *i.e.,* complete eradication of invasive tumor cells in the breast and lymph nodes) versus pathological non-complete response (non-pCR; *i.e.,* exhibiting residual cancer), as ground truth labels for evaluating the developed models. A standard assessment method using the residual cancer burden index (RCBI) was applied for assessment of treatment response. An RCBI score of 0 (*i.e.*, pCR) was defined as the absence of residual invasive and nodal disease [234]. Patients who demonstrated residual disease were classified as non-pCR (*i.e.,* RCBI > 0) [234]. All pathology reviews (pre-treatment and post-surgery histopathology) were performed by a board-certified breast pathologist as part of the patient's standard of care. The Pathological evaluations after surgery demonstrated 25.2% ($n = 52$) of patients with a pCR and 74.8% ($n = 155$) with a non-pCR.

5.2.2 Preprocessing and Dataset Splitting

Using the Sedeen software package [235], an expert pathologist annotated the tumor bed areas on the WSIs. The tumor bed annotations were pre-processed on the three-channel RGB images for patch extraction. Tumor margins were included, when required, to obtain non-overlapping patches with a size of 512×512 pixels (pixel size: 2 μm) from the annotated tumor regions. However, only the patches with more than 10% tumor tissue and less than 10% white background were retained for the study (Figure 5.1). In total, 1,733,421 patches were included in this study.

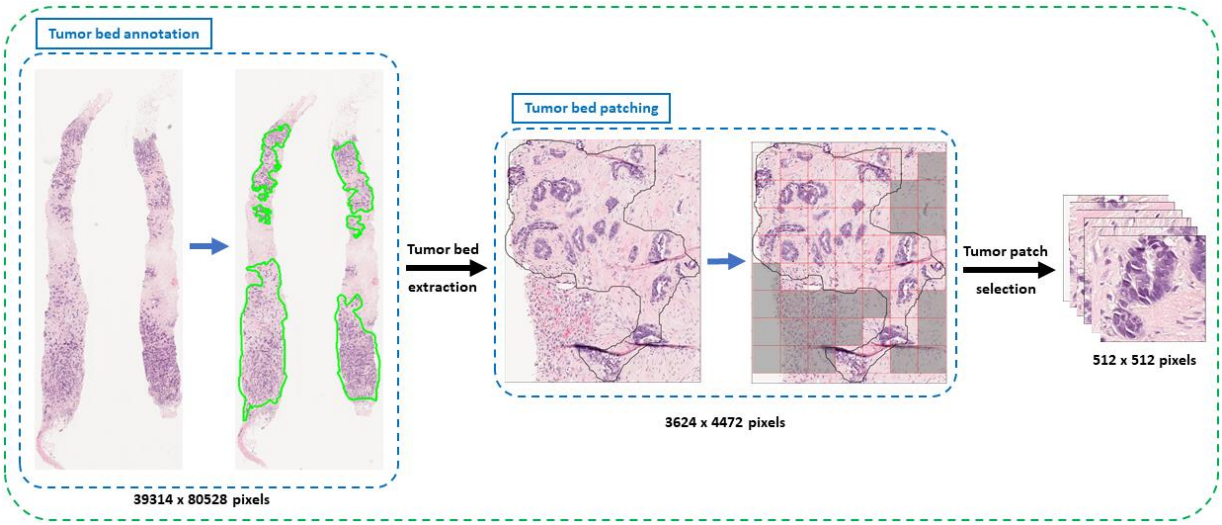
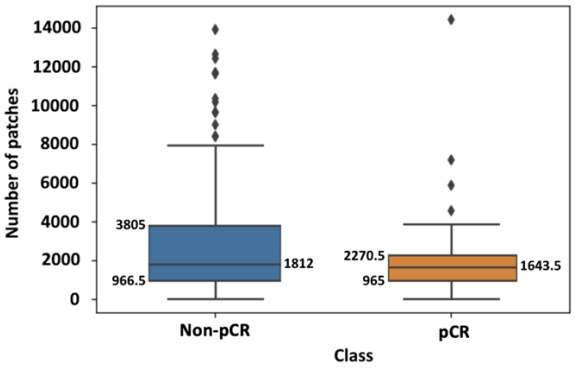


Figure 5.1. Overview of preprocessing steps. The tumor beds were annotated (green contours) in the WSI (40×) by an expert pathologist. Patches of size of 512× 512 pixels were extracted from the tumor beds. Patches with more than 10% tumor tissues and fewer than 10% white background were kept.

Out of the 207 patients in this study, 155 and 52 patients were labeled as the non-pCR and pCR, respectively. The number of patches for each patient varied from 30 to 14418 (median = 1755). The non-pCR and pCR classes included 1,423,210 and 310,211 patches, respectively. As such, the ratio of the pCR to non-pCR class at the patient- and patch-level was 25.2% and 17.9%, respectively. Figure 5.2(a) shows the distribution and quartiles of the number of patches per patient in the pCR and non-pCR classes of the dataset. A stratified random splitting approach was applied to split the

data at patient level into the training (70%; n = 144 with 9,430 annotated tumor beds and 1,559,784 patches) and test sets (30%; n = 63 with 3,574 annotated tumor beds and 173,637 patches). The stratified random splitting was performed considering the response label and the number of extracted patches for each patient. The first quartile, median and the third quartile of the number of patches in each response class (pCR and non-pCR) were used to stratify the patients into eight different bins for random sampling, as shown in Figure 5.2(b). A similar procedure was used to utilize five-fold cross-validation on the training set for optimizing the framework's hyperparameters (described further in Section 5.2.4). The training and test sets obtained using the stratified random splitting approach were assessed, using statistical tests, for possible inhomogeneities in terms of clinical feature distribution between the two sets. The Pearson's chi-squared homogeneity test was used for categorical variables. The continuous variables were assessed using the t-test. Table 1 presents the clinical information of the patients in the training and test sets, along with the statistical test results.



Class (pCR = 0, non-pCR = 1)	Criterion Discretized number of patches	Stratifying bin number: 4×Class + criterion
0	0 (< quartile1)	0
0	1 (>quartile1, <median)	1
0	2 (>median, <quartile3)	2
0	3 (>quartile3)	3
1	0 (< quartile1)	4
1	1 (>quartile1, <median)	5
1	2 (>median, <quartile3)	6
1	3 (>quartile3)	7

Figure 5.2. Stratified random splitting of data at patient level based on the number of patches in each class of the dataset. (a) Box plot presents the distribution and quartiles of the number of patches per patient in each class. (b) The criterion for stratification.

Table 5.1. Demographic and clinical characteristics of the patients. The distribution of each variable was compared between the training and test sets using the Pearson's chi-squared homogeneity test for categorical variables and the t-test for continuous variables; the p-values are reported in the last column.

Patient demographics and clinicopathological characteristics	**Count (%)**				**p-value**
	Training (n=144)		**Test (n=63)**		
	pCR (n=36)	**non-pCR (n=108)**	**pCR (n=16)**	**non-pCR (n=47)**	p = 0.91
Median Age (Years)	47.5	51	49.5	51	p = 0.61
Menopausal Status					
Pre/Peri-menopausal	15 (42%)	56 (52%)	9 (56%)	24 (51%)	p = 0.69
Post-menopausal	21 (58%)	52 (48%)	7 (44%)	23 (49%)	
Tumor Size					
Mean Tumor Size (mm; ± SD)	39.8 ± 21.1	55.8 ± 29.5	37.1 ± 23.9	49.6 ± 30.7	p = 0.22
Nodal Status (N Stage)					
No Positive Lymph Nodes (N0)	14 (39%)	22 (21%)	7 (44%)	8 (17%)	p = 0.59
1-3 Positive Lymph Nodes (N1)	21 (58%)	75 (69%)	8 (50%)	33 (70%)	
4-9 Positive Lymph Nodes (N2)	1 (3%)	10 (9%)	1 (6%)	5 (11%)	
≥10 Positive Lymph Nodes (N3)	0 (0%)	1 (1%)	0 (0%)	1 (2%)	
Receptor Status					
ER Positive	13 (36%)	79 (73%)	3 (19%)	34 (72%)	p = 0.81
PR Positive	12 (33%)	68 (63%)	2 (13%)	30 (64%)	p = 0.27
HER2 Positive	24 (67%)	38 (35%)	10 (63%)	14 (30%)	p = 0.51
Histology					
Invasive Ductal Carcinoma	36 (100%)	96 (89%)	16 (100%)	44 (94%)	p = 0.53
Invasive Lobular Carcinoma	0(0%)	12 (11%)	0 (0%)	3 (6%)	
Nottingham Grade					
1	1 (3%)	4 (4%)	0 (0%)	1 (2%)	p = 0.47
2	9 (25%)	39 (36%)	2 (13%)	22 (47%)	
3	26 (72%)	65 (60%)	14 (87%)	24 (51%)	
Other Clinical Information					
Inflammatory Breast Cancer	2 (6%)	11 (10%)	2 (13%)	5 (11%)	p = 0.64

5.2.3 Response Prediction Framework

Figure 5.3 demonstrates the scheme of the proposed hierarchical deep-learning framework for therapy response prediction. The patient-wise sampler in the framework addresses the unbalance issue of the training data at patient and patch levels. The sampler applies a weighted sampling strategy for under-sampling the majority class in the training set based on the total number of patches and the number of patients in each class using Equation (5.1).

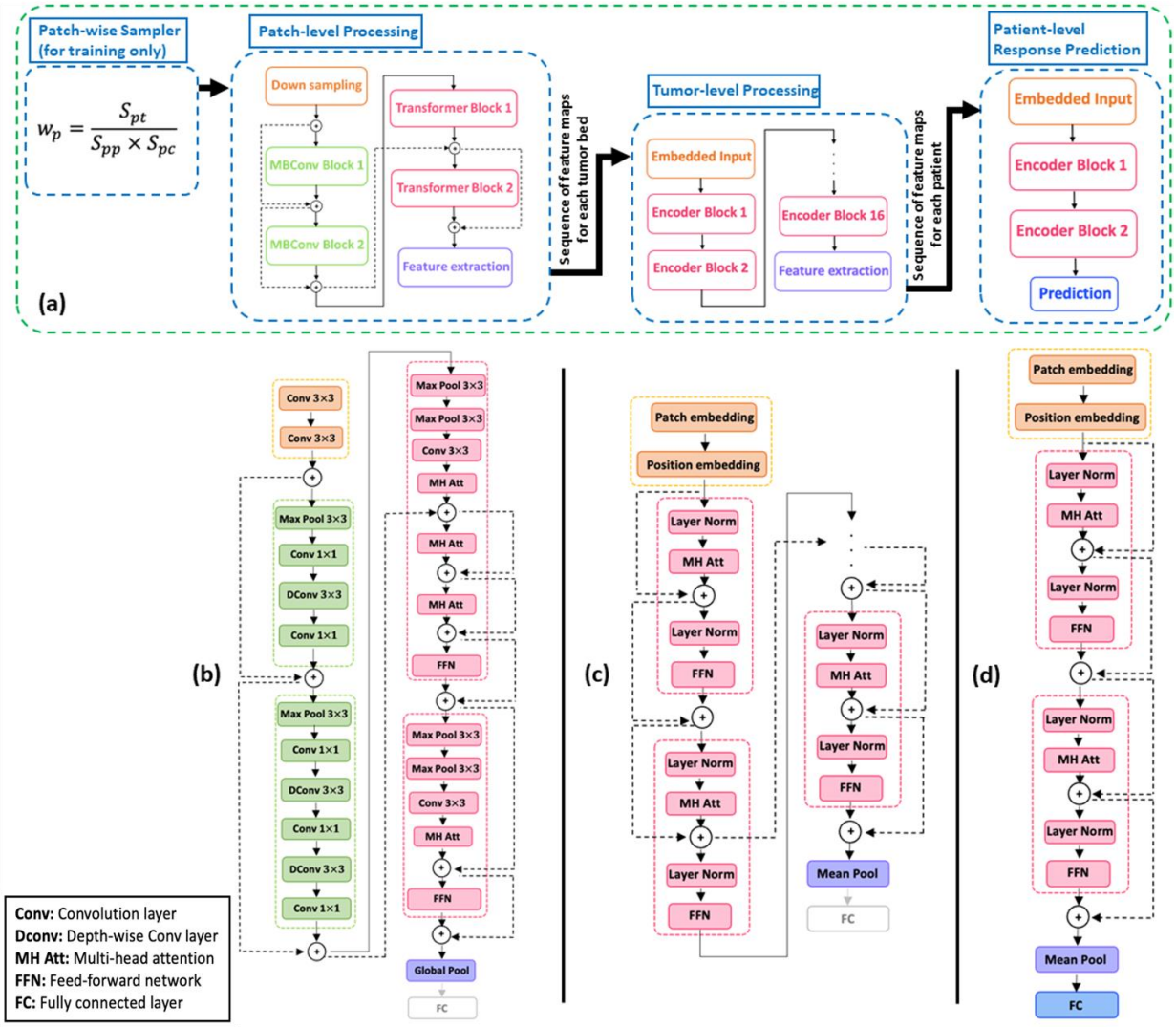


Figure 5.3. Scheme of proposed hierarchical deep-learning framework for NAC outcome prediction. The framework (a) consists of a patch-level sampler for training, and three level of processing. Detailed architectures of the processing modules are shown for patch-level processing module (b), tumor-level processing module (c), and patient-level response prediction module (d). The component colors in (b), (c) and (d) shows the associated block in (a).

$$w_p = \frac{S_{pt}}{S_{pp} \times S_{pc}} \tag{5.1}$$

In Equation (5.1), w_p is the calculated weight for each patient, S_{pt} is the total number of patches in the training set, S_{pp} is the total number of patches for the patient, and S_{pc} is the total number of patches in the associated class of the patient (pCR or non-pCR). After calculating the weights for all

patients in the training set, they are normalized such that the sum of all weights adds up to one. All patches of each patient are assigned equal weights for selection by the sampler. The sampler size (number of patches in each epoch) is tuned as a hyperparameter during the system's training process.

The framework includes a hierarchical flow of the patch-level processing, tumor-level processing, and the patient-level response prediction. It utilizes a self-attention-guided convolutional network architecture followed by two customized ViT network architectures for a hierarchical deep-learning-based analysis of the digital histopathology images for NAC response prediction. In the patch-level processing module, a modified CoAtNet model with two convolutional components followed by two self-attention modules (Figure 5.3(b)) extracts the descriptive features from histopathology image patches that are downsampled to 256×256 pixels using a two-layer convolutional block. A feature map of size 768 associated with each patch is collected from the last layer (global pooling layer) before the fully connected layer in the network. The sequence of feature maps for each tumor bed is generated by stacking the sorted patch-level feature maps of the corresponding tumor bed annotation in the associated histopathology image. Specifically, the feature maps are sorted based on the position of their associated patch in the tumor bed from top left to bottom right. The generated feature map sequence is fed as input to the first ViT model that includes sixteen encoder blocks (Figure 5.3(c)) to explore the relations in aggregated features of each tumor bed. The initial positional encoding vector for the tumor-level processing module is defined based on the generated sequence of the sorted patches. To adapt the sequence of feature maps for the ViT architecture, the input size is defined as the (sequence length $\times$ size of feature map vector). Since the number of patches varies for different tumor beds, a threshold is set for the maximum length of the sequence (tuned as a hyperparameter during the system's training process). For the tumor beds with a smaller number of patches than the threshold, a zero-masking approach is applied, while for the beds with a larger

number of patches, the starting patch of sequence is randomly selected. A similar approach is applied in the patient-level response prediction network (Figure 5.3(d)) to aggregate the information obtained at tumor-level processing and exploiting the global dependencies between the tumor areas for therapy response prediction. A feature map of size 1024 associated with each tumor bed is extracted from the last layer (mean pooling layer) of the tumor-level processing ViT before the fully connected layer. A sequence of the feature maps is generated for each patient to feed into the patient-level response prediction ViT network that includes two encoder blocks. Similar feature map sorting and sequence length thresholding strategies applied for the tumor-level processing ViT are used to create the input sequence of this ViT network.

5.2.4 System Training

The framework's hyperparameters were optimized using a grid search approach with five-fold cross validation on the training set. The sampler size was tuned to 10000 patches for each training epoch. The number of patches per tumor bed, and the number of tumor bed regions per patient varied between 1 to 3879 (median = 6), and 1 to 285 (median = 44), respectively. Accordingly, the threshold for maximum length of the sequence in the tumor-level and patient-level processing modules was tuned to 10 and 50, respectively. These thresholds resulted in an input size of 10×768 and 50×1024 for the associated networks, respectively. The batch size was tuned to 200 for all three networks. A maximum number of 100 epochs was used for training each network. The learning rate was tuned to 0.05 for the patch-level and tumor-level processing networks and 0.01 for the patient-level response prediction network. After the hyperparameter tuning, 20% of the training set patients (n = 29) were selected as the validation set using the stratified random splitting approach and the framework with optimized hyperparameters was trained using the remaining patients in the training

set (80% of the training set, n = 115). Early stopping was utilized for all networks based on the validation loss during the training process to prevent overfitting.

5.2.4 Evaluation

The performance of the proposed framework was assessed on the independent unseen test set using the accuracy, sensitivity, specificity, loss, F1-score, and area under the receiver operating characteristics (ROC) curve (AUC). A threshold value of 0.5 was used as the cut-off to calculate the sensitivity and specificity. To evaluate the performance and effectiveness of the proposed hierarchical framework, ablation experiments were performed with models that only incorporated the patch-level processing, patch-level + tumor-level, and patch-level + patient-level processing modules. In another set of experiments, maximum voting was applied on the output of the patch-level and patch-level + tumor-level models to obtain the tumor-level and patient-level response prediction results, for comparison with those of the hierarchical models. Separate sets of experiments were conducted using two other frameworks that utilized an Xception model coupled with CBAM attention as a state-of-the-art CNN-based model, and a pure self-attended architecture with a ViT model as their patch-level processing network.

A gradient class activation map (Grad-CAM) approach [236] was applied for visualization of the trained patch-level processing attention mechanisms. The Grad-CAM approach provides information on salient regions in an image for a specific class to permit interpreting the network decisions based on the model attention. The attention heatmaps were generated for the patches presented to the trained framework based on their predicted label. The generated heatmaps were stitched together using their position information to create complete attention maps for individual tumor areas. The visualization heatmaps were reviewed to assess and compare the efficacy of attention mechanisms in different networks.

5.3 Results

Table 5.2 presents the evaluation results of different balancing strategies in the patch-level module. The results show the superiority of the proposed patient-wised sampler compared to the weighted loss function approach and label-wised sampler in the patch-level module using the CoAtNet architecture. While the weighted loss function approach could achieve the best training loss of 0.55 compared to the label-wised and patient-wised sampler with the training losses of 0.58 and 0.56, respectively, the difference between the sensitivity and specificity results is higher in this approach. The patient-wised sampler strategy could achieve the best accuracy, sensitivity and specificity on the training, validation and test sets, while the label-wised sampler could generally outperform the weighted loss function approach in terms of these metrics.

Table 5.2. Comparison between different training set balancing strategies at the patch-level module with CoAtNet architecture. The best value in each column is in bold. Acc: accuracy, Sens: sensitivity, Spec: specificity, AUC: area under the ROC curve.

Balancing Strategy	Training Set				Validation Set				Independent Test Set			
	Loss	Acc	Sens	Spec	Loss	Acc	Sens	Spec	Loss	Acc	Sens	Spec
Weighted Loss Function	**0.55**	0.75	0.71	**0.80**	0.64	0.76	0.70	0.78	0.64	0.76	0.71	0.78
Label-wised Sampler	0.58	0.76	0.72	0.78	0.61	0.77	0.72	0.79	0.60	0.78	0.74	0.80
Patient-wised Sampler	0.56	**0.78**	**0.76**	**0.80**	**0.58**	**0.78**	**0.74**	**0.80**	**0.54**	**0.81**	**0.78**	**0.82**

Table 5.3 shows the performance of the proposed framework compared with other similar models. The results demonstrate that the CoAtNet architecture as the patch-level processing module outperformed the Xception model with CBAM attention and the ViT model, with an accuracy of 81% on the test set, compared to accuracies of 79% and 78% respectively. Results of the ablation experiments demonstrate that the three-level hierarchical frameworks could outperform the patch-level only and the two-level (patch-level + tumor-level and patch-level + patient-level) processing

frameworks. The patch-level + tumor-level and patch-level + patient-level processing frameworks consisting of cascaded ViT models resulted in an AUC of 0.78 and 0.77, respectively, on the test set, while the similar architecture with three-level hierarchy resulted in an AUC of 0.80. Also, the two-level hierarchical models (patch-level + tumor-level and patch-level + patient-level) with the Xception + CBAM architecture followed by the ViT module could achieve an AUC of 0.80 and 0.82, respectively, whereas the similar model with three levels of processing could achieve an AUC of 0.86. The tumor-level and patient-level results obtained through maximum voting on the outputs of the patch-level and patch-level + tumor-level models demonstrate a better performance of the corresponding hierarchical models in response prediction. The best performance was achieved by the proposed framework with three-level hierarchy consisting of the CoAtNet architecture as a patch-level processing module and two ViT architectures for the tumor-level processing and patient-level response prediction. This model resulted in an accuracy of 86% on the test set, and a sensitivity, specificity, F1-score and AUC of 87%, 83%, 90% and 0.89 respectively.

Figure 5.4 compares the AUC of the two-level and three-level hierarchical architectures with different patch-level processing modules (ViT, Xception+CBAM, and CoAtNet). The AUCs are in the range of 0.77 and 0.89 with the best results associated with the three-level hierarchical framework using the CoAtNet as the patch-level processing component.

Table 5.3. Performance of different architectures in predicting NAC response using pre-treatment digitized histopathology slides of core needle biopsies. The best value in each column is in bold. Acc: accuracy, Sens: sensitivity, Spec: specificity, AUC: area under the ROC curve.

Model		Training Set				Validation Set				Independent Test set					
		Loss	Acc	Sens	Spec	Loss	Acc	Sens	Spec	Loss	Acc	Sens	Spec	AUC	F1-score
ViT + (None/ViT) + (None/ViT)	**Patch-level**	0.61	0.76	0.76	0.75	0.61	0.77	0.78	0.74	0.59	0.78	0.79	0.75	0.78	0.84
	Max-voting on Patch-level (Tumor-level Results)	-	0.73	0.73	0.73	-	0.74	0.75	0.72	-	0.74	0.75	0.73	-	0.82
	Max-voting on Patch-level (Patient-level Results)	-	0.72	0.73	0.72	-	0.72	0.73	0.71	-	0.73	0.74	0.72	-	0.81
	Patch-level + Tumor-level	0.60	0.77	0.78	0.75	0.60	0.78	0.79	0.75	0.59	0.79	0.80	0.75	0.78	0.85
	Max-voting on Patch-level + Tumor-level (Patient-level Results)	-	0.76	0.77	0.75	-	0.76	0.77	0.74	-	0.78	0.79	0.75	-	0.84
	Patch-level + Patient-level	0.59	0.79	0.79	0.78	0.60	0.78	0.79	0.76	0.57	0.80	0.81	0.78	0.77	0.86
	Patch-level + Tumor-level + Patient-level	0.57	0.80	0.80	0.79	0.58	0.79	0.80	0.76	0.55	0.82	0.83	0.79	0.80	0.87
Xception (+CBAM) + (None/ViT) + (None/ViT)	**Patch-level**	0.57	0.77	0.80	0.74	0.58	0.77	0.79	0.73	0.56	0.79	0.80	0.76	0.79	0.85
	Max-voting on Patch-level (Tumor-level Results)	-	0.76	0.78	0.75	-	0.77	0.78	0.74	-	0.78	0.79	0.76	-	0.84
	Max-voting on Patch-level (Patient-level Results)	-	0.76	0.77	0.76	-	0.76	0.77	0.74	-	0.76	0.77	0.75	-	0.83
	Patch-level + Tumor-level	0.55	0.80	0.81	0.77	0.56	0.80	0.81	0.75	0.55	0.81	0.82	0.77	0.80	0.86
	Max-voting on Patch-level + Tumor-level (Patient-level Results)	-	0.78	0.79	0.77	-	0.79	0.80	0.75	-	0.79	0.80	0.76	-	0.85
	Patch-level + Patient-level	0.54	0.82	0.83	0.79	0.53	0.82	0.84	0.79	0.53	0.83	0.84	0.81	0.82	0.88
	Patch-level + Tumor-level + Patient-level	0.52	0.84	0.84	**0.84**	**0.51**	**0.84**	**0.85**	**0.82**	0.50	0.85	0.86	**0.83**	0.86	**0.90**
CoAtNet + (None/ViT) + (None/ViT)	**Patch-level**	0.56	0.78	0.80	0.76	0.58	0.78	0.80	0.74	0.54	0.81	0.82	0.78	0.79	0.86
	Max-voting on Patch-level (Tumor-level Results)	-	0.76	0.77	0.76	-	0.77	0.78	0.74	-	0.78	0.80	0.77	-	0.85
	Max-voting on Patch-level (Patient-level Results)	-	0.76	0.77	0.75	-	0.76	0.77	0.74	-	0.77	0.79	0.76	-	0.84
	Patch-level + Tumor-level	0.54	0.80	0.83	0.78	0.55	0.80	0.81	0.76	0.53	0.82	0.83	0.78	0.81	0.87
	Max-voting on Patch-level + Tumor-level (Patient-level Results)	-	0.78	0.80	0.75	-	0.79	0.80	0.76	-	0.80	0.81	0.78	-	0.86
	Patch-level + Patient-level	0.52	0.83	0.85	0.80	0.56	0.81	0.83	0.78	0.51	0.85	0.86	0.80	0.84	0.89
	Patch-level + Tumor-level + Patient-level	**0.48**	**0.85**	**0.86**	**0.84**	0.52	**0.84**	**0.85**	0.81	**0.48**	**0.86**	**0.87**	**0.83**	**0.89**	**0.90**

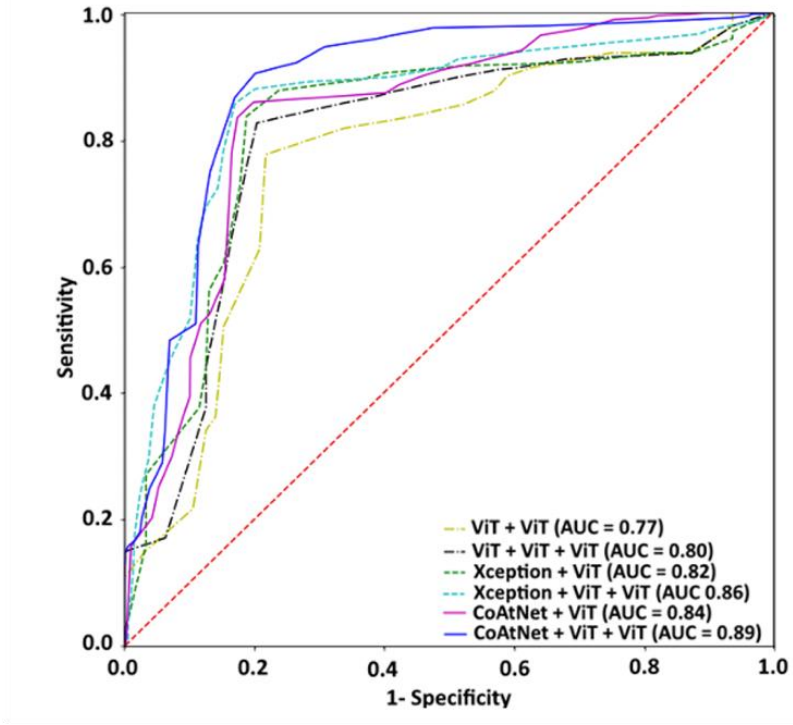


Figure 5.4. Receiver operating characteristic (ROC) curves on the independent test set for the predictive models developed with two-level (patch-level + patient-level: ViT + ViT, Xception + ViT, and CoAtNet + ViT) and three-level (patch-level + tumor-level + patient-level: ViT + ViT + ViT, Xception + ViT + ViT, and CoAtNet + ViT + ViT) hierarchical framework. Different patch-level processing modules were utilized for comparison.

In a separate experiment, the clinical features described in Section 4.2.1 were incorporated into the patient-level response prediction module as a new token input to the ViT architecture to evaluate the possible contribution of the clinical features to the response prediction models. The categorical features were encoded using one-hot vectors, and the continuous features were normalized between 0 and 1 using the training set data, before inputting the model. The results demonstrated that incorporating the clinical features into the framework could not improve the performance of the prediction models.

Figure 5.5 presents the attention heatmaps obtained for two representative tumor regions: one with a pCR and the other one with a non-pCR outcome after NAC. The heatmaps were generated for each

patch in the tumor bed region (contoured by an expert pathologist) using the Grad-CAM approach and overlaid on the original histopathology image for visualization. A comparison between the heatmaps associated with different patch-level processing modules shows that the CoAtNet architecture has demonstrated more focus on the tumor regions compared to the other two networks.

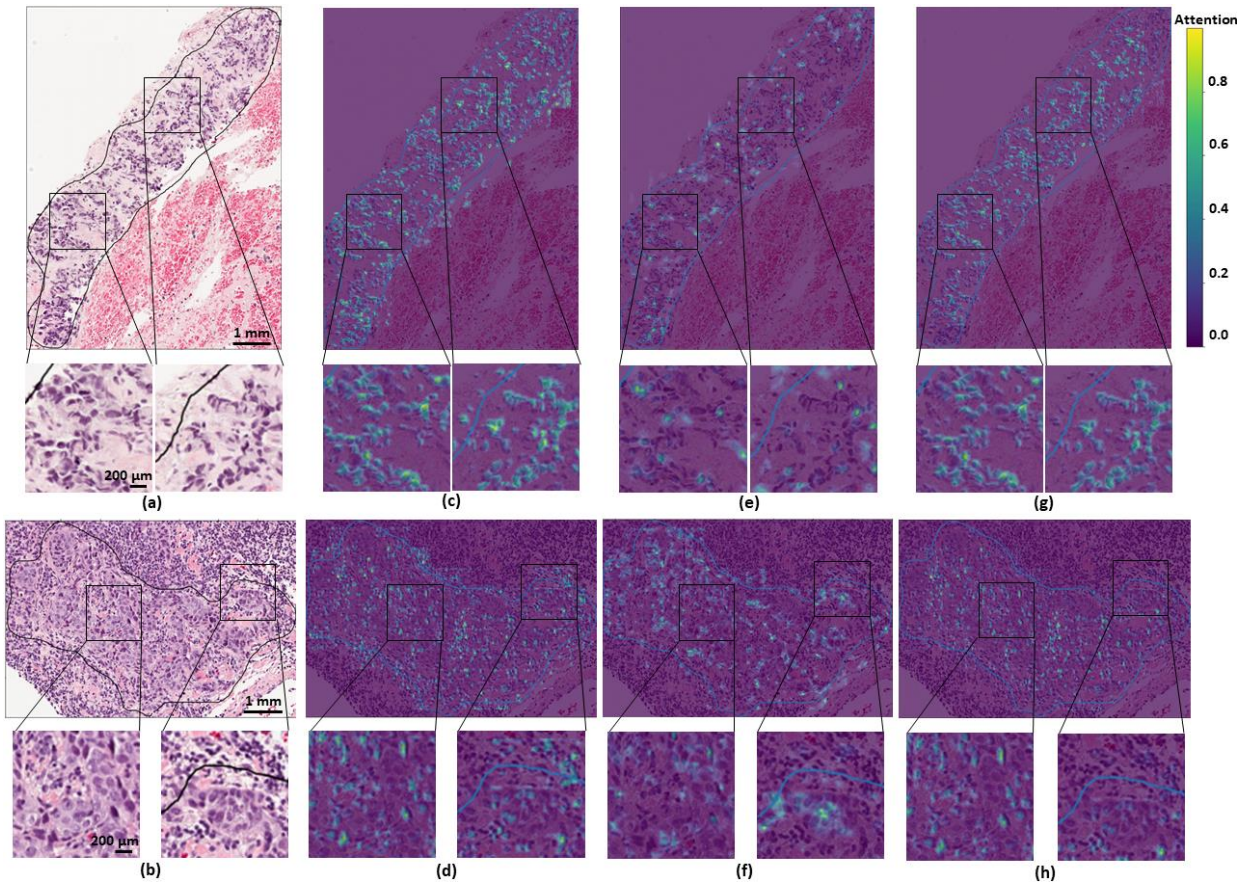


Figure 5.5. Comparison of Grad-CAM attention heatmaps associated with different patch-level processing modules for two representative tumor regions with a pCR (top row) and non-pCR (bottom row) outcome after NAC, respectively: (a, b) tumor areas extracted from the H&E-stained WSIs (40×), (c, d) attention heatmaps of ViT, (e, f) attention heatmaps of Xception + CBAM, (g, h) attention heatmaps of CoAtNet. The contours show the tumor bed annotations drawn by an expert pathologist. The CoAtNet architecture demonstrates more focus on tumor area, compared to the other networks.

5.4 Discussion and Conclusion

In this chapter, a hierarchical deep learning framework was proposed for the first time to predict breast cancer response to NAC using digital histopathology images of pre-treatment biopsy specimens. The proposed model consists of a patch-level processing module followed by a tumor-level processing module and a patient-level response prediction module. A self-attention-guided convolutional network based on the CoAtNet architecture has been adapted for the patch-level processing step, with two ViT models for aggregating the sequence of feature maps at the tumor-level and predicting the therapy response, respectively. The patch-level processing module was applied to capture the local correlations within the tumor microenvironment and extract the feature maps carrying relevant information of the pathology image patches. The tumor-level processing step has been used to aggregate the local information for each tumor region. The patient-level prediction module utilized the sequence of information for various tumor regions to derive the global relations and predict the patient response to NAC. The proposed model could predict the NAC response of the patients in an independent test set with a sensitivity, specificity, and F1-score of 87%, 83%, and 90%, respectively. A comparison between the attention heatmaps generated by various patch-level processing architectures demonstrates that the CoAtNet architecture paid more attention to the tumor areas compared to surrounding the healthy tissues. The performance of the proposed hierarchical model shows that the combination of convolutional blocks with self-attention modules adapted as the patch-level processing component can effectively extract local information within the tumor patches, while the relations between these features in each tumor area and at patient-level for NAC response prediction can be successfully modelled using the vision transformer modules. The model performance on the validation and test sets is slightly better compared to the training set. A possible reason for this pattern can be due to the fact that for addressing the imbalance issue of the dataset, a

number of regulations were applied during the training process such as using a patient-wise sampler and an early stopping strategy based on the performance on the validation set. Also, as the validation and test sets are imbalance, slight improvements in predicting the majority class (non-pCR) would have a magnified effect on the overall performance of the model.

The proposed hierarchical framework provides an effective approach for analyzing the WSIs at high-resolution. Results of the ablation experiments in this study demonstrate that the three-level hierarchical model could outperform the patch-level only and the two-level hierarchical models in predicting the NAC response. The performance of a purely convolutional architecture, a completely self-attention-based model, and a combination of convolutional and self-attention components in extracting informative features from the tumor patches were compared. The results demonstrate that coupling the convolutional and self-attention modules is a more effective architecture for extracting local features. The positional embedding approach proposed in this study, enables the framework to extract global relations within and between different tumor areas and predict the pCR/non-pCR outcome for each patient using multi-head attention modules.

The obtained results in this study show that the local features extracted at the patch and tumor levels carry meaningful information for NAC response prediction. However, the relations within the aggregated patch-level information and the global dependencies between the tumor areas should also be taken into account for a more accurate response prediction at the tumor and patient levels. This is supported by the results of the comparative evaluations performed with maximum voting for predicting NAC response at the tumor and patient levels. Specifically, while the models with maximum voting receive the information from all patches or tumors in a WSI, they have demonstrated inferior performance compared to their hierarchical-model counterparts. Further, aggregating the patch- or tumor-level information using the maximum voting strategy has resulted

in slightly lower accuracy in response prediction compared to the individual patch/tumor level results. These observations highlight the importance of a systematized strategy to fuse the patch/tumor-level information on the tumor/patient level for NAC response prediction. The size and number of the tumor beds are quite variable among the biopsy samples, leading to large variations in the number of patches per tumor bed and WSI. Such variations can potentially lead to inferior aggregated results on the tumor level and patient level based on maximum voting, compared to the corresponding patch-level and patch-level + tumor-level models.

Information on cellular interactions and activities within the tumor microenvironment can potentially be captured from imaging. It may be possible to enhance diagnostic and response-guided therapy approaches by identifying biomarker signatures obtained through mapping tumor subcomponents and assessing biological heterogeneity in digital pathology images.[195] Previous studies have investigated the use of radiomic features from different medical imaging data collected at early stages of diagnosis for predicting the NAC response in breast cancer [198], [200], [217]. The results obtained in those investigations demonstrate that intra-tumor heterogeneity quantified on imaging at pre-treatment can be associated with response to NAC. However, these images may not be acquired routinely for breast cancer diagnosis. As such, integrating this information into a therapy response prediction framework may necessitate additional imaging data collection and processing that would not be always possible.

The study presented in Chapter 4 of this dissertation has demonstrated the potential of hand-crafted pathomic features coupled with conventional ML models in predicting breast cancer response to NAC [237]. The ML models were developed using the training data acquired from 111 patients, where the best model achieved an accuracy of 84% (sensitivity: 85%; specificity: 82%) on the test set (38 patients). The observations of that research are in agreement with the results obtained in this

study. However, extracting hand-crafted features could possibly be affected by feature extraction protocols that influence their reproducibility. Also, while the hand-crafted feature-based conventional ML models have a decent potential in analyzing imaging data, they are bounded with the information provided by a set of mathematically defined features. The data-driven deep-learning models such as the one implemented in this study have shown higher performance in analyzing large-scale data, and better scalability with growing datasets. In a very recent study, deep learning methods have been applied to predict breast cancer response to NAC using three parallel pathology images as the input [238]. The digital pathology images were obtained from the histopathology slides with H&E staining and Ki67 and phosphohistone H3 (PHH3) immunohistochemistry. The models were developed using a training set of 43 patients and evaluated on a test set with 30 patients. Using maximum voting on the patch-level results, their best model achieved a test accuracy of 93% on the patient level for detecting pCR to NAC. Those results, albeit obtained on a relatively small dataset, support the utility of deep learning models in conjunction with digital pathology images of tumor biopsies for NAC response prediction. However, the models developed in that study require multi-modal pathology images with immunohistochemistry that are not routinely performed on pre-treatment tumor biopsies in the clinic.

The data in this study was acquired from a single institution retrospectively. Future multi-institutional studies with external validation are required for more rigorous evaluation of the developed framework for NAC response prediction. The current framework relies on the manual annotation of the tumor beds by expert pathologists on WSIs that is time consuming in the clinical setting. Future works can possibly address this issue by developing automated tumor annotation methods and investigating their efficacy when integrated with the NAC response prediction framework. The framework proposed in this study directly analyses the pathology image patches

with no explicit pre-processing step for nucleus detection and tumor cell classification. Future studies can investigate potential impacts of incorporating such pre-processing steps in the framework to extract tumor microenvironment features of the cancer cells specifically. The methods proposed in this study can potentially be adapted for the analysis of digital pathology WSIs in other applications. This includes diagnostic and prognostic applications for different cancer types such as the breast [239], prostate [240], liver [241], and the lung cancer [242].

In conclusion, this study presented an automated hierarchical framework for analyzing digital histopathology images of biopsy specimens and predicting NAC response at pre-treatment with promising results. The effectiveness of combining convolutional blocks with self-attention modules for hierarchical analysis of high-resolution histopathology images was demonstrated. The results of this study pave the way toward a response-guided therapy paradigm for individual breast cancer patients and motivate future studies on larger multi-institutional datasets for further investigation of the proposed methodologies.

Chapter 6

Conclusion and Future Directions

6.1 Summary and Conclusion

In this dissertation project, various machine learning and quantitative image analysis methods have been developed, optimized, and investigated to analyze digital histopathology images automatically. These methods have been integrated into innovative systems for tissue characterization, to streamline the histopathology assessments workflow, and therapy response prediction in breast cancer, using the data available at pre-treatment. The overarching goal of the project is to pave the way toward a precision oncology paradigm for breast cancer, potentially leading to enhanced therapy outcomes, and improved patient survival and quality of life. To pursue the project's objectives four main studies have been conducted and investigated, as summarized below.

First, a novel multi-scale attention-guided deep learning model was proposed and investigated for breast tissue characterization workflow to classify digitized biopsy specimens into normal, benign, *in situ* and invasive carcinoma tissues, automatically. In the next step, a new cascaded U-Net-based architecture was developed, optimized, and evaluated for segmenting the nuclei accurately in whole slide digital histopathology images. Ultimately, two novel approaches were investigated to predict breast cancer response to NAC prior to the start of treatment using digital histopathology images. In the first approach, a novel NAC response prediction framework was developed and investigated using hand-crafted pathomic features in conjunction with conventional machine learning methods. In the

second approach, an innovative end-to-end deep learning platform was introduced and comprehensively investigated for NAC response prediction using digital histopathology images of pre-treatment tumor biopsy specimens. The investigations conducted in this dissertation were presented in four chapters. A summary of the key research contributions, main findings, and conclusions for each chapter is presented below.

6.1.1 Chapter 2: Breast Tissue Characterization in Digital Histopathology Images

In Chapter 2, a multi-scale attention-guided deep learning model has been proposed to characterize breast tissue in digital histopathology images according to four histological types, including normal, benign, *in situ* carcinoma and invasive carcinoma. The framework includes two parallel convolutional neural networks with a modified VGG16 architecture. The first network analyzes the whole-sample images at low magnification. The second network focuses on the patches extracted from the whole-sample images at high magnification. In the low-magnification network, a global average pooling layer is added at the end of the network to extract the class activation maps for the attention model. A long short-term memory (LSTM) network is adapted as a recurrent attention mechanism to increase the contribution of the relevant parts of each image for classification. In the high-magnification network, the probability vectors are averaged over all patches extracted from an image to obtain the probability vectors associated with the four histological types for each sample. The probability vectors for each sample from the high-magnification network and the attention model are fused using a multilayer perceptron network to generate a classification label. Obtained results on the test set demonstrated an average accuracy of 97.5% ± 1.0% for the proposed model. An average accuracy of 94.5%, 93.5%, and 96.3% was obtained, respectively, for the separate high- and low-magnification networks, and the multi-scale model without an attention mechanism. The

results demonstrate that a multi-scale strategy coupled with an attention mechanism can improve the accuracy of deep learning models in classifying digital histopathology images.

6.1.2 Chapter 3: Nuclei Segmentation in Digital Histopathology Images

A novel cascaded architecture has been developed and evaluated to segment the nuclei in digital histopathology images accurately. The models proposed previously for this task can hardly find the exact border of nuclei, and in practice, the models cannot usually segment the touching or very closed nuclei accurately. To overcome this challenge, the segmentation framework proposed in this chapter consists of a U-Net-based model with customized pixel-wised weighted loss function, followed by a U-Net-based model with a VGG16 backbone and a soft Dice loss function. The output of the weighted U-Net model, which is a probability mask for the input image, is passed to the Vanilla U-Net model to generate the final binary mask. The proposed cascaded model could outperform the other state-of-the-art models with an AJI of 0.72 and a F1-score of 0.83 on the MoNuSeg test set. The results demonstrate the effectiveness of cascading U-Net based architectures in delineating the challenging nuclei accurately.

6.1.3 Chapter 4: Quantitative Pathomic Features with Conventional Machine Learning Models for Therapy Response Prediction

Chapter 4 has presented, for the first time, a machine learning framework for quantitative analysis of digital histopathology images acquired at pre-treatment to predict breast cancer response to NAC. In this framework, the hand-crafted pathomic features are extracted from the segmented nuclei within the tumor regions. For this purpose, a number of pre-processing steps are performed on the whole slide histopathology images. The annotated tumor regions are processed to extract non-overlapping tiles, and the proposed nuclei segmentation model is utilized to segment the nuclei. Several pathomic features are extracted from each segmented image tile for analysis. The clinical

and pathomic features are analyzed through a feature reduction and selection process on the training set to develop optimal biomarkers for NAC response prediction. Independent experiments have been performed on different initial subsets of the features to compare their performance. Gradient boosting machine (GBM) models have been adapted and optimized as conventional multi-feature machine learning classifiers for feature selection and optimal biomarker development, and for pCR prediction using the optimal pathomic biomarkers developed. The predictive model developed with optimal feature set outperformed the other models, with an accuracy of 84%, a sensitivity of 85% and a specificity of 82% on the independent test set. The results demonstrate the potential of quantitative digital histopathology features integrated with ML methods in predicting breast cancer response to NAC.

6.1.4 Chapter 5: A Hierarchical Attention-guided Deep Learning Architecture for Therapy Response Prediction using Pre-treatment Tumor Biopsies

The second approach proposed for predicting NAC response using pre-treatment digital histopathology images has been presented in this chapter. A hierarchical self-attention-guided deep learning framework has been introduced to predict pathological complete response to chemotherapy in breast cancer patients using pre-treatment tumor biopsies. The digital histopathology images are processed through the proposed hierarchical framework consisting of patch-level and tumor-level processing modules followed by a patient-level response prediction component. A combination of convolutional layers and transformer self-attention blocks are utilized in the patch-level processing architecture to generate optimized feature maps. The feature map sequences are defined based on the patch positions within the tumor beds and the bed positions within the biopsy slide for analysis through two adapted vision transformer architectures for tumor-level processing and patient-level response prediction components, respectively. Obtained results on the independent test set showed

an accuracy, sensitivity and specificity of 0.86, 83%, and 87%, respectively, for the proposed hierarchical framework in predicting pCR to NAC *a priori*. Experimental results demonstrate that incorporating the clinical features into the framework (by inputting a new embedding vector consisting of the one-hot-encoded clinical features to the patient-level ViT architecture) could not improve the prediction model's performance. This finding is in agreement with the observations of the study presented in Chapter 4, where none of the clinical features were selected along with the quantitative pathomic features in the optimal feature set for the final predictive model. The results obtained in this project, demonstrate the high potential of the proposed hierarchical deep-learning methodology in analyzing very large-size histopathology images to predict pathological response to NAC at pre-treatment.

6. 2 Limitations and Future Directions

This dissertation project has proposed and investigated different computational and machine learning techniques to streamline or potentially improve various steps in the clinical management of breast cancer. The promising results reported in Chapters 2 to 4 for the proposed frameworks are encouraging and demonstrate the high potential of these methods for paving the way toward a personalized oncology paradigm for breast cancer patients. However, further studies on larger patient cohorts with multi-institutional data and external validation are needed to evaluate the efficacy and robustness of the proposed pipeline more rigorously before they can be recommended and approved for clinical applications. Especially, due to the nature of deep learning architectures and the number of parameters that should be optimized in these models, a larger training set is required to ensure high generalization ability of the models in clinic. Availability of larger datasets could potentially improve the performance of the deep learning models developed in this research

and allow researchers to investigate more expansive, complicated, and possibly powerful architectures for such applications in future. Data acquisition for the studies on NAC response prediction requires an extensive period of time due to the long clinical process involved from diagnosis to definitive response and outcome determination for each patient. Therefore, collecting large yet complete datasets for these studies would be challenging. Also, due to the privacy concerns, clinical institutions can hardly share their collected data with other researchers. This would make the multi-institutional data acquisition more challenging. Another impediment in data collection for the studies aiming at ML model development and optimization is related to the current rate of the complete pathological response to NAC in LABC patients (about 30%). Hence, collecting more data would not necessarily resolve the imbalance issue of dataset in such studies.

Nowadays, deep learning approaches are growing fast, and specifically, several new state-of-the-art architectures are being developed for image analysis [243], [244]. As lack of data is always a barrier in medical image analysis, exploring new efficient architectures with fewer parameters would be an interesting future path for this research. Such architectures may permit developing end-to-end models with all components being trained simultaneously to analyze the entire WSIs for tissue characterization or therapy response prediction with no human interactions.

The focus of this study was on analyzing of the tumor regions on WSIs annotated by the expert pathologists. However, the manual annotation of WSIs is time consuming and could be associated with human bias and error. Developing an automated annotating system could accelerate the process of histopathology image analysis, potentially with high confidence. In this dissertation, an automated framework for tissue characterization was proposed in Chapter 2. However, the proposed model could classify the histology type of the tissue in a relatively small patch. The recently proposed ViT-Adapter-L network [245] has yielded state-of-the-art performance in various dense prediction tasks,

including object detection, instance segmentation, and semantic segmentation. The proposed model utilizes the ViT architecture as the backbone of its framework to learn representations from large-scale multi-modal data. This network can potentially be adapted for the task of tumor region annotation in high-resolution WSIs. Another recent study has proposed a novel ViT architecture, namely Hierarchical Image Pyramid Transformer (HIPT) [246], to analyze the high-resolution histopathology images. The proposed model uses two levels of self-supervised learning to learn high-resolution image representations by exploiting the natural hierarchical structure inherent in WSIs. These studies have shown the potential of the deep-learning-based segmentation models for annotating the tumor regions in WSIs. Such methods can be adapted and further investigated in future to assess their performance and reliability when integrated with the systems proposed in this study for tissue characterization and NAC response prediction.

Other complementary information can be integrated into the proposed response prediction frameworks to enhance, possibly, the performance of the models. Previous studies have shown the correlation of specific genes with breast tumor behavior [247]. Accordingly, investigating machine learning models that input the digital histopathology images along with patient's genomic data, and exploring possible mutual correlations and contribution of this information to therapy response prediction in breast cancer, can be an interesting future direction for this research project. Collecting genomic data, that are currently not acquired as part of the standard of care, from breast cancer patients along with digital histopathology data is a necessity for such investigations.

The methodologies and algorithms developed and investigated in this dissertation project can potentially be extended to other applications for automated analysis of digital histopathology images, tissue characterization and prediction of therapy response and outcome in different types of malignancies, including lung cancer, prostate cancer, and colorectal cancer. The promising results

reported in this dissertation encourage future studies on the development of computational data-driven models as valuable decision support tools for the diagnosis and prognosis of various primary and metastatic cancers treated with different modalities including systemic therapies and/or radiotherapy.

6.3 Closing Remarks

Various computational frameworks have been introduced and investigated in this dissertation project to develop intelligent systems for automated tissue characterization and treatment response prediction in breast cancer using pre-treatment digital histopathology images. The results obtained in this research have shown the high potential of machine learning approaches, including state-of-the-art deep learning architectures in high-throughput analysis of digital histopathology images for different tasks, including tissue characterization, nuclei segmentation, and treatment response prediction. However, further investigations, including multi-institutional data are required to assess the efficacy, robustness, and reliability of the proposed systems in clinic more rigorously.

References

[1] Canadian Cancer Society, “Cancer statistics at a glance,” 2020. https://cancer.ca/en/research/cancer-statistics/cancer-statistics-at-a-glance?gclid=Cj0KCQjwr4eYBhDrARIsANPywCgRwk91-oylaked8wQQF000XejyEncAEmV-u21bjoWZcJYfPLlu2F0aAlVmEALw_wcB (accessed Aug. 21, 2022).

[2] World Health Organization (WHO), “Breast Cancer,” 2021. https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed Oct. 30, 2022).

[3] S. Lei *et al.*, “Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020,” *Cancer Commun*, vol. 41, no. 11, pp. 1183–1194, 2021, doi: 10.1002/cac2.12207.

[4] J. M. Seely and T. Alhassan, “Screening for breast cancer in 2018—what should we be doing today?,” *Current Oncology*, vol. 25, p. 115, Jun. 2018, doi: 10.3747/co.25.3770.

[5] G. M. Reedy, “Current year estimates for breast cancer,” *American Cancer Society*, 2020. https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html. (accessed Jul. 01, 2020).

[6] G. R. et Al, “Breast Cancer,” *American Cancer Society*. https://www.cancer.org/cancer/breast-cancer.html (accessed Jul. 06, 2020).

[7] G. M. Reedy, “Current year estimates for breast cancer,” *American Cancer Society*, 2020. https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html. (accessed Jun. 30, 2020).

[8] National Breast Cancer Foundation, “Breast Tumors.” https://www.nationalbreastcancer.org/breast-tumors/ (accessed Jun. 30, 2020).

[9] J. O’Donnell-Tormey and Matthew Tontonoz, *Cancer and the Immune System: The Vital Connection*. Cancer Research Institute, 2016.

[10] Centers for disease control and prevention, “What Is Breast Cancer?” https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm (accessed Feb. 06, 2020).

[11] “LCIS and Breast Cancer Risk,” *Breast Cancer Community*. https://www.breastcancer.org/symptoms/types/lcis/cancer_risk (accessed Mar. 01, 2020).

[12] H.-P. Sinn and H. Kreipe, “A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition,” *Breast Care*, vol. 8, no. 2, pp. 149–154, 2013, doi: 10.1159/000350774.

[13] “LCIS and Breast Cancer Risk,” *Breast Cancer Community*. https://www.breastcancer.org/symptoms/types/lcis/cancer_risk (accessed Feb. 29, 2020).

[14] National Breast Cancer Foundation, “Types of Breast Cancer.” https://www.nationalbreastcancer.org/types-of-breast-cancer/ (accessed Nov. 30, 2019).

[15] B. Iqbal and A. Buch, “Hormone receptor (ER, PR, HER2/neu) status and proliferation index marker (Ki-67) in breast cancers: Their onco-pathological correlation, shortcomings and future trends,” *Medical Journal of Dr. D.Y. Patil University*, vol. 9, no. 6, p. 674, 2016, doi: 10.4103/0975-2870.194180.

[16] D. C. Allred, “Issues and updates: evaluating estrogen receptor-α, progesterone receptor, and HER2 in breast cancer,” *Modern Pathology*, vol. 23, no. S2, pp. S52–S59, 2010, doi: 10.1038/modpathol.2010.55.

[17] Y. Feng *et al.*, “Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis,” *Genes Dis*, vol. 5, no. 2, pp. 77–106, 2018, doi: 10.1016/j.gendis.2018.05.001.

[18] B. David R., C. Michael A., C. Carolyn C., A. G. Fritz, and F. L. Greene., *AJCC cancer staging manual*. Springer, 2010.

[19] J. Koh and M. J. Kim, “Introduction of a New Staging System of Breast Cancer for Radiologists: An Emphasis on the Prognostic Stage,” *Korean J Radiol*, vol. 20, no. 1, p. 69, 2019, doi: 10.3348/kjr.2018.0231.

[20] A. Taghian and S. D Merajver, “Patient education: Locally advanced and inflammatory breast cancer (Beyond the Basics),” *UpToDate*. https://www.uptodate.com/contents/locally-advanced-and-inflammatory-breast-cancer-beyond-the-basics (accessed Oct. 14, 2021).

[21] K. Tryfonidis, E. Senkus, M. J. Cardoso, and F. Cardoso, “Management of locally advanced breast cancer—perspectives and future directions,” *Nat Rev Clin Oncol*, vol. 12, no. 3, pp. 147–162, 2015.

[22] A. Sadeghi-Naini *et al.*, “Quantitative Ultrasound Evaluation of Tumor Cell Death Response in Locally Advanced Breast Cancer Patients Receiving Chemotherapy,” *Clinical Cancer Research*, vol. 19, no. 8, pp. 2163–2174, 2013, doi: 10.1158/1078-0432.CCR-12-2965.

[23] D. C. Allred, S. K. Mohsin, and S. A. Fuqua, “Histological and biological evolution of human premalignant breast disease.,” *Endocrine-related cancer Endocr Relat Cancer Endocr. Relat. Cancer*, vol. 8, no. 1, pp. 47–61, 2001.

[24] A. Adeyinka *et al.*, “Analysis of Gene Expression in Ductal Carcinoma in Situ of the Breast1,” *Clinical Cancer Research*, vol. 8, no. 12, pp. 3788–3795, Dec. 2002.

[25] J. A. Somarelli *et al.*, “Molecular Biology and Evolution of Cancer: From Discovery to Action,” *Mol Biol Evol*, vol. 37, no. 2, pp. 320–326, 2020, doi: 10.1093/molbev/msz242.

[26] B. Erbas, E. Provenzano, J. Armes, and D. Gertig, “The natural history of ductal carcinoma in situ of the breast: a review,” *Breast Cancer Res Treat*, vol. 97, no. 2, pp. 135–144, 2006, doi: 10.1007/s10549-005-9101-z.

[27] M. R. Ataollahi, J. Sharifi, M. R. Paknahad, and A. Paknahad, “Breast cancer and associated factors: a review.,” *J Med Life*, vol. 8, no. 4, pp. 6–11, 2015.

[28] J. Coignard *et al.*, “A case-only study to identify genetic modifiers of breast cancer risk for BRCA1/BRCA2 mutation carriers,” *Nat Commun*, vol. 12, no. 1, p. 1078, 2021, doi: 10.1038/s41467-020-20496-3.

[29] Y.-S. Sun *et al.*, “Risk Factors and Preventions of Breast Cancer,” *Int J Biol Sci*, vol. 13, no. 11, pp. 1387–1397, 2017, doi: 10.7150/ijbs.21635.

[30] S. W. Duffy *et al.*, “Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer,” *Eur J Cancer*, vol. 88, pp. 48–56, 2018, doi: 10.1016/j.ejca.2017.10.022.

[31] L. M. Morimoto *et al.*, “Obesity, body size, and risk of postmenopausal breast cancer: The women’s health initiative (United States),” *Cancer Causes and Control*, vol. 13, no. 8, pp. 741–751, 2002, doi: 10.1023/A:1020239211145.

[32] J. L. Kelsey, M. D. Gammon, and E. M. John, “Reproductive Factors and Breast Cancer,” *Epidemiol Rev*, vol. 15, no. 1, pp. 36–47, 1993, doi: 10.1093/oxfordjournals.epirev.a036115.

[33] E. Warner *et al.*, “Prospective Study of Breast Cancer Incidence in Women With a BRCA1 or BRCA2 Mutation Under Surveillance With and Without Magnetic Resonance Imaging,” *Journal of Clinical Oncology*, vol. 29, no. 13, pp. 1664–1669, 2011, doi: 10.1200/JCO.2009.27.0835.

[34] H. Noh, S. Hong, and B. Han, “Learning Deconvolution Network for Semantic Segmentation,” May 2015.

[35] S. Shapiro, “Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. International Breast Cancer Screening Network (IBSN) and the European Network of Pilot Projects for Breast Cancer Screening,” *Int J Epidemiol*, vol. 27, no. 5, pp. 735–742, 1998, doi: 10.1093/ije/27.5.735.

[36] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, “Computer-aided detection and classification of microcalcifications in mammograms: a survey,” *Pattern Recognit*, vol. 36, no. 12, pp. 2967–2991, 2003, doi: 10.1016/S0031-3203(03)00192-4.

[37] M. Kriege *et al.*, “Efficacy of MRI and Mammography for Breast-Cancer Screening in Women with a Familial or Genetic Predisposition,” *New England Journal of Medicine*, vol. 351, no. 5, pp. 427–437, 2004, doi: 10.1056/NEJMoa031759.

[38] D. Saslow *et al.*, “American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography,” *CA Cancer J Clin*, vol. 57, no. 2, pp. 75–89, 2007, doi: 10.3322/canjclin.57.2.75.

[39] C. Kuhl *et al.*, “Prospective Multicenter Cohort Study to Refine Management Recommendations for Women at Elevated Familial Risk of Breast Cancer: The EVA Trial,” *Journal of Clinical Oncology*, vol. 28, no. 9, pp. 1450–1457, 2010, doi: 10.1200/JCO.2009.23.0839.

[40] J. R. Scheel, J. M. Lee, B. L. Sprague, C. I. Lee, and C. D. Lehman, “Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts,” *Am J Obstet Gynecol*, vol. 212, no. 1, pp. 9–17, Jan. 2015, doi: 10.1016/j.ajog.2014.06.048.

[41] G. L. Menezes, “Magnetic resonance imaging in breast cancer: A literature review and future perspectives,” *World J Clin Oncol*, vol. 5, no. 2, p. 61, 2014, doi: 10.5306/wjco.v5.i2.61.

[42] C. Loukas, S. Kostopoulos, A. Tanoglidi, D. Glotsos, C. Sfikas, and D. Cavouras, “Breast Cancer Characterization Based on Image Classification of Tissue Sections Visualized under Low Magnification,” *Comput Math Methods Med*, vol. 2013, pp. 1–7, 2013, doi: 10.1155/2013/829461.

[43] D. Ivan, V. Selinko, A. A. Sahin, N. Sneige, and L. P. Middleton, “Accuracy of core needle biopsy diagnosis in assessing papillary breast lesions: histologic predictors of malignancy,” *Modern Pathology*, vol. 17, no. 2, pp. 165–171, Feb. 2004, doi: 10.1038/modpathol.3800014.

[44] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, “Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model,” *Sci Rep*, vol. 7, no. 1, p. 4172, 2017, doi: 10.1038/s41598-017-04075-z.

[45] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever, “Breast Cancer Histopathology Image Analysis: A Review,” *IEEE Trans Biomed Eng*, vol. 61, no. 5, pp. 1400–1411, 2014, doi: 10.1109/TBME.2014.2303852.

[46] R. D. Cardiff, C. H. Miller, and R. J. Munn, “Manual Hematoxylin and Eosin Staining of Mouse Tissue Sections,” *Cold Spring Harb Protoc*, vol. 2014, no. 6, p. pdb-prot073411, 2014, doi: 10.1101/pdb.prot073411.

[47] A. L. Martel *et al.*, “An Image Analysis Resource for Cancer Research: PIIP—Pathology Image Informatics Platform for Visualization, Analysis, and Management,” *Cancer Res*, vol. 77, no. 21, pp. e83–e86, 2017, doi: 10.1158/0008-5472.CAN-17-0323.

[48] P. Bankhead *et al.*, “QuPath: Open source software for digital pathology image analysis,” *Sci Rep*, vol. 7, no. 1, p. 16878, 2017, doi: 10.1038/s41598-017-17204-5.

[49] P. M. Ravdin *et al.*, “Computer Program to Assist in Making Decisions About Adjuvant Therapy for Women With Early Breast Cancer,” *Journal of Clinical Oncology*, vol. 19, no. 4, pp. 980–991, 2001, doi: 10.1200/JCO.2001.19.4.980.

[50] T.-A. Moo, R. Sanford, C. Dang, and M. Morrow, “Overview of Breast Cancer Therapy,” *PET Clin*, vol. 13, no. 3, pp. 339–354, 2018, doi: 10.1016/j.cpet.2018.02.006.

[51] American Cancer Society, “Hormone Therapy for Breast Cancer.” https://www.cancer.org/cancer/breast-cancer/treatment/hormone-therapy-for-breast-cancer.html (accessed Aug. 21, 2022).

[52] T. Whelan *et al.*, “Mastectomy or Lumpectomy? Helping Women Make Informed Choices,” *Journal of Clinical Oncology*, vol. 17, no. 6, pp. 1727–1727, Jun. 1999, doi: 10.1200/JCO.1999.17.6.1727.

[53] CTCA, “What’s the difference: Adjuvant and neoadjuvant therapies,” *Cancer Treatment Centers of America*, 2019. https://www.cancercenter.com/community/blog/2019/01/whats-the-difference-adjuvant-and-neoadjuvant-therapies (accessed Aug. 21, 2022).

[54] W. F. Symmans *et al.*, “Assessment of Residual Cancer Burden and Event-Free Survival in Neoadjuvant Treatment for High-risk Breast Cancer,” *JAMA Oncol*, vol. 7, no. 11, p. 1654, 2021, doi: 10.1001/jamaoncol.2021.3690.

[55] A. F. Schott and D. F. Hayes, “Defining the Benefits of Neoadjuvant Chemotherapy for Breast Cancer,” *Journal of Clinical Oncology*, vol. 30, no. 15, pp. 1747–1749, May 2012, doi: 10.1200/JCO.2011.41.3161.

[56] M. Markman, “Recurrent breast cancer,” *Cancer treatment centers of America*, 2021. https://www.cancercenter.com/cancer-types/breast-cancer/types/rare-breast-cancer-types/recurrent-breast-cancer (accessed Nov. 28, 2021).

[57] D. Nguyen, J. Yu, W. C. Reinhold, and S. X. Yang, “Association of Independent Prognostic Factors and Treatment Modality With Survival and Recurrence Outcomes in Breast Cancer,” *JAMA Netw Open*, vol. 3, no. 7, p. e207213, 2020, doi: 10.1001/jamanetworkopen.2020.7213.

[58] L. M. Spring *et al.*, “Pathologic Complete Response after Neoadjuvant Chemotherapy and Impact on Breast Cancer Recurrence and Survival: A Comprehensive Meta-analysis,” *Clinical Cancer Research*, vol. 26, no. 12, pp. 2838–2848, 2020, doi: 10.1158/1078-0432.CCR-19-3492.

[59] N. Goossens, S. Nakagawa, X. Sun, and Y. Hoshida, “Cancer biomarker discovery and validation.,” *Transl Cancer Res*, vol. 4, no. 3, pp. 256–269, 2015, doi: 10.3978/j.issn.2218-676X.2015.06.04.

[60] A. Sadeghi-Naini *et al.*, “Quantitative Ultrasound Evaluation of Tumor Cell Death Response in Locally Advanced Breast Cancer Patients Receiving Chemotherapy,” *Clinical Cancer Research*, vol. 19, no. 8, pp. 2163–2174, 2013, doi: 10.1158/1078-0432.CCR-12-2965.

[61] M. S. Bae, M. Seo, K. G. Kim, I.-A. Park, and W. K. Moon, “Quantitative MRI morphology of invasive breast cancer: correlation with immunohistochemical biomarkers and subtypes,” *Acta radiol*, vol. 56, no. 3, pp. 269–275, 2015, doi: 10.1177/0284185114524197.

[62] J. M. Boone, A. L. C. Kwan, K. Yang, G. W. Burkett, K. K. Lindfors, and T. R. Nelson, “Computed Tomography for Imaging the Breast,” *J Mammary Gland Biol Neoplasia*, vol. 11, no. 2, pp. 103–111, 2006, doi: 10.1007/s10911-006-9017-1.

[63] W. B. Eubank and D. A. Mankoff, “Current and future uses of positron emission tomography in breast cancer imaging,” *Semin Nucl Med*, vol. 34, no. 3, pp. 224–240, 2004, doi: 10.1053/j.semnuclmed.2004.03.007.

[64] R. Gupta, T. Kurc, A. Sharma, J. S. Almeida, and J. Saltz, “The Emergence of Pathomics,” *Curr Pathobiol Rep*, vol. 7, no. 3, pp. 73–84, 2019, doi: 10.1007/s40139-019-00200-x.

[65] W. T. Tran *et al.*, “Personalized Breast Cancer Treatments Using Artificial Intelligence in Radiomics and Pathomics,” *J Med Imaging Radiat Sci*, vol. 50, no. 4, pp. S32–S41, 2019, doi: 10.1016/j.jmir.2019.07.010.

[66] X. Wang *et al.*, “RaPtomics: integrating radiomic and pathomic features for predicting recurrence in early stage lung cancer,” in *Medical Imaging 2018: Digital Pathology*, 2018, p. 21. doi: 10.1117/12.2296646.

[67] N. A. Alymani, M. D. Smith, D. J. Williams, and R. D. Petty, “Predictive biomarkers for personalised anti-cancer drug use: Discovery to clinical implementation,” *Eur J Cancer*, vol. 46, no. 5, pp. 869–879, 2010, doi: 10.1016/j.ejca.2010.01.001.

[68] K. Bera, V. Velcheti, and A. Madabhushi, “Novel Quantitative Imaging for Predicting Response to Therapy: Techniques and Clinical Applications,” *American Society of Clinical Oncology Educational Book*, no. 38, pp. 1008–1018, 2018, doi: 10.1200/EDBK_199747.

[69] E. Scalco and G. Rizzo, “Texture analysis of medical images for radiotherapy applications,” *Br J Radiol*, vol. 90, no. 1070, p. 642, 2017, doi: 10.1259/bjr.20160642.

[70] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, “Machine Learning methods for Quantitative Radiomic Biomarkers,” *Sci Rep*, vol. 5, no. 1, p. 13087, 2015, doi: 10.1038/srep13087.

[71] D. Zhang, G. Lu, and others, “A comparative study on shape retrieval using Fourier descriptors with different shape signatures,” in *Proc. of international conference on intelligent multimedia and distance education (ICIMADE01)*, 2001, pp. 1–9.

[72] D. Zwillinger and S. Kokoska, *CRC standard probability and statistics tables and formulae*. Crc Press, 1999.

[73] A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, “Image biomarker standardisation initiative,” 2016, doi: 10.1148/radiol.2020191145.

[74] H. Sharma, N. Zerbe, S. Lohmann, K. Kayser, O. Hellwich, and P. Hufnagl, “A review of graph-based methods for image analysis in digital histopathology,” *Diagn Pathol*, vol. 1, no. 1, 2015.

[75] W. T. Tran *et al.*, “Personalized Breast Cancer Treatments Using Artificial Intelligence in Radiomics and Pathomics,” *J Med Imaging Radiat Sci*, vol. 50, no. 4, pp. S32–S41, 2019, doi: 10.1016/j.jmir.2019.07.010.

[76] G. L. Banna *et al.*, “The Promise of Digital Biopsy for the Prediction of Tumor Molecular Features and Clinical Outcomes Associated With Immunotherapy,” *Front Med (Lausanne)*, vol. 6, 2019, doi: 10.3389/fmed.2019.00172.

[77] F. Aeffner *et al.*, “Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association,” *J Pathol Inform*, vol. 10, no. 1, p. 9, 2019, doi: 10.4103/jpi.jpi_82_18.

[78] R. Gupta, T. Kurc, A. Sharma, J. S. Almeida, and J. Saltz, “The Emergence of Pathomics,” *Curr Pathobiol Rep*, vol. 7, no. 3, pp. 73–84, 2019, doi: 10.1007/s40139-019-00200-x.

[79] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," *IEEE Rev Biomed Eng*, vol. 2, pp. 147–171, 2009, doi: 10.1109/RBME.2009.2034865.

[80] M. Dabass, R. Vig, and S. Vashisth, "Review of Histopathological Image Segmentation via current Deep Learning Approaches," *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pp. 1–6, Dec. 2018, doi: 10.1109/CCAA.2018.8777616.

[81] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," *IEEE Rev Biomed Eng*, vol. 2, pp. 147–171, 2009, doi: 10.1109/RBME.2009.2034865.

[82] R. J. Chen *et al.*, "Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis," *IEEE Trans Med Imaging*, pp. 1–1, 2020, doi: 10.1109/TMI.2020.3021387.

[83] Y. Alzahrani and B. Boufama, "Biomedical Image Segmentation: A Survey," *SN Comput Sci*, vol. 2, no. 4, p. 310, 2021, doi: 10.1007/s42979-021-00704-7.

[84] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

[85] "2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)," *IEEE*. https://ieeexplore.ieee.org/xpl/conhome/7150573/proceeding (accessed Oct. 01, 2021).

[86] H. Noh, S. Hong, and B. Han, "Learning Deconvolution Network for Semantic Segmentation," *The IEEE International Conference on Computer Vision (ICCV)*, May 2015.

[87] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," *J Digit Imaging*, vol. 32, no. 4, pp. 582–596, Aug. 2019, doi: 10.1007/s10278-019-00227-x.

[88] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," *MIDL 2018 Conference*, 2018.

[89] J. Fang, Q. Zhou, and S. Wang, "Segmentation Technology of Nucleus Image Based on U-Net Network," *Sci Program*, vol. 2021, pp. 1–10, 2021, doi: 10.1155/2021/1892497.

[90] J. Arevalo, F. A. Gonzalez, R. Ramos-Pollan, J. L. Oliveira, and M. A. Guevara Lopez, "Convolutional neural networks for mammography mass lesion classification," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 797–800. doi: 10.1109/EMBC.2015.7318482.

[91] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model," *Sci Rep*, vol. 7, no. 1, p. 4172, 2017, doi: 10.1038/s41598-017-04075-z.

[92] 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. doi: 10.1109/IJCNN.2016.7727519.

[93] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[94] T. Araújo *et al.*, “Classification of breast cancer histology images using Convolutional Neural Networks,” *PLoS One*, vol. 12, no. 6, p. e0177544, 2017, doi: 10.1371/journal.pone.0177544.

[95] K. Nazeri, A. Aminpour, and M. Ebrahimi, “Two-Stage Convolutional Neural Network for Breast Cancer Histology Image Classification,” 2018, pp. 717–726. doi: 10.1007/978-3-319-93000-8_81.

[96] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, “DCAN: Deep contour-aware networks for object instance segmentation from histology images,” *Med Image Anal*, vol. 36, pp. 135–146, 2017, doi: 10.1016/j.media.2016.11.004.

[97] H. Ahmady Phoulady, D. B. Goldgof, L. O. Hall, and P. R. Mouton, “Nucleus segmentation in histology images with hierarchical multilevel thresholding,” 2016, p. 979111. doi: 10.1117/12.2216632.

[98] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images,” *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016, doi: 10.1109/TMI.2016.2538465.

[99] P. Afshar, A. Mohammadi, K. N. Plataniotis, A. Oikonomou, and H. Benali, “From Hand-Crafted to Deep Learning-based Cancer Radiomics: Challenges and Opportunities,” *IEEE Signal Process Mag*, pp. 132–160, 2018, doi: 10.1109/MSP.2019.2900993.

[100] A. R. Podgorsak *et al.*, “Automatic radiomic feature extraction using deep learning for angiographic parametric imaging of intracranial aneurysms,” *J Neurointerv Surg*, p. neurintsurg-2019-015214, 2019, doi: 10.1136/neurintsurg-2019-015214.

[101] P. Vaidya *et al.*, “RaPtomics - Integrating Radiomic and Pathomic Features for Predicting Recurrence in Early Stage Lung Cancer,” no. March 2018, 2019, doi: 10.1117/12.2296646.

[102] J. Saltz *et al.*, “Towards Generation, Management, and Exploration of Combined Radiomics and Pathomics Datasets for Cancer Research.,” *AMIA Jt Summits Transl Sci Proc*, vol. 2017, pp. 85–94, 2017.

[103] D. L. Pham, C. Xu, and J. L. Prince, “Current Methods in Medical Image Segmentation,” *Annu Rev Biomed Eng*, vol. 2, no. 1, pp. 315–337, 2000, doi: 10.1146/annurev.bioeng.2.1.315.

[104] A. Rehman, M. Ahmed Butt, and M. Zaman, “A Survey of Medical Image Analysis Using Deep Learning Approaches,” in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 1334–1342. doi: 10.1109/ICCMC51019.2021.9418385.

[105] K. G. Kim, “Book Review: Deep Learning,” in *Healthcare Informatics Research*, vol. 22, 2016, pp. 351–354. doi: 10.4258/hir.2016.22.4.351.

[106] A. Zadeh Shirazi, E. Fornaciari, N. S. Bagherian, L. M. Ebert, B. Koszyca, and G. A. Gomez, “DeepSurvNet: deep survival convolutional network for brain cancer survival rate classification based on histopathological images,” *Med Biol Eng Comput*, vol. 58, no. 5, pp. 1031–1045, 2020, doi: 10.1007/s11517-020-02147-3.

[107] K. Saednia, W. T. Tran, and A. Sadeghi-Naini, “Automatic characterization of breast lesions using multi-scale attention-guided deep learning of digital histology images,” *Comput Methods Biomech Biomed Eng Imaging Vis*, pp. 1–9, 2022, doi: 10.1080/21681163.2022.2058415.

[108] N. Coudray *et al.*, “Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning,” *Nat Med*, vol. 24, no. 10, pp. 1559–1567, 2018, doi: 10.1038/s41591-018-0177-5.

[109] A. Echle, N. T. Rindtorff, T. J. Brinker, T. Luedde, A. T. Pearson, and J. N. Kather, “Deep learning in cancer pathology: a new generation of clinical biomarkers,” *Br J Cancer*, vol. 124, no. 4, pp. 686–696, 2021, doi: 10.1038/s41416-020-01122-x.

[110] H. Sung *et al.*, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.

[111] American Cancer Society, “Breast Cancer.” https://www.cancer.org/cancer/breast-cancer.html (accessed Nov. 02, 2022).

[112] B. N. Joe and E. A. Sickles, “The Evolution of Breast Imaging: Past to Present,” *Radiology*, vol. 273, no. 2S, pp. S23–S44, Nov. 2014, doi: 10.1148/radiol.14141233.

[113] C. W. Elston and I. O. Ellis, “Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up.,” *Histopathology*, vol. 41, no. 3a, pp. 151–151, Sep. 2002, doi: 10.1046/j.1365-2559.2002.14691.x.

[114] 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. doi: 10.1109/IJCNN.2016.7727519.

[115] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med Image Anal*, vol. 42, no. Dec, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[116] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med Image Anal*, vol. 42, no. Dec, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[117] M. Z. Alom, C. Yakopcic, Mst. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network," *J Digit Imaging*, vol. 32, no. 4, pp. 605–617, Aug. 2019, doi: 10.1007/s10278-019-00182-7.

[118] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. W. Pluim, "Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images," *PLoS One*, vol. 8, no. 7, p. e70221, Jul. 2013, doi: 10.1371/journal.pone.0070221.

[119] J. W. Wei, L. J. Tafe, Y. A. Linnik, L. J. Vaickus, N. Tomita, and S. Hassanpour, "Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks," *Sci Rep*, vol. 9, no. 1, p. 3358, Dec. 2019, doi: 10.1038/s41598-019-40041-7.

[120] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 2424–2433.

[121] K. Nazeri, A. Aminpour, and M. Ebrahimi, "Two-Stage Convolutional Neural Network for Breast Cancer Histology Image Classification," *International Conference Image Analysis and Recognition*, pp. 717–726, 2018, doi: 10.1007/978-3-319-93000-8_81.

[122] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[123] T. Araújo *et al.*, "Classification of breast cancer histology images using Convolutional Neural Networks," *PLoS One*, vol. 12, no. 6, p. e0177544, 2017, doi: 10.1371/journal.pone.0177544.

[124] K. Nazeri, A. Aminpour, and M. Ebrahimi, "Two-Stage Convolutional Neural Network for Breast Cancer Histology Image Classification," *International Conference Image Analysis and Recognition*, pp. 717–726, 2018, doi: 10.1007/978-3-319-93000-8_81.

[125] C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo, and J. Liu, "Breast cancer histopathology image classification through assembling multiple compact CNNs," *BMC Med Inform Decis Mak*, vol. 19, no. 1, p. 198, Dec. 2019, doi: 10.1186/s12911-019-0913-x.

[126] "15th International Conference on Image Analysis, Recognition: ICIAR 2018 grand challenge."

[127] A. Golatkar, D. Anand, and A. Sethi, "Classification of Breast Cancer Histology Using Deep Learning," Springer, Cham, 2018, pp. 837–844. doi: 10.1007/978-3-319-93000-8_95.

[128] J. Vizcarra, R. Place, L. Tong, D. Gutman, and M. D. Wang, "Fusion In Breast Cancer Histology Classification," in *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 2019, pp. 485–493. doi: 10.1145/3307339.3342166.

[129] A. Vaswani *et al.*, "Attention is All you Need," *Adv Neural Inf Process Syst*, pp. 5998–6008, 2017.

[130] W. Wang and J. Shen, "Deep Visual Attention Prediction," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2368–2378, May 2018, doi: 10.1109/TIP.2017.2787612.

[131] A. Vaswani *et al.*, "Attention is All you Need," *Adv Neural Inf Process Syst*, pp. 5998–6008, 2017.

[132] J. Schlemper *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Med Image Anal*, vol. 53, pp. 197–207, Apr. 2019, doi: 10.1016/j.media.2019.01.012.

[133] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018.

[134] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *3rd International Conference on Learning Representations, ICLR 2015*.

[135] K. Xu *et al.*, "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention," *International conference on machine learning*, pp. 2048–2057, Feb. 2015.

[136] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141. doi: 10.1109/CVPR.2018.00745.

[137] L. Chen *et al.*, "SCA-CNN: Spatial and Channel-Wise Attention in Convolutional Networks for Image Captioning," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6298–6306. doi: 10.1109/CVPR.2017.667.

[138] N. Tomita, B. Abdollahi, J. Wei, B. Ren, A. Suriawinata, and S. Hassanpour, "Attention-Based Deep Neural Networks for Detection of Cancerous and Precancerous Esophagus Tissue on Histopathological Slides," *JAMA Netw Open*, vol. 2, no. 11, p. e1914645, Nov. 2019, doi: 10.1001/jamanetworkopen.2019.14645.

[139] M.-Y. Ban, W.-D. Tian, and Z.-Q. Zhao, "Real-Time Object Detection Based on Convolutional Block Attention Module," in *ICIC 2020. Lecture Notes in Computer Science*, 2020, pp. 41–50. doi: 10.1007/978-3-030-60796-8_4.

[140] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *3rd International Conference on Learning Representations (ICLR2015)*, 2015.

[141] "15th International Conference on Image Analysis, Recognition: ICIAR 2018 grand challenge."

[142] M. C. Seiler and F. A. Seiler, "Numerical Recipes in C: The Art of Scientific Computing," *Risk Analysis*, vol. 9, no. 3, pp. 415–416, Sep. 1989, doi: 10.1111/j.1539-6924.1989.tb01007.x.

[143] M. Lin, Q. Chen, and S. Yan, "Network In Network," *ICLR 2014 conference*, Dec. 2014.

[144] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning Deep Features for Discriminative Localization,” *IEEE conference on computer vision and pattern recognition*, pp. 2921–2929, 2016.

[145] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning Deep Features for Discriminative Localization,” *IEEE conference on computer vision and pattern recognition*, pp. 2921–2929, 2016.

[146] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *ICLR 2015 Conference*, Dec. 2015.

[147] K. Xu *et al.*, “Show, Attend and Tell: Neural Image Caption Generation with Visual Attention,” *International conference on machine learning*, pp. 2048–2057, 2015.

[148] A. Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network,” *Physica D*, Aug. 2018.

[149] D. Chicco, N. Tötsch, and G. Jurman, “The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation,” *BioData Min*, vol. 14, no. 1, p. 13, 2021, doi: 10.1186/s13040-021-00244-z.

[150] T. le Dinh, S.-G. Kwon, S.-H. Lee, and K.-R. Kwon, “Breast Tumor Cell Nuclei Segmentation in Histopathology Images using EfficientUnet++ and Multi-organ Transfer Learning,” *Journal of Korea Multimedia Society*, vol. 24, pp. 1000–1011, 2021.

[151] X. Xie, J. Chen, Y. Li, L. Shen, K. Ma, and Y. Zheng, “Instance-Aware Self-supervised Learning for Nuclei Segmentation,” *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020.*, vol. 12265, pp. 341–350, 2020, doi: 10.1007/978-3-030-59722-1_33.

[152] A. L. Martel, S. Nofech-Mozes, S. Salama, S. Akbar, and M. Peikari, “Assessment of Residual Breast Cancer Cellularity after Neoadjuvant Chemotherapy using Digital Pathology [Data set],” *The Cancer Imaging Archive*, 2019, doi: 10.7937/TCIA.2019.4YIBTJNO.

[153] N. Kumar *et al.*, “A Multi-Organ Nucleus Segmentation Challenge,” *IEEE Trans Med Imaging*, vol. 39, no. 5, pp. 1380–1391, 2020, doi: 10.1109/TMI.2019.2947628.

[154] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.*, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[155] R. Zhang, L. Du, Q. Xiao, and J. Liu, “Comparison of Backbones for Semantic Segmentation Network,” *J Phys Conf Ser*, vol. 1544, no. 1, p. 012196, 2020, doi: 10.1088/1742-6596/1544/1/012196.

[156] S. Jadon, “A survey of loss functions for semantic segmentation,” in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2020, pp. 1–7. doi: 10.1109/CIBCB48159.2020.9277638.

[157] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[158] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, 2012, vol. 25.

[159] C. E. DeSantis *et al.*, “Breast cancer statistics, 2019,” *CA Cancer J Clin*, vol. 69, no. 6, pp. 438–451, 2019, doi: 10.3322/caac.21583.

[160] A. Ahmad, *Breast Cancer Statistics: Recent Trends*. Springer, Cham, 2019. doi: 10.1007/978-3-030-20301-6_1.

[161] O. Falou *et al.*, “Evaluation of Neoadjuvant Chemotherapy Response in Women with Locally Advanced Breast Cancer Using Ultrasound Elastography,” *Transl Oncol*, vol. 6, no. 1, pp. 17–24, 2013, doi: 10.1593/tlo.12412.

[162] A. Sadeghi-Naini *et al.*, “Quantitative Ultrasound Evaluation of Tumor Cell Death Response in Locally Advanced Breast Cancer Patients Receiving Chemotherapy,” *Clinical Cancer Research*, vol. 19, no. 8, pp. 2163–2174, 2013, doi: 10.1158/1078-0432.CCR-12-2965.

[163] L. Sannachi *et al.*, “Breast Cancer Treatment Response Monitoring Using Quantitative Ultrasound and Texture Analysis: Comparative Analysis of Analytical Models,” *Transl Oncol*, vol. 12, no. 10, pp. 1271–1281, 2019, doi: 10.1016/j.tranon.2019.06.004.

[164] A. Sadeghi-Naini *et al.*, “Quantitative Ultrasound Evaluation of Tumor Cell Death Response in Locally Advanced Breast Cancer Patients Receiving Chemotherapy,” *Clinical Cancer Research*, vol. 19, no. 8, pp. 2163–2174, 2013, doi: 10.1158/1078-0432.CCR-12-2965.

[165] C. Sousa *et al.*, “Neoadjuvant radiotherapy in the approach of locally advanced breast cancer,” *ESMO Open*, vol. 5, no. 2, p. e000640, 2020, doi: 10.1136/esmoopen-2019-000640.

[166] S. M. Scholl *et al.*, “Neoadjuvant versus adjuvant chemotherapy in premenopausal patients with tumours considered too large for breast conserving surgery: Preliminary results of a randomised trial: S6,” *Eur J Cancer*, vol. 30, no. 5, pp. 645–652, 1994, doi: 10.1016/0959-8049(94)90537-1.

[167] S. Chuthapisith, J. M. Eremin, M. El-Sheemy, and O. Eremin, “Neoadjuvant chemotherapy in women with large and locally advanced breast cancer: Chemoresistance and prediction of response to drug therapy,” *The Surgeon*, vol. 4, no. 4, pp. 211–219, 2006, doi: 10.1016/S1479-666X(06)80062-4.

[168] G. N. Hortobagyi, “Comprehensive management of locally advanced breast cancer.,” *Cancer*, vol. 66, no. 6 Suppl, pp. 1387–91, 1990, doi: 10.1002/1097-0142(19900915)66:14+<1387::aid-cncr2820661414>3.0.co;2-i.

[169] D. Sethi, R. Sen, S. Parshad, J. Sen, S. Khetarpal, and M. Garg, “Histopathologic changes following neoadjuvant chemotherapy in locally advanced breast cancer,” *Indian J Cancer*, vol. 50, no. 1, p. 58, 2013, doi: 10.4103/0019-509X.112301.

[170] S. H. Giordano, “Update on Locally Advanced Breast Cancer,” *Oncologist*, vol. 8, no. 6, pp. 521–530, 2003, doi: 10.1634/theoncologist.8-6-521.

[171] A. Romero *et al.*, “Correlation between response to neoadjuvant chemotherapy and survival in locally advanced breast cancer patients.,” *Ann Oncol*, vol. 24, no. 3, pp. 655–61, 2013, doi: 10.1093/annonc/mds493.

[172] L. M. Spring *et al.*, “Pathologic Complete Response after Neoadjuvant Chemotherapy and Impact on Breast Cancer Recurrence and Survival: A Comprehensive Meta-analysis,” *Clinical Cancer Research*, vol. 26, no. 12, pp. 2838–2848, 2020, doi: 10.1158/1078-0432.CCR-19-3492.

[173] S. J. Cleator, A. Makris, S. E. Ashley, R. Lal, and T. J. Powles, “Good clinical response of breast cancers to neoadjuvant chemoendocrine therapy is associated with improved overall survival,” *Annals of Oncology*, vol. 16, no. 2, pp. 267–272, 2005, doi: 10.1093/annonc/mdi049.

[174] I. C. Smith *et al.*, “Neoadjuvant Chemotherapy in Breast Cancer: Significantly Enhanced Response With Docetaxel,” *Journal of Clinical Oncology*, vol. 20, no. 6, pp. 1456–1466, 2002, doi: 10.1200/JCO.2002.20.6.1456.

[175] B. dos Anjos Pultz, F. A. C. da Luz, P. R. de Faria, A. P. L. Oliveira, R. A. de Araújo, and M. J. B. Silva, “Far Beyond the Usual Biomarkers in Breast Cancer: A Review,” *J Cancer*, vol. 5, no. 7, pp. 559–571, 2014, doi: 10.7150/jca.8925.

[176] F. Aeffner *et al.*, “Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association,” *J Pathol Inform*, vol. 10, no. 1, p. 9, 2019, doi: 10.4103/jpi.jpi_82_18.

[177] O. Jimenez-del-Toro *et al.*, “Analysis of Histopathology Images:From Traditional Machine Learning to Deep Learning,” in *Biomedical Texture Analysis*, Elsevier, 2017, pp. 281–314. doi: 10.1016/B978-0-12-812133-7.00010-7.

[178] P. Vaidya *et al.*, “RaPtomics - Integrating Radiomic and Pathomic Features for Predicting Recurrence in Early Stage Lung Cancer,” *Medical Imaging 2018: Digital Pathology*, vol. 10581, no. March 2018, p. 105810M, 2019, doi: 10.1117/12.2296646.

[179] J. Saltz *et al.*, “Towards Generation, Management, and Exploration of Combined Radiomics and Pathomics Datasets for Cancer Research.,” *AMIA Jt Summits Transl Sci Proc*, vol. 2017, pp. 85–94, 2017.

[180] W. F. Symmans *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, doi: 10.1200/JCO.2007.10.6823.

[181] W. F. Symmans *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, doi: 10.1200/JCO.2007.10.6823.

[182] A. L. Martel *et al.*, “An Image Analysis Resource for Cancer Research: PIIP—Pathology Image Informatics Platform for Visualization, Analysis, and Management,” *Cancer Res*, vol. 77, no. 21, pp. e83–e86, 2017, doi: 10.1158/0008-5472.CAN-17-0323.

[183] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241.

[184] A. L. Martel, S. Nofech-Mozes, S. Salama, S. Akbar, and M. Peikari, “Assessment of Residual Breast Cancer Cellularity after Neoadjuvant Chemotherapy using Digital Pathology [Data set].,” *The Cancer Imaging Archive.*, 2019, doi: 10.7937/TCIA.2019.4YIBTJNO.

[185] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, “A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology,” *IEEE Trans Med Imaging*, vol. 36, no. 7, pp. 1550–1560, 2017, doi: 10.1109/TMI.2017.2677499.

[186] D. A. Gutman *et al.*, “The Digital Slide Archive: A Software Platform for Management, Integration, and Analysis of Histology for Cancer Research,” *Cancer Res*, vol. 77, no. 21, pp. e75–e78, 2017, doi: 10.1158/0008-5472.CAN-17-0629.

[187] J. J. M. van Griethuysen *et al.*, “Computational Radiomics System to Decode the Radiographic Phenotype,” *Cancer Res*, vol. 77, no. 21, pp. e104–e107, 2017, doi: 10.1158/0008-5472.CAN-17-0339.

[188] Zwillinger, Daniel, and S. Kokoska, “CRC standard probability and statistics tables and formulae,” *Crc Press*, 1999.

[189] A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, “Image biomarker standardisation initiative,” *Radiology*, 2020, doi: 10.1148/radiol.2020191145.

[190] S. Bhattacharjee *et al.*, “Multi-Features Classification of Prostate Carcinoma Observed in Histological Sections: Analysis of Wavelet-Based Texture and Colour Features,” *Cancers (Basel)*, vol. 11, no. 12, p. 1937, 2019, doi: 10.3390/cancers11121937.

[191] T. Chen and C. Guestrin, “XGBoost,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[192] N. v. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[193] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, “Automated grading of breast cancer histopathology using spectral clusteringwith textural and architectural image

features," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2008, pp. 496–499. doi: 10.1109/ISBI.2008.4541041.

[194] C. Sun and W. G. Wee, "Neighboring gray level dependence matrix for texture classification," *Comput Vis Graph Image Process*, vol. 23, no. 3, pp. 341–352, 1983, doi: 10.1016/0734-189X(83)90032-4.

[195] A. Heindl, S. Nawaz, and Y. Yuan, "Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology," *Laboratory Investigation*, vol. 95, no. 4, pp. 377–384, 2015, doi: 10.1038/labinvest.2014.155.

[196] R. J. Chen *et al.*, "Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis," *IEEE Trans Med Imaging*, pp. 1–1, 2020, doi: 10.1109/TMI.2020.3021387.

[197] P. Mobadersany *et al.*, "Predicting cancer outcomes from histology and genomics using convolutional networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 13, pp. E2970–E2979, 2018, doi: 10.1073/pnas.1717139115.

[198] S. Ha, S. Park, J.-I. Bang, E.-K. Kim, and H.-Y. Lee, "Metabolic Radiomics for Pretreatment 18F-FDG PET/CT to Characterize Locally Advanced Breast Cancer: Histopathologic Characteristics, Response to Neoadjuvant Chemotherapy, and Prognosis," *Sci Rep*, vol. 7, no. 1, p. 1556, 2017, doi: 10.1038/s41598-017-01524-7.

[199] H. Moghadas-Dastjerdi, H. R. Sha-E-Tallat, L. Sannachi, A. Sadeghi-Naini, and G. J. Czarnota, "A priori prediction of tumour response to neoadjuvant chemotherapy in breast cancer patients using quantitative CT and machine learning," *Sci Rep*, vol. 10, no. 1, p. 10936, 2020, doi: 10.1038/s41598-020-67823-8.

[200] C. Kolios *et al.*, "MRI texture features from tumor core and margin in the prediction of response to neoadjuvant chemotherapy in patients with locally advanced breast cancer," *Oncotarget*, vol. 12, no. 14, pp. 1354–1365, 2021, doi: 10.18632/oncotarget.28002.

[201] B. Zhao *et al.*, "Reproducibility of radiomics for deciphering tumor phenotype with imaging," *Sci Rep*, vol. 6, no. 1, p. 23428, 2016, doi: 10.1038/srep23428.

[202] A. Pfob *et al.*, "Intelligent Vacuum-Assisted Biopsy to Identify Breast Cancer Patients With Pathologic Complete Response (ypT0 and ypN0) After Neoadjuvant Systemic Treatment for Omission of Breast and Axillary Surgery," *Journal of Clinical Oncology*, 2022, doi: 10.1200/JCO.21.02439.

[203] A. Pfob *et al.*, "Identification of breast cancer patients with pathologic complete response in the breast after neoadjuvant systemic treatment by an intelligent vacuum-assisted biopsy," *Eur J Cancer*, vol. 143, pp. 134–146, 2021, doi: 10.1016/j.ejca.2020.11.006.

[204] K. Tryfonidis, E. Senkus, M. J. Cardoso, and F. Cardoso, "Management of locally advanced breast cancer—perspectives and future directions," *Nat Rev Clin Oncol*, vol. 12, no. 3, pp. 147–162, 2015, doi: 10.1038/nrclinonc.2015.13.

[205] S. H. Giordano, “Update on Locally Advanced Breast Cancer,” *Oncologist*, vol. 8, no. 6, pp. 521–530, 2003, doi: 10.1634/theoncologist.8-6-521.

[206] M. Dhanushkodi *et al.*, “Locally Advanced Breast Cancer (LABC): Real-World Outcome of Patients From Cancer Institute, Chennai,” *JCO Glob Oncol*, no. 7, pp. 767–781, 2021, doi: 10.1200/GO.21.00001.

[207] O. Falou *et al.*, “Evaluation of Neoadjuvant Chemotherapy Response in Women with Locally Advanced Breast Cancer Using Ultrasound Elastography,” *Transl Oncol*, vol. 6, no. 1, pp. 17–24, 2013, doi: 10.1593/tlo.12412.

[208] L. Sannachi *et al.*, “Breast Cancer Treatment Response Monitoring Using Quantitative Ultrasound and Texture Analysis: Comparative Analysis of Analytical Models,” *Transl Oncol*, vol. 12, no. 10, pp. 1271–1281, 2019, doi: 10.1016/j.tranon.2019.06.004.

[209] A. Pfob *et al.*, “Identification of breast cancer patients with pathologic complete response in the breast after neoadjuvant systemic treatment by an intelligent vacuum-assisted biopsy,” *Eur J Cancer*, vol. 143, pp. 134–146, 2021, doi: 10.1016/j.ejca.2020.11.006.

[210] N. M. Rueth *et al.*, “Underuse of Trimodality Treatment Affects Survival for Patients With Inflammatory Breast Cancer: An Analysis of Treatment and Survival Trends From the National Cancer Database,” *Journal of Clinical Oncology*, vol. 32, no. 19, pp. 2018–2024, 2014, doi: 10.1200/JCO.2014.55.1978.

[211] S. M. Scholl *et al.*, “Neoadjuvant versus adjuvant chemotherapy in premenopausal patients with tumours considered too large for breast conserving surgery: Preliminary results of a randomised trial: S6,” *Eur J Cancer*, vol. 30, no. 5, pp. 645–652, 1994, doi: 10.1016/0959-8049(94)90537-1.

[212] X. Meng, X. Chang, X. Wang, and Y. Guo, “Development and Validation a Survival Prediction Model and a Risk Stratification for Elderly Locally Advanced Breast Cancer,” *Clin Breast Cancer*, 2022, doi: 10.1016/j.clbc.2022.06.002.

[213] D. Sethi, R. Sen, S. Parshad, J. Sen, S. Khetarpal, and M. Garg, “Histopathologic changes following neoadjuvant chemotherapy in locally advanced breast cancer,” *Indian J Cancer*, vol. 50, no. 1, p. 58, 2013, doi: 10.4103/0019-509X.112301.

[214] W. Haque, V. Verma, S. Hatch, V. Suzanne Klimberg, E. Brian Butler, and B. S. Teh, “Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy,” *Breast Cancer Res Treat*, vol. 170, no. 3, pp. 559–567, 2018, doi: 10.1007/s10549-018-4801-3.

[215] A. Tudorica *et al.*, “Early Prediction and Evaluation of Breast Cancer Response to Neoadjuvant Chemotherapy Using Quantitative DCE-MRI,” *Transl Oncol*, vol. 9, no. 1, pp. 8–17, 2016, doi: 10.1016/j.tranon.2015.11.016.

[216] A. Tahmassebi *et al.*, “Impact of Machine Learning With Multiparametric Magnetic Resonance Imaging of the Breast for Early Prediction of Response to Neoadjuvant Chemotherapy

and Survival Outcomes in Breast Cancer Patients," *Invest Radiol*, vol. 54, no. 2, pp. 110–117, 2019, doi: 10.1097/RLI.0000000000000518.

[217] H. Moghadas-Dastjerdi, H. R. Sha-E-Tallat, L. Sannachi, A. Sadeghi-Naini, and G. J. Czarnota, "A priori prediction of tumour response to neoadjuvant chemotherapy in breast cancer patients using quantitative CT and machine learning," *Sci Rep*, vol. 10, no. 10936, 2020, doi: 10.1038/s41598-020-67823-8.

[218] H. Taleghamar, S. A. Jalalifar, G. J. Czarnota, and A. Sadeghi-Naini, "Deep learning of quantitative ultrasound multi-parametric images at pre-treatment to predict breast cancer response to chemotherapy," *Sci Rep*, vol. 12, no. 1, p. 2244, 2022, doi: 10.1038/s41598-022-06100-2.

[219] W. T. Tran *et al.*, "Predicting breast cancer response to neoadjuvant chemotherapy using pretreatment diffuse optical spectroscopic texture analysis.," *Br J Cancer*, vol. 116, no. 10, pp. 1329–1339, 2017, doi: 10.1038/bjc.2017.97.

[220] H. Taleghamar, H. Moghadas-Dastjerdi, G. J. Czarnota, and A. Sadeghi-Naini, "Characterizing intra-tumor regions on quantitative ultrasound parametric images to predict breast cancer response to chemotherapy at pre-treatment.," *Sci Rep*, vol. 11, no. 1, p. 14865, 2021, doi: 10.1038/s41598-021-94004-y.

[221] M. Hayashi, Y. Yamamoto, and H. Iwase, "Clinical imaging for the prediction of neoadjuvant chemotherapy response in breast cancer.," *Chin Clin Oncol*, vol. 9, no. 3, p. 31, 2020, doi: 10.21037/cco-20-15.

[222] V. Romeo *et al.*, "Assessment and Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer: A Comparison of Imaging Modalities and Future Perspectives.," *Cancers (Basel)*, vol. 13, no. 14, 2021, doi: 10.3390/cancers13143521.

[223] N. Dimitriou, O. Arandjelović, and P. D. Caie, "Deep Learning for Whole Slide Image Analysis: An Overview," *Front Med (Lausanne)*, vol. 6, no. 264, 2019, doi: 10.3389/fmed.2019.00264.

[224] D. Clunie *et al.*, "Digital Imaging and Communications in Medicine Whole Slide Imaging Connectathon at Digital Pathology Association Pathology Visions 2017," *J Pathol Inform*, vol. 9, no. 1, p. 6, 2018, doi: 10.4103/jpi.jpi_1_18.

[225] F. Chollet, "Xception: Deep Learning With Depthwise Separable Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[226] 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 (CVPR)*, 2016.

[227] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009, pp. 248–255.

[228] A. Vaswani *et al.*, “Attention is All you Need,” in *Advances in Neural Information Processing Systems*, 2017, vol. 30, pp. 5998–6008.

[229] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” in *Third International Conference on Learning Representations (ICLR2015)*, 2015.

[230] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *International Conference on Learning Representations*, 2021.

[231] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in Vision: A Survey,” *ACM Comput Surv*, vol. 54, no. 10s, pp. 1–41, 2022, doi: 10.1145/3505244.

[232] S. Paul and P.-Y. Chen, “Vision Transformers Are Robust Learners,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, pp. 2071–2081, 2022, doi: 10.1609/aaai.v36i2.20103.

[233] Z. Dai, H. Liu, Q. v Le, and M. Tan, “CoAtNet: Marrying Convolution and Attention for All Data Sizes,” in *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 3965–3977.

[234] W. F. Symmans *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, doi: 10.1200/JCO.2007.10.6823.

[235] A. L. Martel *et al.*, “An Image Analysis Resource for Cancer Research: PIIP—Pathology Image Informatics Platform for Visualization, Analysis, and Management,” *Cancer Res*, vol. 77, no. 21, 2017, doi: 10.1158/0008-5472.CAN-17-0323.

[236] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74.

[237] K. Saednia *et al.*, “Quantitative digital histopathology and machine learning to predict pathological complete response to chemotherapy in breast cancer patients using pre-treatment tumor biopsies,” *Sci Rep*, vol. 12, no. 1, p. 9690, 2022, doi: 10.1038/s41598-022-13917-4.

[238] H. Duanmu *et al.*, “A spatial attention guided deep learning system for prediction of pathological complete response using breast cancer histopathology images,” *Bioinformatics*, vol. 38, no. 19, pp. 4605–4612, 2022, doi: 10.1093/bioinformatics/btac558.

[239] A. Lagree *et al.*, “Assessment of Digital Pathology Imaging Biomarkers Associated with Breast Cancer Histologic Grade.,” *Curr Oncol*, vol. 28, no. 6, pp. 4298–4316, 2021, doi: 10.3390/curroncol28060366.

[240] W. Bulten *et al.*, “Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge.,” *Nat Med*, vol. 28, no. 1, pp. 154–163, 2022, doi: 10.1038/s41591-021-01620-2.

[241] A. Taylor-Weiner *et al.*, “A Machine Learning Approach Enables Quantitative Measurement of Liver Histology and Disease Monitoring in NASH.,” *Hepatology*, vol. 74, no. 1, pp. 133–147, 2021, doi: 10.1002/hep.31750.

[242] W. S. Shim *et al.*, “DeepRePath: Identifying the Prognostic Features of Early-Stage Lung Adenocarcinoma Using Multi-Scale Pathology Images and Deep Convolutional Neural Networks.,” *Cancers (Basel)*, vol. 13, no. 13, 2021, doi: 10.3390/cancers13133308.

[243] A. Singh and P. Singh, “Image Classification: A Survey,” *Journal of Informatics Electrical and Electronics Engineering (JIEEE)*, vol. 1, no. 2, pp. 1–9, 2020, doi: 10.54060/JIEEE/001.02.002.

[244] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in Vision: A Survey,” *ACM Comput Surv*, vol. 54, no. 10s, pp. 1–41, 2022, doi: 10.1145/3505244.

[245] Z. Chen *et al.*, “Vision Transformer Adapter for Dense Predictions,” 2022.

[246] R. J. Chen *et al.*, “Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning,” 2022.

[247] T. Iwamoto and L. Pusztai, “Predicting prognosis of breast cancer with gene signatures: are we lost in a sea of data?,” *Genome Med*, vol. 2, no. 11, p. 81, 2010, doi: 10.1186/gm202.

[248] N. Kumar *et al.*, “A Multi-Organ Nucleus Segmentation Challenge,” *IEEE Trans Med Imaging*, vol. 39, no. 5, pp. 1380–1391, 2020, doi: 10.1109/TMI.2019.2947628.

[249] Y. Zhou, O. F. Onder, Q. Dou, E. Tsougenis, H. Chen, and P.-A. Heng, “CIA-Net: Robust Nuclei Instance Segmentation with Contour-Aware Information Aggregation,” *International Conference on Information Processing in Medical Imaging*, pp. 682–693, 2019, doi: 10.1007/978-3-030-20351-1_53.

[250] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, vol. 11045, pp. 3–11, 2018, doi: 10.1007/978-3-030-00889-5_1.

[251] X. Xie, J. Chen, Y. Li, L. Shen, K. Ma, and Y. Zheng, “Instance-Aware Self-supervised Learning for Nuclei Segmentation,” *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pp. 341–350, 2020, doi: 10.1007/978-3-030-59722-1_33.