An-Najah National University
Faculty of Graduate Studies

COMPUTER VISION AND MACHINE LEARNING APPROACHES FOR INTERPRETATION OF BREAST HISTOPATHOLOGY

By
Omar Husni Mustafa Odeh

Supervisor
Dr. Anas Toma

This Thesis is Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Artificial Intelligence, Faculty of Graduate Studies, An-Najah National University, Nablus - Palestine.

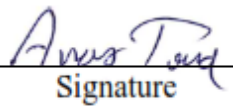
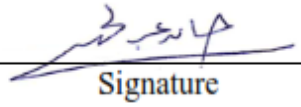
2024

COMPUTER VISION AND MACHINE LEARNING APPROACHES FOR INTERPRETATION OF BREAST HISTOPATHOLOGY

By

Omar Husni Mustafa Odeh

This Thesis was defended successfully on 12/09/2024 and approved by:

Dr. Anas Toma Supervisor	 Signature
Dr. Suhail Odeh External Examiner	 Signature
Dr. Hamed Abdelhaq Internal Examiner	 Signature

Dedication

To my beloved parents,
for their unconditional love and support,
and to all my family and friends
who have been a constant source of strength and
inspiration.

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Dr. Anas Toma, for his invaluable guidance, support, and expertise throughout the course of this research. His insights and suggestions have been key in shaping this work.

I am also deeply thankful to Dr. Hanood Abu Rass for her assistance and contributions to the medical aspects of this study. Her knowledge and expertise have been immensely beneficial. Furthermore, my appreciation extends to the entire team at the Histopathology Department. Their support in data collection and willingness to share their expertise played an important role in the completion of this research.

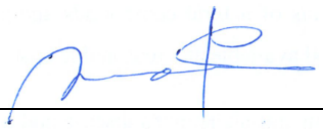
Declaration

I, the undersigned, declare that I submitted the thesis entitled:

COMPUTER VISION AND MACHINE LEARNING APPROACHES FOR INTERPRETATION OF BREAST HISTOPATHOLOGY

I declare that the work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification.

Student's Name Omar Husni Mustafa Odeh

Signature: 

Date: 12/09/2024

List of Contents

List of Tables

List of Figures

List of Appendices

COMPUTER VISION AND MACHINE LEARNING APPROACHES FOR INTERPRETATION OF BREAST HISTOPATHOLOGY

By
Omar Husni Mustafa Odeh
Supervisor
Anas Toma

Abstract

Breast cancer, along with other cancer types, represents a major health issue worldwide, especially in regions with limited medical resources, such as Palestine where it's harder to diagnose and treat cancer patients. The traditional diagnostic methods which rely mainly on the manual examination of histology slides in the lab by experts are both time consuming and require high skills, which often leads to delays in diagnosis and treatment. This research aims to leverage Whole Slide Imaging (WSI) and advanced image processing techniques, including machine learning and deep learning, to develop a tool to support breast cancer diagnosis in Palestine. A collection of histopathology images from breast cancer cases, provided by An-Najah Hospital in Nablus, Palestine, was used in this study. An automated classification model was developed, capable of identifying normal, benign, in situ, and invasive breast cancer tissues at magnifications of x20 and x40. It was empowered using advanced preprocessing techniques to normalize the colors and sizes of the training data. The classification model resulted in a remarkable performance of InceptionV3 and Gradient Boosting models with accuracies of up to 98.7% and 98% respectively in different settings. The same algorithm was applied to a BACH challenge dataset, resulting in a high accuracy of 96.7%, surpassing the results of other studies. The study also covers whole slide image (WSI) analysis, emphasizing the complex approach required for the proper detection of malignancy and tumor size estimation using a customized segmentation approach which resulted in an 84% Jaccard index score. Additionally, it discusses novel approaches in localization of tumor beds and grading of treatment responses, aiming to provide a methodological basis for improving patient care outcomes in challenging environments like Palestine.

Keywords: Breast Cancer Diagnosis, Automated Image Analysis, Whole Slide Imaging (WSI), Deep Learning in Pathology, Healthcare in Palestine.

Chapter One

Introduction and Theoretical Background

1.1 Introduction

Breast cancer is the most prevalent cancer worldwide and a leading cause of death among women (1). This issue is particularly critical in Palestine, where it accounts for 36.6 % of all female-specific cancer cases, as illustrated in Figure A.1 Appendix A and, 18.2 % of cancer cases across both genders. It also contributes to 11.3 % of total cancer deaths, according to the Palestinian Ministry of Health (2). Traditional manual diagnosis methods, based on examining histology images, present challenges due to their time-intensive nature and the high expertise required. Besides, the shortage of medical graduates specializing in pathology in Palestine aggravates these challenges, leading to delays in diagnosis and treatment (3).

To understand the important role of histology in breast cancer diagnosis, the nature of histology slides must be explored. They are thin parts of tissues or samples that are placed on a glass slide, produced by a series of procedures that include fixation, embedding, cutting, and staining. This process is important in staining the cells in preparation for microscopy examination. These slides are used by pathologists to evaluate the tissue structure and the detail of cells, which are the key factors in the diagnosis of cancerous alterations and the assessment of the tumor grade (4).

Microscopy of histology slides is a fundamental step in diagnosing breast cancer. In this case, pathologists use microscopes that provide magnification powers ranging from 4X to 40X to examine tissue and cellular structures. This close examination assists in identifying the general features of cancer such as absence of tissue differentiation, changes in cell size and form, and abnormal nuclei. Detecting these anomalies is very helpful in identifying the cancer, type, grading, and consequently finding the correct diagnosis and the most appropriate treatment methods. This is a broad analysis of the cellular and tissue organization of breast cancer through the use of microscopy and its application in the diagnosis and treatment of the disease (5).

The breakthrough of WSI has brought a new way in which pathologists make diagnoses. WSI is a technique that involves to the subsequent process of slide conversion into high-resolution digital images, allowing pathologists to examine these slides on their computers (6). This invention contributes several advantages, such as the simplified sharing of images, digital storage, and using image analyzing and machine learning methods (7). For instance, in areas such as Palestine where patients with cancer are deprived of being screened by advanced diagnostic tools, WSI will play a great role in improving cancer care.

The diagnostic issues in Palestine have been addressed by the recent progress in automatic image analysis and machine learning, including deep learning approaches, which have brought new prospects in breast cancer research. Revolutionary technologies include histology image interpretation, tumor growth assessment, and treatment response evaluation. This improvement in the diagnostic tools is important providing better quality of treatment planning choices for the patients (8).

By WSI, which lies at the forefront of breast cancer diagnosis, dealing with invasive breast cancer is no longer a problem. Digital scanning of whole-section histologic slides is a technology that permits 360-degree tissue excision sampling from slides. This technique allows for scrutinizing the distribution of cancerous cells and their shape which will provide precise diagnoses of the stage of cancer invasion. WSI plays a major role in the early detection of breast cancer, marking a significant stride in diagnostic precision (9).

The identification of tumor beds in histology images, an important part of the diagnostic process, is used for assessing the extent of cancer. Proper localization of the area involved is necessary for preparing surgical and therapeutic measures. In addition, the capacity to categorize treatment responses by image analysis has an enormous value, revealing the effectiveness of treatments and helping oncologists make proper decisions (10).

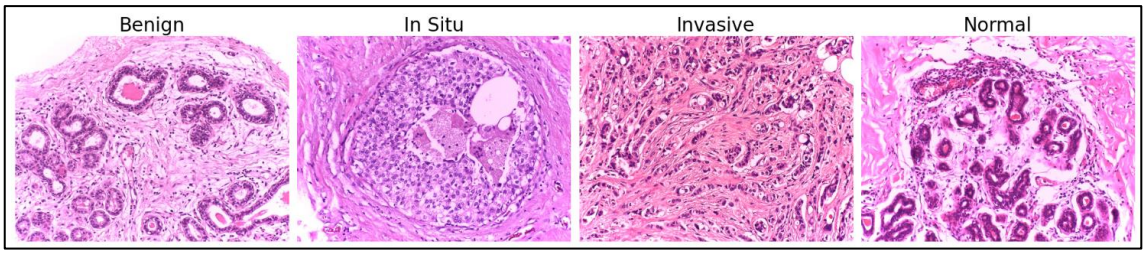
The method of identifying the tumor bed or malignancy involves applying a mask to the histology slide where the area of interest is emphasized. A mask as shown in Figure A.2 in Appendix A, in this context, is a black and white image (binary image) that outlines the region of interest, distinguishing the tumor bed or malignancy from the background. This is accompanied by the creation of algorithms that can effectively identify and measure the tumor bed or malignancy region (11).

The dataset used in this study is based on a primary dataset provided by An-Najah University which is a collection of breast cancer histology images divided into four categories (invasive, *in situ*, benign, and normal). This dataset will be very useful in the training and validation of the classification model. Furthermore, more data from different sources will give a broader view of tumor extent and response to therapy.

In the case of breast cancer, as shown in Figure 1.1, the terms *in situ*, benign, invasive, and normal are used to describe the state of cells and tissues. When cancerous cells are still localized in the place they originated from, it is called "*in situ*", for example Ductal Carcinoma *In Situ* (DCIS) where the abnormal cells are found in the lining of a breast duct but have not spread (12). The word "benign" is used to describe non-cancerous growths that do not invade nearby tissues or spread to other parts of the body, for example, fibroadenomas, which are benign breast tumors that are not dangerous and are frequently found in women. "Invasive" means that the cancer cells have penetrated beyond the ducts and have invaded the surrounding breast tissue, and Invasive Ductal Carcinoma (IDC) is the most prevalent type of breast cancer. "Normal" means that there are no pathological changes in the breast tissue, no cancerous formations, and the cells are morphologically and functionally fully normal. It is important to recognize these differences in order to diagnose, treat, and manage breast issues appropriately (13).

Figure 1.1

Breast tissue classification



This thesis begins with an introduction to the research problem, outlining the key objectives and the significance of the study. Following this, a comprehensive literature review is presented, summarizing existing research in the field and identifying gaps that this work aims to address. The methodology section then details the approaches and techniques employed in the research, including any specific models or custom methods used. The results section offers a snapshot of the anticipated findings and their relevance.

Finally, the conclusion summarizes the main outcomes of the research and provides recommendations for future work or practical applications.

In summary, this thesis attempts to utilize the capabilities of automatic image analysis and classification methods to enhance the diagnosis and treatment of breast cancer in Palestine. By using these state-of-the-art technologies, the aim is to improve patient care and results significantly. The potential of these methods as important aids for doctors to have early, accurate diagnosis, and effective treatment plans is huge. This kind of improvement not only improves the capacities of medical workers but also plays a major role in reducing breast cancer mortality rates. This research is a significant contribution to the area of medical technology and has far reaching effects on public health in Palestine and the world at large.

1.2 Problem Statement

Manual diagnosis of Breast cancer is a laborious process that consumes a lot of effort and time. As a result, many pathologists across the world get overloaded by the number of slides in the lab. They must take their own time to make sure of a correct diagnosis. Another delaying factor is that only 2% of medical graduates specialize in pathology. In addition, there is a lot of useful information regarding whole-slide imaging that a human reader cannot fully extract.

The identification of malignancy in whole slide images is one of the critical steps of breast cancer diagnosis. This is useful in the identification of the disease and coming up with proper treatment plans for the same. However, the level of precision and speed necessary for the best treatment planning goes beyond what the present manual methods can offer. AI-enabled systems can assist doctors in providing quick assessments of the invasiveness of a tumor.

The identification of therapeutic interferences is one of the areas of practice that can benefit from the application of automated tools. They serve as an aid to doctors in the evaluation and classification of patient's responses to therapy, an important issue in the development of individualized programs. This involves assessing the response of a patient's cancer to certain treatments, which in turn involves decisions about further care and alterations to the treatment regimen.

To conclude, the creation of automated systems for the diagnosis of breast cancer is not only a great innovation in technology but also a helping tool for doctors. They offer important aid by helping doctors evaluate and control patient care more efficiently.

1.3 Aims of the Study

- Develop an Automated Classification Model for breast cancer histology images at magnifications of x20 and x40 using deep learning techniques and feature extraction methods to classify images into normal, benign, *in situ*, and invasive categories. Specifically, the model will be applied to the new An-Najah dataset.
- Analyze Whole Slide Images for full tumor assessment, accurately determining the Malignant.
- Classify, detect, and evaluate therapy response for various breast cancer therapies, aiding in personalized treatment planning.

1.4 Research questions

- What are the most effective deep learning techniques for classifying breast cancer histology images into normal, benign, *in situ*, and invasive categories?
- How does the performance of automated classification models developed using the An-Najah dataset compare with existing models trained on other datasets?
- What role do feature extraction methods play in enhancing the accuracy of breast cancer histology image classification, and how can these methods be optimized?
- How can deep learning algorithms be optimized for the effective detection of malignancy in Whole Slide Images (WSI), considering the data size and intricate details present in these images?
- What is the importance of developing accurate deep learning models for the classification of therapy responses in cancer treatment?

1.5 Importance of the Study

- Development of an automatic classification model to increase the speed of breast cancer diagnosis based on histology images reducing the workload on the pathologists.

- Enhancing Tumor Analysis: Enhancement of the nuclei segmentation methods in high-resolution histology images for a deeper understanding of tumor biology that will enable accurate diagnosis and comprehension of the tumor behavior.
- Tumor Assessment: Image analysis of the whole slide for evidence of tumor spread and heterogeneity is critical for proper staging of the disease and targeting of treatment strategies.
- Optimizing Treatment Planning: Categorization of therapy responses to develop targeted treatment strategies, and hence, more effective patient-specific treatments.
- Providing a dataset for the classification of histopathology tissue into four categories: *in situ*, invasive, normal, and benign. This dataset is designed to be globally accessible for use in various research studies.

1.6 Hypotheses of the Study

The hypotheses of this study are aligned with the specific objectives:

- Automated Classification Model Efficiency: It is hypothesized that the employment of Convolutional Neural Network (CNN) and deep learning approaches in histology image classification will significantly enhance the diagnostic precision and efficacy.
- Detail-Oriented Whole Slide Analysis: Whole slide image analysis is supposed to provide a full assessment of tumor spread and morphology, which will influence treatment decisions and planning.
- Effective Therapy Response Classification: It is believed that precise classification of therapy responses by modern methods will result in the development of more personalized and effective treatment approaches that will improve the outcomes in breast cancer patients.

1.7 Overview of Methodology

The methodology of this research is divided into three parts, each targeting a specific aspect of breast cancer diagnosis and treatment evaluation using histology images:

- Classification and Detailed Analysis of Histology Images at 20X and 40X Magnifications: This method of image categorizing involves labeling these histology images as normal, benign, *in situ*, and invasive at 20X magnification, with the help

of advanced deep learning techniques like convolutional neural networks (CNNs) for more accurate classification. Moreover, at 40X magnification, the view changes to the accurate segmentation of nuclei and, subsequently, the extraction of features, which in turn are used to classify the image. The enumerated points are the main pieces for taking the diagnostic tests and performing detailed analysis at the cell level.

- Whole Slide Image Analysis: This part talks about the analysis of whole slide images, using precise deep learning models like NasteNet, U-Net, PSPNet, and ENet to identify invasive areas, as well. The main objective is to judge the Malignancy and monitor the spatial distribution of cancer tissues.
- Therapy Response Classification: The third section of the paper is dedicated to the evaluation of the efficacy of treatment methods. The research aims to distinguish between the effects of therapy, according to patients' responses, which helps therapists in adjusting their treatments to reach the best possible therapy for each patient.

1.8 Methods and techniques

This section defines the different methods and approaches that were used in this study. It describes the main algorithm, the pre-processing steps, the auxiliary algorithms used in some parts, and the comparison of the used architectures.

1.8.1 Preprocessing and color Normalization

Color normalization is a technique that is used in image processing to make all the images look similar by altering the color characteristics. This technique is useful in our applications where color representation consistency is important (14).

The approach presented by Reinhard et al. works with the color space of images that are converted to the L*a*b* color space; the L* is the lightness, whereas the a* and b* are chromaticity coordinates. This color space is more perceptually uniform and, therefore, appropriate for color manipulation (14).

1.8.2 Image augmentation

Image augmentation is a technique of creating a new training dataset from the existing dataset by applying various transformations to the images. Some are as basic as flipping images, which refers to flipping them from the current position to a horizontal or a vertical position, and rescaling, which is just zooming in and out. Some of these include shifting the image horizontally or vertically, or, in other words, rotating the image by shifting the pixel intensity along the X or Y axes, changing the brightness and contrast of the image, adding noise, and cropping the image at different positions. These techniques help in creating new versions of the images and hence, the model does not get overfitted to the unseen data (15).

Some of the augmentation techniques include shearing, which involves rotating the images, and color jittering which adjusts changing the color, saturation, and brightness of images. It is also possible to apply such operations as blurring or sharpening, elastic transformations to imitate shape variations, and cut-out operations for partial occlusion of the image. These augmentations help prevent overfitting since the model is exposed to more situations and improve the model's performance when encountering variations (15).

1.8.3 Comparative Analysis of Deep Learning Classification Algorithms

1. Inception V3: A Convolutional Neural Network Architecture

Inception V3 is an improved model of the simple Inception known as GoogLeNet that aims at improving the capabilities of Convolutional Neural Networks and the amount of memory they occupy while at the same time increasing the level of accuracy (16). This study uses Inception V3 for two primary purposes: Histopathological image classification and histopathology image segmentation. In classification, Inception V3 can distinguish between different tissue types, as well as the diseases that are associated with these tissues, which is the core of arriving at the diagnosis.

Segmentation is done using U-net architecture and Inception V3 is used as the encoder in this case to enable the ability to segment specific structures within the histopathology images. This has shown that Inception V3 is an effective model that can be employed in different strategies in medical image analysis.

2. VGG16

VGG16 was introduced by Simonyan and Zisserman in 2014 and is considered one of the most important works in deep learning because of its deep, consistent architecture that focuses on simplicity. It has a total of 13 convolutional layers with a small 3×3 filter size to maintain a uniform feature extraction mechanism across the depth. This method increases the model's potential to learn intricate features, which is why it is highly useful for image identification. Due to its effectiveness, especially when used for transfer learning, the VGG16 is considered a benchmark model for future works and implementations in many fields (17).

3. ResNet50

He et al. 2015 proposed ResNet50, through which they introduced a new method of training very deep neural networks using residual blocks and skip connections. This architecture helps to avoid the issue of vanishing gradients and therefore, the network can learn from many layers. The ResNet50 model has 50 layers consisting of convolutional and fully connected layers, and it employs deep residual learning to ease the training of deep networks while ensuring that the signal does not get lost in the layers. The architecture and learning capability of deeper networks make ResNet50 versatile, and besides image classification, it can be applied to image segmentation(18).

4. EfficientNetB0

Another model in the CNN category is EfficientNetB0 developed by Tan and Le in 2019 which is an improvement in the CNN due to an optimal trade-off of the network's dimensional sizes. The model shows a compound scaling strategy that is logically accustomed to scaling up and down the depth, width, and resolution of the network, under the guidance of a set of empirically determined scaling coefficients. When combined with using mobile inverted bottleneck convolutions and Squeeze and Excitation blocks, this sheds new light on the improved efficiency and accuracy of the models. As a sub-model of the EfficientNet series designed for combating difficult image recognition tasks, the EfficientNetB0 is an ideal start for creating much larger and more potent models (19).

1.8.4 Comparative Analysis of Machine learning Classification Algorithms

1. Gradient Boosting

Gradient Boosting is a powerful algorithm used in machine learning for both regression and classification. It improves weak learners, especially decision trees, through a process of multiple passes with the objective of minimizing a loss function, a concept that was proposed by J. H. Friedman in 2001. It deals with more challenging parts of the data and enhances the precision in an iterative manner by adding predictors one by one to correct previous errors. This method is slightly more advanced than other ensemble methods as it focuses on error reduction and gives improved results with large and complicated data sets. The versatility of Gradient Boosting using different data types and loss functions allows custom solutions to be developed for a variety of modeling scenarios, which makes it a very versatile tool in predictive analytics (20).

2. Random Forest

The strength of Random Forest lies in ensemble learning, and it is equally proficient at both classification and regression. It creates many decision trees on different data subsets and combines their outputs, to improve accuracy and prevent overfitting. Besides improving predictive accuracy, it reveals feature importance that improves understanding of the dataset. Random Forest has the ability to combine the strengths of many learners and simple deployment, therefore, it is a cornerstone in predictive modeling methodologies (21).

3. Support Vector Machine (SVM)

SVM is a foundation for classification, regression, and detecting outliers in supervised learning. It tactically increases class separation in the feature space, using the kernel trick to effectively deal with the data that is non-linearly separable by elevating it into higher dimensions. This focus on maximizing class margins provides accuracy and prevents overfitting, critical in high-dimensional settings. The flexibility provided by SVM through kernel function customization makes it an ideal candidate for precision-intensive applications even though it is computationally expensive in larger data sets (22).

4. Extremely Randomized Trees

The method of Extremely Randomized Trees (Extra Trees) stands out in the field of ensemble learning by adding more randomness into tree construction and choosing random splits to reduce variance and slightly increase bias. This randomness enhances the training process and makes the model less sensitive to data noise and thus, improves the trade-off between the bias and variance for better accuracy. Extra Trees enhances feature selection in a very subtle manner, which is helpful in feature selection as it underlines the features naturally. It outperforms in complex and noisy data, and having less computational complexity makes it suitable for several classification and regression techniques (23).

5. K-Nearest Neighbors

The basis for the K-Nearest Neighbors (KNN) is the concept of proximity where the result is influenced by the 'k' nearest points. KNN is a type of 'lazy learning' where it does not compute anything until the moment when it must make a prediction, and it keeps the dataset in memory. The choice of the appropriate 'k' to some extent defines the noise level and model's propensity to under or overfit. The KNN algorithm's versatility from image classification to recommendation systems recommends it as an easy-to-use and efficient solution for many predictive problems; however, it has the drawback of being computationally intensive for large amounts of data (24).

1.8.5 Deep Learning Approaches to Image Segmentation

1. U-Net

The U-Net model proposed by Ronneberger et al. in 2015 is one of the most creative models in biomedical imaging due to its U-structure and is specifically suitable for image segmentation. This design is backed up by skip connections that combine high-level features learned in the early layers with low-level features learned in the later layers. These relations are necessary for the preservation and establishment of spatial information that is important for detailed segmentation. The model uses chains of 3x3 convolutions without padding, followed by ReLU activation and 2x2 max pooling for effective downsampling, doubling the feature channels at each step to improve its feature-capturing capabilities (25).

2. Attention gates

The use of attention in Convolutional Neural Networks (CNNs) enables the network to pay more attention to the important aspects of the input data which improves the performance of the network in tasks such as image classification and segmentation. There are several types of attention, such as spatial attention, which focuses on the important areas of an image, and channel attention, which emphasizes important feature channels. Spatial attention generates an attention map through the convolutional layers and activation functions, while channel attention uses global pooling and fully connected layers to come up with channel-wise attention coefficients. Self-attention or intra-attention is another type of attention mechanism that is widely used in transformer architecture and it computes the attention score between every position in the input sequence as a way of representing the sequence (26).

Attention gates, which are a form of attention, enhance the performance of CNN by guiding the analysis towards the right features. For example, in U-Net, attention gates regulate the feature maps of the encoder with the gating signal of the decoder. This step entails determining the attention coefficients from the input feature maps and gating signals that regulate the input feature maps through element-wise multiplication. Finally, the feature maps are fed into the final layers of the network to increase the focus on relevant information to enhance the model's performance and reduce its complexity. It also enhances the ability of the network to highlight the significant regions and provides valuable information concerning the regions of the input data that the network has given special attention to (26).

3. ENet

ENet presented by Paszke et al. in 2016 is a network for applications that need fast and accurate segmentation, especially for mobile or embedded platforms. At the first few layers, by performing swift downsampling at the initial stages, ENet reduces the size of the input data and narrows down the field of view to the region of interest. This efficiency is accomplished by using bottleneck modules that break down the convolutions to depth-wise separable convolutions, hence lowering the computational complexity while at the same time reducing the model size hence it is suitable for real-time use (27).

4. PSPNet

Proposed in 2017 by Zhao and others, the Pyramid Scene Parsing Network, or PSPNet, is the state-of-the-art method for the scene parsing problem with a pyramid pooling module. This module gathers characteristics at different levels; therefore, it is possible to gather context data from the local and larger levels which is essential for the analysis of complex scenes. This procedure of pooling at several scales and then up-sampling the features back to the original dimension helps PSPNet to capture contextual information, thus enhancing the ability of the network to analyze a scene (28).

5. Nested U-Net (U-Net ++)

The Nested UNet or UNet++ is an enhanced version of the original U-Net architecture which was proposed by Zhou et al. in 2018 and it incorporates nested and dense skip pathways. This complicated architecture greatly decreases the gap between the encoding and decoding sections of the network, helping to learn features at multiple scales. U-Net ++ has dense connectivity where every layer is connected to every other layer in the forward sequence, enabling the effectiveness of the reuse of features and feature propagation in the entire network to give better segmentation results (29).

1.8.6 GrabCut Algorithm

The GrabCut algorithm is an algorithm used in computer vision that seeks to segment an image into foreground and background. It employs Graph Cuts optimization and integrates color features and texture information for better segmentation (30).

1.8.7 Evaluation methods

Performance evaluation of the model uses confusion matrices (31) . This technique enables us to measure the model's accuracy by comparing the real class labels with the predictions by the model. It provides an overall view of the model's performance in different classes and helps to recognize the parts where the model performs well and where it struggles.

1- Accuracy

In classification tasks, this metric evaluates the overall correctness of the model and is calculated as the ratio of correctly predicted observations to the total observations. It is defined as the sum of True Positives and True Negatives divided by the total number of cases.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1.1}$$

Where: TP = True Positives, TN = True Negatives, FP = False Positives and FN = False Negatives

In segmentation tasks, accuracy measures the proportion of correctly classified pixels to the total number of pixels in an image. It evaluates how well the model segments the entire image.

$$Accuracy = \frac{Number\ of\ Correctly\ Classified\ Pixels}{Total\ Number\ of\ Pixels} \tag{1.2}$$

Where:

Number of Correctly Classified Pixels: The sum of True Positives and True Negatives in the segmented image.

2- Precision

In classification tasks, this metric evaluates the number of true positives among all positive calls. It is defined as the number of True Positives divided by the sum of True Positives and False Positives.

In segmentation tasks, precision measures the model's accuracy in identifying the target class pixels among all pixels it has classified as that class. It indicates the proportion of correctly segmented pixels compared to all pixels labeled as belonging to the target class.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1.3}$$

Where:

In classification, **TP** = True Positives and **FP** = False Positives

In segmentation, **TP**: Pixels correctly identified as belonging to the target class, **FP**: Pixels incorrectly identified as belonging to the target class when they are not.

3- Recall (Sensitivity)

In classification, this metric evaluates the number of true positives among all actual positives. It is expressed as True Positives over the sum of True Positives and False Negatives.

In segmentation tasks, recall measures the model's ability to correctly identify all pixels belonging to a particular class in an image. It is calculated as the ratio of correctly segmented pixels to the total number of actual pixels belonging to that class.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1.4}$$

Where:

In classification, **TP** = True Positives and **FN** = False Negatives

In segmentation, TP: Pixels correctly identified as belonging to the target class, FN: Pixels that are part of the target class but not identified by the segmentation model.

4- F1 Score

The F1 Score is the harmonic mean of Precision and Recall, which gives a balanced measure of both metrics. It is defined as twice the product of Precision and Recall divided by the sum of Precision and Recall.

$$\text{F1 Score} = 2\frac{\text{Precision X Recall}}{\text{Precision} + \text{Recall}} \tag{1.5}$$

5- Weighted average

Weighted average metrics are calculated by first computing the metric (such as Precision, Recall, or F1-Score) for each class independently, Then, these metrics are averaged across all classes, with each class's contribution weighted according to the number of samples it contains (32).

$$Weighted_{Recall} = \frac{1}{N}\sum_{i=1}^{N} Recall_i\, W_i \tag{1.6}$$

$$Weighted_{Percision} = \frac{1}{N}\sum_{i=1}^{N} Percision_i\ \ W_i \tag{1.7}$$

$$Weighted_{F1score} = \frac{1}{N}\sum_{i=1}^{N} F1score_i W_i \tag{1.8}$$

Where N is the number of classes, W represents the weights.

The Jaccard Index (33) and Dice Coefficient (34) are two of the most used measures for comparing the similarity of two sets, especially in the case of image segmentation in computer vision and medical image analysis. Here is the analysis of the respective:

6- Jaccard Index

The Jaccard Index commonly known as the Intersection over Union (IoU) is a measure that is essential when checking the accuracy of the segmentation we have done. It provides a numerical value for the similarity between the segmentation masks generated by the model and the actual ground truth, providing a measure of the model's effectiveness in identifying the regions of interest.

$$\text{Jaccard Index (IoU)} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{1.9}$$

Where:

Area of Overlap = The region where the predicted segmentation and the ground truth segmentation both indicate the presence of an object.

Area of Union = The total area covered by both the predicted segmentation and the ground truth segmentation.

7- Dice Coefficient

The Dice Coefficient is like the Jaccard Index, but it calculates twice the numerators of the overlap between the predicted and the ground truth masks in relation to the total masks' sizes. It's especially valuable in the context of medical image analysis for its sensitivity to small structures.

$$\text{Dice Coefficient} = 2\frac{\text{Area of Overlap}}{\text{Total Area of Prediction} + \text{Total Area of Ground Truth}} \quad (1.10)$$

Where:

Area of Overlap = The region where the predicted segmentation and the ground truth segmentation both indicate the presence of an object.

Total Area of Prediction = The total area covered by the predicted segmentation.

Total Area of Ground Truth = The total area covered by the ground truth segmentation.

1.9 Literature Review

In the dynamic and fastest growing area of breast cancer research, several fronts have witnessed some positive developments. From the beginning, the emergence of machine learning and deep learning technologies has changed the way breast cancer is classified, in terms of distinguishing between benign and malignant tumors. This includes collecting different data sets and new models for better classification. Along with that, the segmentation of malignant regions in whole slide images has achieved awesome progress, with deep learning algorithms like Convolutional Neural Networks providing fortune accuracy for diagnosis and treatment planning, which is critically needed.

Moreover, in 2023, the effect of treatment response, especially in the prediction of responses to neoadjuvant chemotherapy, has been delivered by the creation of a Gradient Boosting Machine model. This approach utilizes advanced atomic features to make more precise predictions of the treatment outcomes, therefore, forming a new and highly promising way of cancer therapies. These developments together are a big jump on the breast cancer battlefield, where patients will get better prognoses due to using precision medicine as well as advanced computing.

1.9.1 Breast Cancer Classification

Breast cancer diagnosis is a modernistic field that forms solid interest for both medical organizations and researchers. It started to receive more and more attention from researchers due to the major evolution which happened in the computer infrastructure over the past few years. Numerous studies have been conducted to develop new machine learning techniques and deep learning networks to build high accuracy models that can classify benign and malignant masses.

In 2018, Hussain, L., Aziz, W., Saeed, S., Rathore, S., and Rafique, proposed a method to classify breast cancer into malignant or benign by presenting different techniques and extracting various features such as texture, morphological, entropy based, SIFT& elliptic Fourier descriptors (EFDs). The used dataset is Screening Mammography by the University of South Florida. Then they applied different classifiers such as Support vector machines (SVM), decision tree (J48), and NB. The result shows naïve Bayes given the highest accuracy (35).

In 2018, a team of researchers used Machine learning techniques such as Naïve Bayes (NB), k-nearest neighbor (KNN), and random forest (RF) to classify breast cancer into two classes malignant and benign. The used dataset is the Wisconsin dataset of breast cancer and data Features extracted from a digitized image of a fine needle aspirate (FNA) of a breast mass. The results have shown that KNN methods achieved more accuracy = 98.27 and higher precision = 98.27% (36). In The same year, Amrane, M., Oukid, S., Gagaoua, I. and Ensari, T. presented different Machine learning techniques such as Naïve Bayes (NB) and k-nearest neighbor (KNN) to classify breast cancer into two classes malignant or benign. The used dataset is the Wisconsin dataset of breast cancer. The results have shown that KNN methods achieved more accuracy =97.5109 (37).

In 2020, Mohammed, S.A., Darrab, S., Noaman, S.A., and Saake. They proposed an approach to improve the accuracy in three machine learning models which are Naïve Bayes (NB), Decision Tree (J48), and Sequential Minimal Optimization (SMO). They achieved that by applying the resample filter several times. The used dataset is the Wisconsin Breast Cancer (WBC) and Breast Cancer dataset (BCD). The result shows high accuracy on the Breast Cancer dataset (BCD) after applying resampling filtering seven times (38).

In 2018, Guo, Y., Dong, H., Song, F., Zhu, C., and Liu, J, proposed different deep learning models such as Google Net, patch voting, and Hybrid CNN unit to detect breast cancer and classify it into four classes Benign, *in situ*, Invasive & Normal. The dataset was Provided by the Grand Challenge on Breast Cancer Histology Images 2018. The results have shown that Hybrid CNN unit is more accurate than others (39). In the same year, a team of researchers used transfer learning to classify breast cancer histology images into benign, *in situ*, invasive, and normal. The used models are Araujo et al (40), Nawaz et al (41), Alex Net, Google Net, and ResNet50, and they were applied that on Dataset of histology images published by Araujo and team. The results showed that ResNet50 gives more accuracy than others proposed models (42).

In 2020, a team of researchers proposed a new technique using Deep Neural Network with Support Value (DNNS) to classify breast cancer into two classes, malignant or benign. The presented method is based on support value in DNN to achieve better performance. The used dataset has been provided by M. G Cancer Hospital & Research Institute and contains 8009 histology images. The results have shown high improvement compared to other methods from previous studies where the proposed by DNNS method achieved Accuracy = 97.2%, Precision = 97.9% and Recall = 97.01% (43).

In 2018, Awan et al proposed a method that leverages convolutional neural networks (CNNs) combined with contextual information for breast cancer classification using histology images into four categories: *in situ*, benign, invasive, and normal, particularly designed for small datasets. The method consists of two main steps: using activation features of CNN trained for patch-based classification and then training a separate classifier using features of overlapping patches for image-based classification. This framework outperformed the state of the art methods, achieving an accuracy of 83% on the BACH challenge dataset (44).

In 2019, using the same dataset, Yao et al, proposed a deep learning model combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) with an attention mechanism to classify histopathology images into four categories: normal, benign, *in situ* carcinomas, and invasive carcinomas. The model adopted is a perceptron attention mechanism for integrating the various features that it extracts, switchable normalization rather than the common batch normalization, and targeted dropout as a

regularization approach. The model was tested and trained on the BACH dataset, achieving an accuracy of 0.92 (45).

In 2022, Prabira Kumar Sethy & Santi Kumari Behera proposed a classification model based on k-nearest neighbor (KNN) combined with deep and handcrafted features to diagnose breast cancer using histological images. The study considered four malignancy levels: normal, benign, *in situ*, and invasive. The classification was examined using three classifiers with three sets of deep features (extracted from the fc6 layer of pretrained networks: alexnet, vgg16, and vgg19) and three handcrafted features (GLCM, HOG, and LBP). The model based on fine KNN with combined features from vgg16 and LBP achieved an accuracy of 84.2% and an area under the curve (AUC) of 0.85. The likelihood ratio for positive results (LR+) was 12.5, indicating the method's significant contribution to diagnosis and its effectiveness as a diagnostic test (46).

In 2023, Simonyan, E.O., Badejo, J.A., and Weijin, J.S. proposed a CNN-based approach for breast cancer classification using a hybrid dataset from BreakHis and Histo. By leveraging transfer learning with four pre-trained models (DenseNet201, ResNet50, ResNet101, MobileNet-v2), DenseNet201 achieved the highest accuracy of 91.37% and 100% sensitivity at 200x magnification, outperforming state-of-the-art models (47).

In 2024, Liu, Y., Liu, X., and Qi, Y. introduced an adaptive threshold selection method for denoising histopathological images, aiming to enhance the accuracy of breast cancer classification. The approach integrates wavelet denoising and deep learning by optimizing the threshold through back-propagation during CNN training. This method was trained and tested on the BACH dataset from the Grand Challenge on Breast Cancer Histology Images and achieved a classification accuracy of 91.25%. The results demonstrate that this adaptive method surpasses traditional thresholding techniques like Sqtwolog and Minimax in terms of performance (48).

This paper aims to develop a segmentation model that can operate on all pathological images. In addition, to create a new way to detect breast cancer which will be applied to other datasets to measure the reliability of the algorithm, and it will be applied to data from the same patients on different time periods to analyze the treatment.

The table below summarizes the results of the research:

Table 1.1

Summary of previous studies

Authors	Year	Datasets	Models	Results	Remarks
(35)	2018	USF Mammography	SVM, J48, NB	Specificity, sensitivity, AUC = 1 (Bayes)	Features: texture, morphological, SIFT, EFDs. Classes: malignant or benign.
(36)	2018	The Wisconsin	RF, KNN, NB	RF ACC = 94.74, P = 92.18, R = 93.65	Features from digitized image of FNA of breast mass. Classes: malignant or benign.
(37)	2018	WBC	NB, KNN	KNN ACC = 0.9751, NB = 0.9619	Features from digitized image of FNA of breast mass. Classes: malignant or benign.
(38)	2020	WBC & BCD	J48, NB, SMO	WBC: J48 ACC = 98.20, NB = 76.61, SMO = 95.32	Features from digitized image of FNA of breast mass. resample filter is applied multiple times. Classes: malignant or benign.
(39)	2018	BACH Challenge 2018	GoogLeNet, Patch voting, Hybrid CNN	GoogLeNet ACC = 0.8, Patch voting = 0.825, Hybrid CNN = 0.875	Four classes: Benign, *in situ*, Invasive, Normal. 20X magnification
(42)	2018	Araujo's Dataset	Araujo et al (40) Nawaz et al(41) AlexNet GoogleNet, ResNet50	Patch-wise: Araujo et al = 66.7, Nawaz et al = 75.73, AlexNet = 79.84, GoogleNet = 81.07, ResNet50 = 83.60	Image-wise: Araujo et al = 77.8, Nawaz et al = 81.25, AlexNet = 82.3, GoogleNet = 83.6, ResNet50 = 85.0. Classes: Benign, *in situ*, Invasive, Normal. 20X magnification
(43)	2020	M. G Cancer Hospital	DNNS	ACC = 97.2, P = 97.9, R = 97.01	Magnification: 40X, 100X, 200X, 400X. Classes: malignant or benign.
(44)	2018	BACH Challenge	ResNet with svm	ACC=0.83	Magnification: 20X
(45)	2019	BACH Challenge	CNN (DenseNet121 + DenseNet169) + TD + SN + RNN (LSTM) + Attention (Ensemble)	ACC =0.92	Magnification: 20X
(46)	2022	BACH Challenge	KNN with combining the feature of vgg16 and LBP	ACC=0.84, P=0.84, R=0.84	Magnification: 20X
(47)	2023	BreakHis and Histo datasets	DenseNet201	ACC=0.9137	Magnification: 20X
(48)	2024	BACH Challenge	wavelet denoising and deep learning	ACC=0.9125	Magnification: 20X

1.9.2 segment Malignancy Regions in Whole Slide Images of Breast Cancer

The spectrum of medical imaging and cancer diagnostics is full of challenges. Observing recent studies that performed whole image analysis of sliced images by adopting deep learning and machine learning methods is very impressive. In this process, this technology helps identify and evaluate the size of the cancer lesion with high accuracy. The research that focused on cancer detection by each model separately has accomplished a feat that no one expected, bringing the theory of precision medicine to the clinic to set more specific management plans. Measurements of whole slide images that are carried out with exact and reliable results of the tumor areas in oncology formulate a new and unbeaten line in this sphere. The figures show us a better way, specifically that demonstrates a precise picture of the cancer aftermath in the body.

The Cruz-Roa study conducted in 2017 used deep learning to build models that demonstrated impressive accuracy in detecting diverse cancers from whole slide images, with AUC values of over 0.98 for all the cancer types studied. Likewise, computer aided diagnosis might be performed on slides and achieve an excellent accuracy of 100% while still ignoring 65-75% of the slides. This AI model is a tool that will not only make diagnosing pathological conditions faster but also significantly improve patient care (49).

Yusuf Celik and his team with their innovative study of 2020 managed to make a major step forward in the detection of early-stage invasive ductal carcinoma (IDC) using artificial intelligence. The work in this study is based on two pretrained deep transfer learning models, specifically ResNet-50 and DenseNet-161, that are applied to an image patch dataset of 277,524 histopathological samples. The study demonstrated the power of deep learning in medical diagnostics by reporting high performance metrics: the DenseNet-161 model obtained an F-score of 92.38% and a balanced accuracy of 91.57%, while the ResNet-50 model achieved an F-score of 94.11% and a balanced accuracy of 90.96%. Consequently, this implies that the use of some of the state-of-the-art artificial intelligence methods may be crucial in enhancing the reliability as well as the speed of the pathological diagnosis of breast cancer which is a very significant development in the analysis of digital pathology images. Another demonstration of the model is the employment of the BreaKHis dataset for the detection of breast cancer. It showed that the model performed well in the classification of histopathological images into benign and

malignant categories at different magnifications. Thus, this study can be considered a significant contribution to the field of computational pathology as it presents an effective way of improving cancer diagnosis and treatment (50).

Mridha et al. Shaw explored a deep learning solution for the identification of invasive ductal carcinoma (IDC) in breast cancer whole slide images (WSI) in their work conducted in 2021. To this end, they proposed the use of Deep Learning, including the application of ANN and CNN, to improve the diagnostic effectiveness of IDC. The authors evaluated their method on 162 WSI from IDC patients, and the initial average sensitivity their method achieved was 81.56%, which was improved to 86.24% using image augmentation and regularization. Thus, this study focuses on the effectiveness of deep learning in detecting IDC in WSIs, which greatly affects computational pathology (51).

In 2023, the paper of Peyret, R. et al. «Multicenter automatic detection of invasive carcinoma on breast whole slide images» was published which was an advance in the field of digital pathology. This work presents a patch-based convolutional neural network (CNN) to locate invasive carcinoma in breast whole slide images from different medical centers. This algorithm was implemented based on the reference acquisition center data and fine-tuning was performed by transfer learning methodology to apply it to the new target acquisition center with fewer training samples. Having achieved excellent results in terms of accuracy, recall, and precision on the reference and target datasets, the study focuses on the algorithm’s efficiency in maintaining the level of performance across various centers. This achievement supports the fact that AI has the potential to enhance diagnostic work in routine clinical pathology and becomes one of the most important aspects of the search for new and better methods of cancer diagnosis (52).

The table below summarizes the results of the research:

Table 1.2

Summary of Deep Learning in Medical Imaging Studies

Authors	Year	Datasets	Models	Results	Remarks
(49)	2017	44,732 whole slide images	Weakly supervised deep learning	AUC > 0.98 for all analyzed cancer types; 65-75% of slides could be disregarded.	Advances in medical image processing.
(50)	2020	277,524 histopathological image patches and the BreakHis breast cancer dataset for validation.	ResNet-50 and DenseNet-161	DenseNet-161 achieved an F-score of 92.38% and balanced accuracy of 91.57%. ResNet-50 had an F-score of 94.11% and balanced accuracy of 90.96%.	The study highlights the use of AI in accurately detecting invasive ductal carcinoma, demonstrating the effectiveness of deep learning models in enhancing breast cancer diagnosis.
(51)	2021	162 WSI of IDC patients (113 training, 49 testing)	Deep Learning, ANNs, CNNs	Improved model sensitivity from 81.56% to 86.24% in IDC detection after applying image augmentation and regularization.	Enhances IDC diagnosis in breast cancer WSIs.
(52)	2023	Breast whole slide images	Patch-based CNN algorithm.	High performance in detecting invasive carcinoma across different centers.	Demonstrates model generalizability and efficiency in routine pathology practice.

1.9.3 Treatment Response Effect

A study carried out by a group of researchers in 2022 investigated the capability of a Gradient Boosting Machine (GBM) multi-feature machine learning model in predicting the response of breast cancer patients to neoadjuvant chemotherapy (NAC) using features derived from pretreatment biopsies. Among different feature subsets, wavelet and graph-based pathomic features showed a clear advantage in the prediction of NAC response. The best response signature included seven characteristics, four from the wavelet category and three from the graph-based category. The model created by this signature predicted the NAC response with a sensitivity of 85%, specificity of 82%, and an area under the curve (AUC) of 0.90 (53).

1.10 Extension of Previous Studies and contribution

1.10.1 Enhancement of Breast Cancer Classification through Advanced Techniques and Feature Reduction

Leveraging of Previous Models with a Focus on Feature Impact: This study was an enhancement of some earlier deep learning models such as the ones proposed by Guo et al. (2018) and the KNN approach (2022) which has been seen to be efficient in classifying breast cancer. The main objective has been to establish how the reduction of features impacts the performance of these deep learning architectures aimed at producing high levels of classification accuracy.

In this work, we used the InceptionV3 architecture for training and testing, which incorporated techniques like batch majority voting on X20 histology images; the model was trained and tested using the An-Najah dataset. This was a significant development in the enhancement of the model and its capability to deliver results within the framework of our main research area. Thus, the model was validated for the ability to perform well on unseen data using another entirely separate dataset known as the Bach Challenge data. This blind testing phase is important for assessing the model's ability to make predictions for various data sets, thus giving a general idea of the model's performance and its reliability in various diagnostic cases.

A two-stage method has been adopted in this study to classification of X40 histopathology breast cancer images. First, the histology images are Segmented and then categorized after the nuclei have been singled out. On the other hand, evaluation is performed to assess the impact of reduced features measurement on the classification's precision and efficacy. Such a handling of the classification problem is also reasonable because this kind of dual stage is expected not only to enhance classification accuracy by isolating key regions of interest, reducing noise, and focusing on relevant features but also to provide ideas on the choice of the feature space for breast cancer diagnosis in deep learning systems.

1.10.2 Quantitative Analysis of Invasive Tumor Regions in Whole Slide Images

Foundation on Deep Learning Techniques: The research has extended the application of deep learning, originally pioneered by Cruz-Roa et al. (2017) and Yusuf Celik et al. (2020), to not just detect but also quantitatively analyze invasive areas in whole slide images.

Advanced Image Analysis Integration: The integration of deep learning frameworks and methodologies from these studies has been directed towards a more thorough analysis. This approach transcends mere detection, offering a quantifiable measure of invasive regions, thereby enriching the information available for cancer staging and treatment planning.

1.10.3 Evaluation of Treatment Response via Tumor Bed Segmentation

Grading System for Therapy Response: A new classification system into different grades based on tumor bed segmentation has been introduced. This methodology provides an innovative means to assess and quantify the efficacy of various treatment modalities, offering a personalized perspective on cancer therapy effectiveness.

1.10.4 Synthesizing Past and Present

This thesis integrates key findings and methods of previous studies more practically. The orientation of detailed and segmented analyses, which includes the synthesis of many approaches from the literature, has become a breakthrough in breast cancer diagnostics and treatment. This integrative approach not only validates the efficiency of the former models but also adds new layers and opportunities, thus paving the way for more interpretation of cancer in histopathology.

Chapter Two
Methodology and Experimental Design

This chapter presents the methodology and experimental design used in the study, focusing on breast cancer classification and segmentation in histopathology images at different magnifications. It covers data preprocessing, classification, and segmentation techniques, followed by a comparison of algorithms and evaluation methods. The chapter also includes an analysis of treatment response in breast cancer histology, with a focus on data preparation, segmentation, and evaluation metrics used to assess the model's performance.

2.1 Breast Cancer classification in 20X magnification images

The main goal of this study is to categorize histological images into four groups: invasive, *in situ*, benign, and Normal. To reach this goal, data has been gathered from An-Najah University which has samples for each category at both 20x and 40x magnification levels. The obtained data will be employed to design a model that is able to accurately train the classifier.

The evaluation was done on the two different datasets. For the first time in this study, the dataset, which is known as the new An-Najah dataset, has been presented. The second dataset used is the BACH Challenge dataset which is used in the model to figure out the performance in the different scenarios using An-Najah and BACH Challenge datasets. To refer to the full classification process, refer to Figure 2.1.

2.1.1 Data Preprocessing and Normalization

The histopathology images were preprocessed and normalized for analysis using a series of steps designed to make them compatible with different deep learning models. In the beginning, Reinhard normalization was performed for every histopathology image. This approach was chosen due to its ability to normalize color distribution across the images to a shared reference, and at the same time, retain the morphological features within the images. This stage is important for keeping the consistency of visual data throughout the dataset which allows deep learning models to concentrate on the true variances of the data instead of the artifacts introduced by various imaging conditions.

The steps for Reinhard normalization involved are:

1. Convert Images to L*a*b* Color Space

Transform the source and target images into the (L*, a*, b*) color space from the red, green, and blue (RGB) color space. This transformation can be done using equations used to convert colors.

2. Calculate the Mean and Standard Deviation

Calculate the mean and standard deviation of the three channels L*a*b* of the source and target image. Let:

$\mu L_s, \mu a_s, \mu b_s \ and \ \sigma l_s, \sigma a_s, \sigma b_s$ be the means and standard deviations of the source image, and $\mu L_t, \mu a_t, \mu b_t \ and \ \sigma l_t, \sigma a_t, \sigma b_t$ be the means and standard deviations of the target image.

3. Normalize the Source Image

Adjust each pixel in the source image to match the statistical properties of the target image. For a pixel with L*, a*, and b* values (L_s, a_s, b_s), the transformation is:

$$L'_s = \frac{\sigma L_t}{\sigma L_s}(l_s - \mu l_s) + \mu l_t \tag{2.1}$$

$$a'_s = \frac{\sigma a_t}{\sigma a_s}(a_s - \mu a_s) + \mu a_t \tag{2.2}$$

$$b'_s = \frac{\sigma b_t}{\sigma b_s}(b_s - \mu b_s) + \mu b_t \tag{2.3}$$

Here, L_s', a_s', b_s' are the transformed values for the source image pixel.

4. Convert Back to RGB

The final step is to convert back the (L*, a*, b*) image that has been transformed to the RGB color space using the inverse of the transformation matrix used in the first step. This procedure results in the source image having a color histogram that is close to the target image, thus the color appearance of the two images is similar. This technique is very useful in histopathology where color plays a significant role in the analysis and diagnosis of tissues.

We divided the data into training, validation, and testing with 70%, 15%, and 15%, respectively, to provide enough data for the learning of the model while at the same time having enough data to tune the hyperparameters and assess the model's performance accurately.

Afterward, each picture was divided into nine equal segments. This splitting aimed to retain the detail and quality of each sample as capturing the intricate details in histopathology images is essential for the learning of deep learning models. In addition, this process guarantees that all parts of the image are resized to 299x299 pixels. This dimension is applicable to nearly all deep learning architectures and plays an important role for the batch processing in the deep learning methods, which allows using many pretrained models, which work with input images, that have certain sizes.

These methods include the preprocessing and normalization steps, and one of the benefits of these steps is the production of a homogenized dataset, containing all the relevant information necessary for further application of deep learning techniques in the classification of histopathological images.

2.1.2 InceptionV3 model

InceptionV3 is a deep CNN that uses factorized convolutions and auxiliary classifiers to boost accuracy and efficiency, while also optimizing computational resources.

1. Key Features of Inception V3

Inception V3 is a deep learning algorithm which has the following features:

- Inception Modules: InceptionV3 extracts features at multiple scales due to the use of convolution layers with different filter sizes, pooling, and concatenation of layers. Such a modular design also enables the network to extract multiple and more complex features.
- Factorized Convolutions: To minimize computational cost Inception V3 employs several techniques that factorize the convolutions into other smaller operations. For example, a 3x3 filter can be approximated by two filters of size 1x3 and 3x1 respectively.
- Auxiliary Classifiers: To solve the vanishing gradient problem, auxiliary classifiers

are integrated while training the network. These intermediate layers compute the loss and assist in the computation of gradients to improve the training of the network.

- Grid Size Reduction: There are some strategies that have been used in the architecture to reduce the grid size which include the stride convolution and the pooling layers, it makes it possible to handle large images without much computational power.
- Label Smoothing: This is done with the intention of reducing the model's confidence with a single class during training, which also enhances the generalization ability of the model and reduces the chances of overfitting.

2. Inception Module

An Inception module performs multiple convolutions and pooling operations in parallel and concatenates their outputs. This design allows the network to capture various spatial features together. A basic Inception module can be represented as the following equation:

$$\text{Inception Module} = \text{Concat}(\text{Conv1x1}, \text{Conv3x3}, \text{Conv5x5}, \text{Pooling}) \quad (2.4)$$

3. Factorized Convolutions

Factorizing convolutions helps to reduce the computational load. For instance, a 3x3 convolution can be replaced by two separate convolutions:

$$\text{Conv3x3} \rightarrow \text{Conv1x3} + \text{Conv3x1} \quad (2.5)$$

4. Architecture Details

Inception V3 as shown in Figure A.3 in Appendix A consists of multiple blocks of Inception modules, each followed by reduction modules to decrease the spatial dimensions. The architecture ends with fully connected layers and a SoftMax layer for classification as follows:

- Input Layer: Starts with a 299x299 RGB image.
- Stem: A series of convolutional and max-pooling layers that preprocess the input.
- Inception Module A: Repeated three times, consists of various convolutional layers and pooling layers.
- Reduction Module A: Reduces the spatial dimensions using convolutional and

pooling layers.

- Inception Module B: Repeated four times, with more complex convolutional operations.
- Reduction Module B: Further reduces the spatial dimensions.
- Inception Module C: Repeated two times, again with diverse convolutional layers.
- Output Layer: Includes average pooling, dropout, a fully connected layer, and softmax for classification.

5. Training and Usage

Inception V3 is trained using a large dataset (like ImageNet) and can be fine-tuned for specific tasks using transfer learning.

6. Equations and Terms

1- Convolution

The convolution operation is defined as:

$$\mathrm{Conv}(x) = x\,W + b \tag{2.6}$$

where x is the input, W represents the weights, and b is the bias.

2- Pooling

Pooling operations, such as max pooling or average pooling, are used to reduce the spatial dimensions of the feature maps. They are defined as:

$$\mathrm{Pool}(x) = \max(x) \text{or } \mathrm{avg}(x) \tag{2.7}$$

where max or average pooling is applied to the input x.

3- Concatenation

The concatenation operation combines the outputs from different branches in an Inception module. It is defined as:

$$\mathrm{Concat}(x1, x2, \ldots, xn) = [x1, x2, \ldots, xn] \tag{2.8}$$

where $x1, x2, \ldots, xn$ are the outputs from different branches.

2.1.3 Model Configuration - InceptionV3

To overcome the problems of Breast cancer histology image classification, we have modified the InceptionV3 model through the incorporation of custom layers created for our classification purpose. This modification includes adding a Global Average Pooling 2D (GAP) layer, a Dropout layer with a 0.3 dropout rate to avoid overfitting, and a Dense layer with 1024 neurons. This Dense layer uses ReLU activation (54) and is regularized with L1-L2 to improve performance. The architecture ends with a SoftMax activated output layer, customized for the number of classes found in our dataset.

The final output of the proposed model mathematically represented from the InceptionV3 architecture, which is described layer by layer with their respective configurations, demonstrates the tactical improvements made to the InceptionV3 architecture. These modifications are pivotal for achieving high accuracy in the classification of histopathology images into the defined categories: benign, invasive, *in situ*, and normal.

$$\text{Output} = \text{softmax}\left(w_2 \times \left(\text{ReLU}\left(w_1 . GAP\left(F_{base}(X)\right) + b_1\right)\right) + b_2\right) \qquad (2.9)$$

Where:

X = Input to the base feature extractor.

$F_{base}(X)$ = Base feature extraction function.

GAP = Global Average Pooling

w_1 = Weight matrix for the first linear transformation.

b_1 = Bias term for the first linear transformation.

ReLU = Rectified Linear Unit activation function .

w_2 = Weight matrix for the second linear transformation.

b_2 = Bias term for the second linear transformation.

softmax = Softmax activation function.

This neural network architecture effectively extracts essential patterns from the data through feature extraction. Global Average Pooling (GAP) makes the model simpler and

decreases its complexity, so the model does not overfit (55). ReLu activation makes the model non-linear; it helps the model learn the relationship between data inputs and outputs. Finally, the softmax output layer gives clear and interpretable class probability which helps in understanding the model prediction.

To have a deeper look at the architecture of our model, including the concrete configurations and the reasons for each change, additional information is provided in Appendix D of this thesis. The appendix gives the web address for the complete model architecture enabling the reader to understand the technicalities of our modified InceptionV3 model.

2.1.4 Training Process

The training process consists of two phases:

1. Initial Phase: We freeze the base model layers to allow the top custom layers to adapt without modifying the pretrained feature extraction capabilities.
2. Fine-tuning Phase: We unfreeze the top 150 layers for detailed adjustments, optimizing the model.

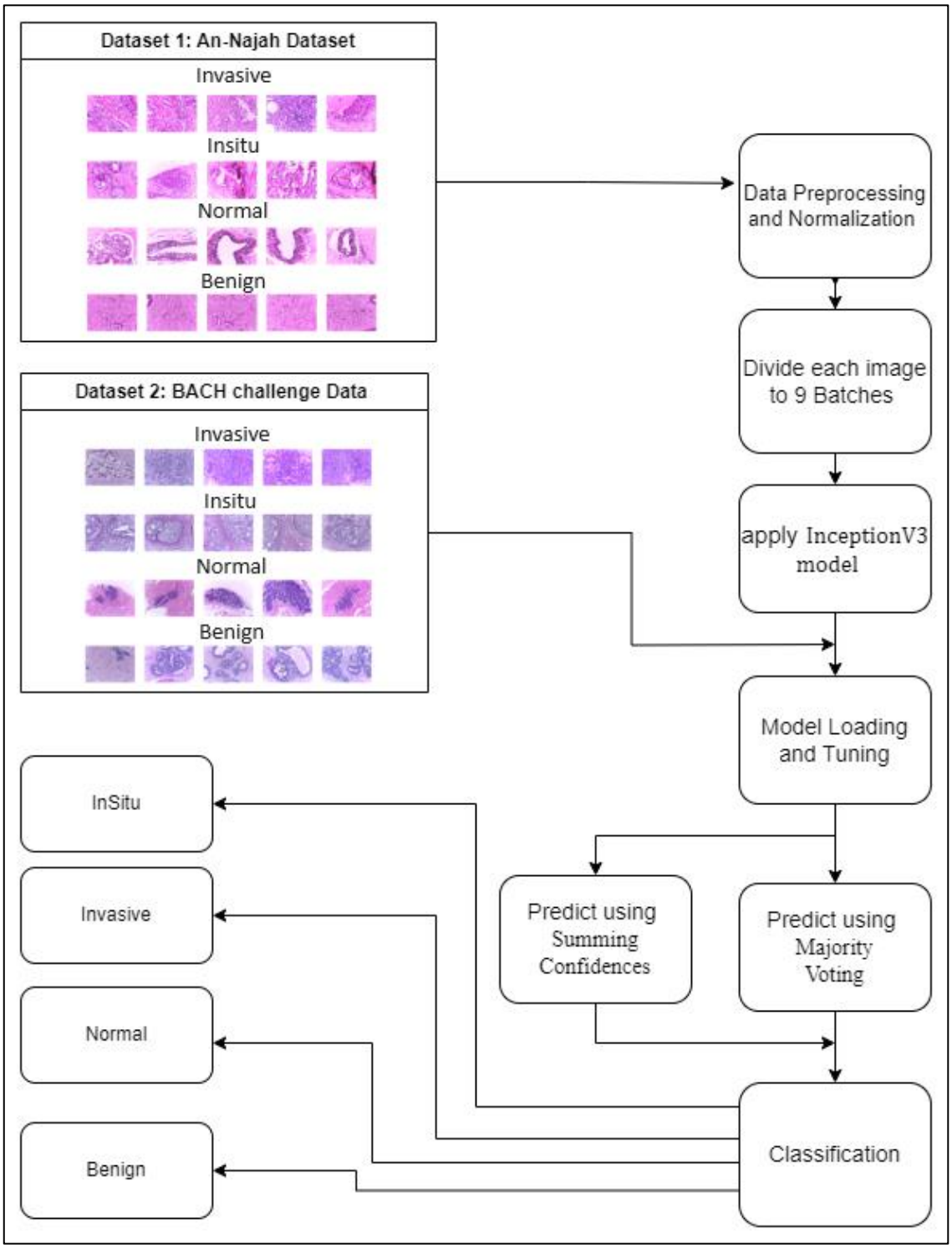
The model employs an Adam optimizer with a reduced initial learning rate and is compiled using categorical cross entropy to address the multi-class classification nature of our task.

2.1.5 Adaptive Learning Rate and Model Checkpoints

A ReduceLROnPlateau scheduler (56) adjusts the learning rate based on validation loss performance, aiding in model convergence. Model checkpoints are utilized to save the state at each epoch, ensuring the best model configuration is retained for evaluation.

Figure 2.1

Flowchart of classification method in 20x Classification



2.1.6 Algorithm Implementation

For the classification algorithm, we implement two ensemble techniques:

1. Majority Voting: For each segment of an image, a prediction is made, and then the class which gets the maximum votes across all segments, is chosen as the final class of the image.
2. Summing Confidences: The confidence score is accumulated for each part of an image to compute the total for each class. The class having the highest sum of confidence is the ultimate prediction.

2.1.7 A Comparison of Deep Learning Algorithms for Classification

We used the following models to compare with the customized model: VGG16, ResNet50, and EfficientNetB0. These architectures were selected for their well-established performance and diverse design philosophies in the field of convolutional neural networks, offering a comparison to assess the customized model's capabilities and efficiency.

2.1.8 Evaluation Method

We employed a confusion matrix to measure our model's performance, evaluating it using the main metrics as shown in section 1.8.7:

1- Accuracy for a whole model.

2- Precision, Recall, and F1 Score for each class.

3- Precision, Recall, and F1 for a whole model using Weighted-Average.

2.2 Breast Cancer classification in 40X magnification images

The classification process involved utilizing the An-Najah dataset, where the approach primarily focused on segmenting nuclei in histopathology images. Subsequently, features were extracted based on the details of these nuclei, followed by classification using a specific model tailored for this purpose. To see the flow chart of this full classification process, refer to Figure 2.2.

2.2.1 Data Preprocessing and Normalization

The concentration was on the preprocessing and normalization of histopathological images, like the steps described in section 2.1.1. Each image was thoroughly processed, including the Reinhard normalization application. This method was selected particularly for its power to standardize the color distribution throughout the dataset without changing the fundamental morphological details of the histopathology images. Normalization is the key process in which deep learning is based on uniform and accurate visual information, making the analysis based on real differences in data and not artifacts created by different imaging conditions. The data is split into 80% for training and 20% for testing.

After the normalization step, the histopathology images were partitioned into 16 identical segments. This segmentation is a strategic approach applied to retain the details and quality of every section of an image. The preservation of the detail within each segment is critical for the segmentation model since it depends on these small details for the correct division of the image into the appropriate classes. Moreover, every segment had to be resized to 256x256 pixels, to satisfy the requirements of the segmentation model utilized in our research. This standard size allows for the processing of images in a batch, a common practice in deep learning that improves computational efficiency and model training.

Such application of the preprocessing and normalization, as described in section 2.1.1, guaranteed that the data was perfectly prepared for the use of deep learning in histopathological image segmentation, preserving the fidelity of image details and meeting the requirements of the segmentation model.

2.2.2 Image Segmentation

The first process of our proposed approach is the identification of nuclei from histological images. A U-Net model with the InceptionV3 architecture provides an efficient segmentation because of the model's superior performance in analyzing histopathology images. The segmentation process includes Data Preparation and Augmentation.

1- Image Preprocessing

The preprocessing stage is important for preparing the images before they are fed into the segmentation model. It involves several steps:

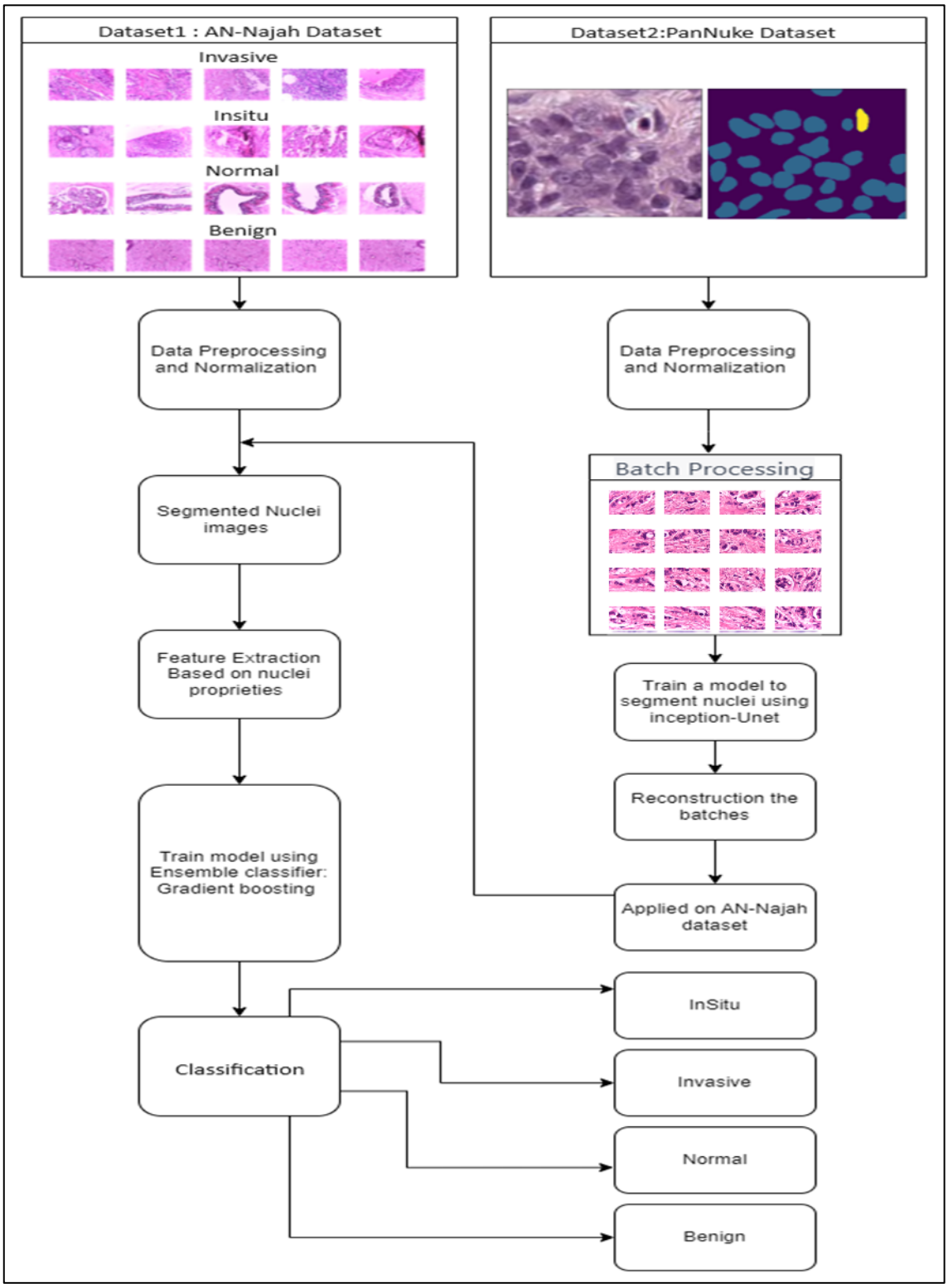
- **Median Filtering and Gaussian Blur:** First, median filtering is used to eliminate noise and at the same time preserve edges, and then a Gaussian blur is applied to smoothen the image. This enhances the quality of the image and makes the image ready for sharpening.
- **Reinhard Normalization:** This technique normalizes the color distribution of the source image to match a target image, ensuring consistent color properties across different images.

2- Image Patch Creation

To handle large images efficiently, they are divided into 16 patches. This step helps the model to work with the images in smaller portions while giving the right level of detail on the images for proper segmentation.

Figure 2.2:

Flowchart of classification method in 40x magnification



3- U-Net architecture

The following describes the architecture of U-Net, including its encoder (contracting path), decoder (expansive path), and bottleneck:

a. Encoder (Contracting Path):

- The left side of the U-Net as shown in Figure A.4 in Appendix A, which captures context and compresses the input image.
- Consists of a series of convolutional and max pooling layers.
- Each block typically has:
  - Two 3x3 convolution layers, each followed by a ReLU activation.
  - A 2x2 max-pooling layer with stride 2 for downsampling.
  - The number of feature channels is doubled at each downsampling step.

b. Bottleneck:

- The bottom layer of the U connects the encoder and decoder.
- Consists of two 3x3 convolutions, each followed by a ReLU activation.
- Often features the highest number of filters.

c. Decoder (Expansive Path):

- The right side of the U-Net as shown in Figure A.4 in Appendix A, which reconstructs the image to the original size.
- Consists of a series of up convolution (transposed convolution) and concatenation layers.
- Each block typically has:
  - An up-sampling layer (2x2 up convolution) to double the spatial dimensions.
  - Concatenation with the corresponding feature map from the encoder path (skip connections).
  - Two 3x3 convolution layers, each followed by a ReLU activation.
  - The number of feature channels is halved at each upsampling step.

d. Output Layer:

- A final 1x1 convolution layer to map each 64-component feature vector to the desired number of classes.
- Followed by a softmax or sigmoid activation function, depending on the segmentation task.

4- Model Training and Evaluation

The model uses the InceptionV3 architecture as its backbone, utilizing pre-trained weights while excluding the top layers to work as the encoder. In this case, the used InceptionV3 encoder is incorporated into the U-Net structure to improve the U-Net segmentation performance due to the encoder's ability to extract features. Once the model is constructed by the combination of the U-Net and InceptionV3 encoder.

The dataset undergoes image resizing to a consistent resolution of 256x256 pixels. Augmentation techniques, such as rescaling, are applied to ensure variations of the model.

Using an Adam optimizer (57), the efficiency of the model is therefore tested using metrics such as the Dice Loss function, the F1 score, and the IOU score. The objective is to generate accurate binary masks delineating the nuclei, governed by the equation:

$$L_{dice} = 1 - 2\frac{X \cap Y}{|X| + |Y|} \tag{2.10}$$

Here, X refers to the predicted mask, and Y refers to the ground truth, underscoring the significance of precision in segmentation.

5- Post preprocessing

In the post-processing and pre-feature extraction stage, the morphological opening filter was used to eliminate the noise from the images. This filtering technique that involves using the erosion of the image followed by a dilation assists in the removal of small noise and other unnecessary features that may be present on the image hence making the extracted features to be even more accurate to the actual structures present in the image.

2.2.3 Feature Extraction

After the segmentation process, an overall feature extraction is conducted on the segmented nuclei to quantify their morphological(58), textural(59), and color parameters, a critical step for elucidating the inherent complexities of histological patterns. However, the focus of our analysis narrows down to a selected set of features that are most pertinent to our study objectives. These features are chosen for their potential to offer significant insights into the nature and behavior of the nuclei across different types (Epithelial, Inflammatory, Neoplastic, Dead, and Connective Tissue). The selected features include:

- Area (A): This area of nuclear, has a basic but significant role of the total pixel number within the nucleus.

- Circularity (C): This check marks the degree of roundness or balance of the nucleus, that is how far from the nucleus is different from a circle.

- Perimeter (P): In addition, the total perimeter of the nucleus is a factor in the analysis of the level of complexity of the shape of the nucleus.

- Average Intensity (AI): The average pixel brightness of a nucleus, which is useful in assessing the global brightness or darkness of a nucleus.

- Texture: Nuclear images were measured using the Gray Level Co-occurrence Matrix (GLCM), which is a tool for measuring pixel intensity variation and provides information about the nucleus texture.

- Number of Nuclei: Segmentation and enumeration of each type of nucleus are inherent to determine the cellular composition and distribution across each histological structure.

The analytical process consolidates those attributes in a dataset format, where each line is the image ID and its related extracted attributes. This organization facilitates a nuanced analysis of each image, allowing for a detailed comparison and classification based on the types of nuclei present: This is a very complicated situation, but we will simplify it into major categories: Epithelial, Inflammatory, Neoplastic, Dead, and Connective Tissue. By this sort of systemic approach of feature extraction and classification, the interpretative attributes of the data are improved and this, in turn, will lead to a deeper

interpretation of the histological images about their cellular content and characteristics.

2.2.4 Comparison of Machine Learning Classification Algorithms

We conducted a comparative analysis of several classifiers for the classification of histology images: Random Forest, Gradient Boosting, SVM (Support Vector Machine), KNN (K-Nearest Neighbors), and Extra Trees. This comparative approach aimed to discern the most efficacious model for the precise categorization of histology images into their designated classes.

2.2.5 Classification and Model Optimization

A Detailed assessment of the performance of each classifier was conducted by tuning the important parameters with Grid Search CV to enhance the classification accuracy (60). The Grid search is a hyperparameter tuning method used to find the optimal set of parameters for a machine learning model. The optimization of parameters was tailored to each model's characteristics:

1. Random Forest and Extra Trees:

- N estimators: Tuned to pick the best number of trees in the forest that balances the complexity of learning and overfitting.
- Max depth: Tuned to regulate the depth of the tree, which denotes the model's ability to capture complex patterns without causing overfitting.
- Min samples leaf: minimum the number of samples in each leaf node, for generalization.

2. Gradient Boosting:

- N estimators: Defines the number of boosting stages to enhance learning efficiency.
- Learning rate: Adjusted to control the impact of each tree on the final outcome, with lower rates generally requiring more trees but potentially leading to more robust models.
- Max depth: Fine-tuned to manage the depth of individual trees, balancing complexity, and overfitting risks.

3. SVM:

- Focused on the C parameter (penalty parameter) to manage the trade-off between smooth decision boundary and classifying training points correctly.
- The kernel type was also optimized to effectively transform the data into higher dimensional spaces for more complex separations.

4. KNN:

- N neighbors: determining the number of neighbors to consider for determining the class of each sample.
- The distance metric for measuring closeness between samples was also an important optimization aspect to ensure the model's accuracy.

2.2.6 Evaluation Method

This section applies the evaluation method explained in section 2.1.6 by using confusion matrices to thoroughly evaluate the performance of our classification model. Using this method, we can quantitatively compare the real class labels with the model's predictions, giving an overall picture of its accuracy and the areas of improvement.

2.3 Segmentation of malignancy regions in a whole slide image

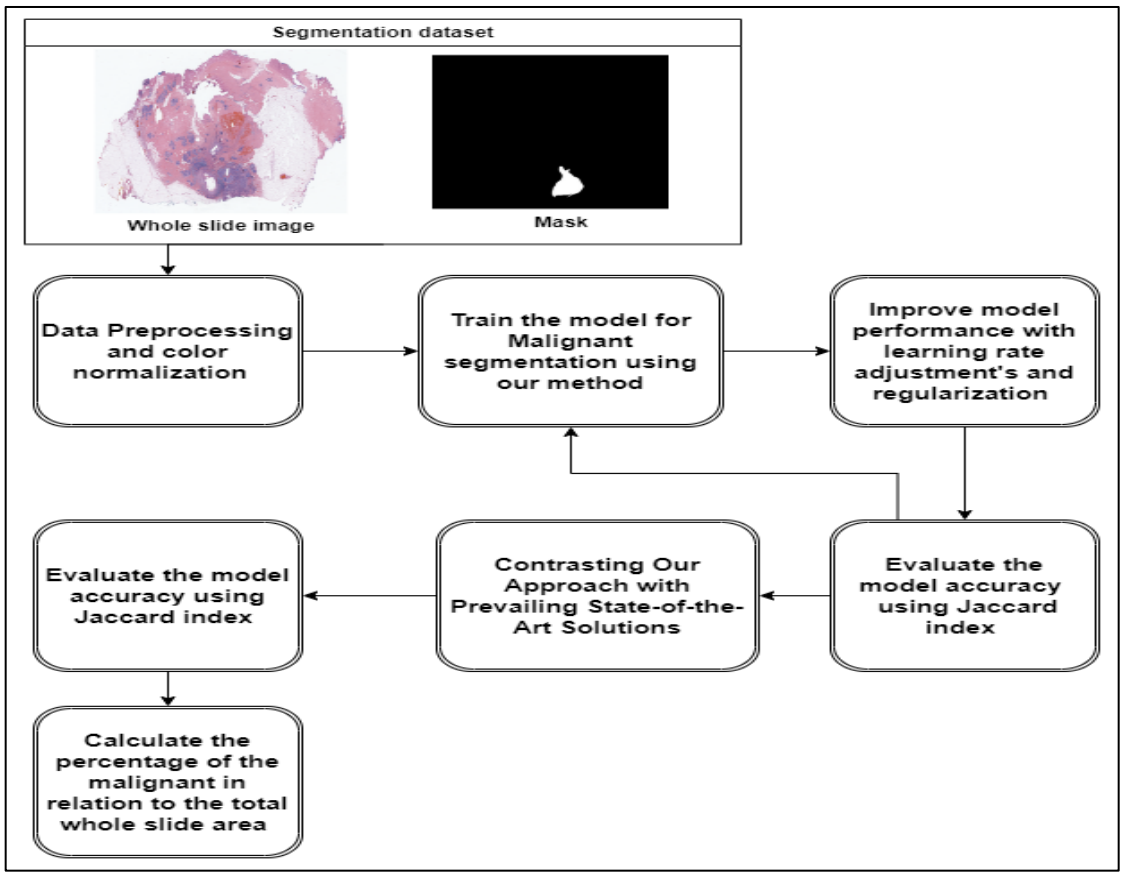
This section describes the methods employed in this work to focus on the problem of segmenting medical images for breast cancer histopathology. Using deep learning techniques alongside state-of-the-art architectural frameworks, this research aims to enhance the efficiency of identifying malignancy within whole slide images. The datasets used in this study, CWRU, TCGA, CINJ, and HUP contain many whole slide images and their corresponding masks for the malignant regions thus forming the basis of this study. Figure 2.3 illustrates the full segmentation process, highlighting the key steps involved. The focal points of our study include:

- **Detection of Malignancy Regions**: The exact identification of malignant regions in these slides is a key objective. This requires differentiating between benign and malignant tissue areas, allowing a targeted analysis of possible cancerous areas.
- **Quantification of Malignancy Area Relative to Slice Area**: We intend to do more than just detecting, we are targeting to measure the malignancy area accurately

concerning the whole slice area. This definition is important in helping the treatment planning process, providing oncologists with a useful parameter for evaluating the intensity and volume of cancerous tissue presence.

Figure 2.3

Flowchart of Segmentation of extent of invasive and area detection



2.3.1 Dataset Description and Preparation

In the methodology section dedicated to the preparation and enhancement of histopathology images for model training, a systematic process was employed to ensure that the images were optimally conditioned for analysis by various deep learning models. This process comprised several key steps, detailed as follows:

Image Resizing: First, all histopathology images were rescaled to a standard size of 512x512 pixels. This normalization is important for the maintaining of uniformity throughout the dataset which helps the models to learn from the images without the bias of the varying sizes.

Normalization: After resizing, the images underwent Reinhard normalization. as mentioned in section 2.1.1. The selected approach is known for providing color distribution uniformity among the images while keeping the important morphological and textural details. This kind of standardization eliminates the differences in the lighting and staining protocols that may result in a great difference in the image appearance.

We divided it into training, validation, and test sets in the ratio of 80:10:10 respectively because it is important for a small data set to have more data for training. This approach provides the model with the most data for learning while at the same time allowing for proper tuning and evaluation.

Data Augmentation: To improve the dataset and generalization of the models, the training images were augmented by rotation. This augmentation technique consists of the rotation of the images from different angles which helps to produce extra training samples. These rotations bring many different points of view, and therefore, help the models to learn more complete representations of histopathological features. This step is useful when they need to establish the reliability of the models against the orientation of the image.

2.3.2 Detailed Model Architecture in our approach

In our model, the architecture intricately combines the strengths of the InceptionV3 with a U-Net structure, enhanced by an advanced attention mechanism within the decoder, specifically adapted for the nuanced task of segmenting histopathology images of breast cancer. This section describes the elements of our architecture, highlighting how each contributes to the main goal of precisely identifying malignant regions.

Part one: InceptionV3 Backbone as the Encoder

Our encoder is based on the InceptionV3 architecture which is well known for its feature extraction capabilities. This backbone is pivotal for analyzing whole slide images through two primary mechanisms:

Multi-Scale Feature Processing: Using its characteristic Inception modules, our model processes inputs at multiple scales simultaneously. These modules use convolutional filters of different sizes (1x1, 3x3, and 5x5) within a single layer, allowing the extraction of a wide variety of features, from small details to larger spatial patterns. Such a multi-

modal approach is necessary to detect subtle signs of malignancy that are scattered over various tissue structures characterizing breast cancer histopathology images.

Enhanced Feature Map via Depth Concatenation: The results of the various streams of the Inception modules are merged using depth concatenation. This method unites detected features over scales, which makes the feature map a multi-dimensional array of context. Such kind of a feature set is important for high precision in the segmentation of malignant regions by combining the low-level textures and high-level semantic cues, which are critical for accurate segmentation.

Part two: Decoder with Integrated Attention Mechanism

Our decoder is augmented with a sophisticated attention mechanism, refining the segmentation outputs with an emphasis on precision and relevance:

Selective Feature Enhancement Through Attention Gates: By using attention gates, the model optimally reinforces the important features of the segmentation task. These gates evaluate and change the salience of features according to their spatial significance, favoring regions suggestive of malignancy. This selective improvement guarantees that the focus of the model remains only on the areas of interest and thus, increases the accuracy of segmentation in differentiating malignant from benign.

Incorporation of Spatial Context for Distinction: The attention mechanism enhances the discriminative ability of the model by embedding spatial context into the feature modulation procedure. It is a dynamic feature weighting approach considering the spatial distribution of histopathological patterns typical of malignant and benign tissues. This adaptive focus improves the model's capability to separate the foreground malignancy regions from the background, thus increasing both the sensitivity and specificity of the segmentation procedure.

The joint architecture of our model, combining the powerful feature extraction of the InceptionV3 backbone with the fine-grained focus of the attention mechanism using U-Net architecture, makes the model suitable for the complexities of medical image segmentation in breast cancer histopathology. This method increases the sensitivity to detect malignant regions and allows quantification of the malignancy extent concerning the slice area, which provides useful information for treatment planning.

To have a deeper dive into the specific configurations, layers, and rationale behind the integration of the attention mechanism with the InceptionV3 backbone, we have detailed this information in Appendix D. This appendix includes a URL that provides access to the full model architecture.

2.3.3 Sophisticated Training Strategies

Our model employs two sophisticated strategies to enhance training efficiency and prevent overfitting: dynamic learning rate adjustments and advanced regularization techniques.

First strategy: Dynamic Learning Rate Adjustments

The training process incorporates dynamic learning rate adjustments to improve model convergence:

Learning Rate Scheduler: A scheduler reduces the learning rate upon plateauing of validation loss, facilitating finer adjustments in weights and improving model generalizability (61).

The learning rate adjustment can be defined by the equation:

$$LR_{new} = LR_{initial} \times dropRate^{\left(floor\left(\frac{(1+epoch)}{epochsDrop}\right)\right)} \quad (2.11)$$

Where:

$LR_{initial}$ is the initial learning rate set at the beginning of training.

dropRate is a factor by which the learning rate is reduced.

epoch is the current training epoch.

epochsDrop is the number of epochs after which the learning rate is reduced.

Second strategy: Regularization Techniques

In addition to L2 regularization, dropout layers are strategically placed within the network to prevent overfitting by randomly omitting a subset of features during training.

The L2 regularization term (62) is defined by the equation:

$$\mathrm{L2}_{\mathrm{reg}} = \frac{\mathrm{lambda}}{2 * \mathrm{N}} \sum_{i}^{n} {W_i}^2 \tag{2.12}$$

Where:

lambda is the regularization strength, controlling the degree of regularization.

N is the total number of weights in the layer.

$\mathbf{w_i}$ represents each weight in the layer.

2.3.4 Comparing the customized model to another CNN architecture.

We used the following models to compare with the customized model: U-Net Architecture for precise biomedical segmentation, ENet tailored for speed in semantic segmentation, PSPNet mastering scene parsing with pyramid pooling, and Nested U-Net (U-Net ++) for enhanced feature learning. Each of these architectures was chosen for their unique strengths in specific areas of segmentation, providing a comprehensive benchmark to evaluate the performance and effectiveness of the customized model.

2.3.5 Tumor Area detection

In our study, after successful tumor region segmentation in histopathology slices, the last step was to determine the percentage of the tumor in the given slice. This was done using an iterative GrabCut algorithm that is important in separating the foreground (tumor areas) from the slide background. The GrabCut algorithm was iteratively applied to eliminate background parts and thus to concentrate only on the tumor bearing areas of the slice. In this way, it was possible to precisely determine the portion of the slice occupied by the tumor. The tumor area percentage was calculated by comparing the area of the segmented tumor mask to the total area of the slice. This framework, however, improves the accuracy of tumor quantification and enables a more elaborate and targeted analysis of histopathological slides.

Steps of the GrabCut Algorithm:

- Initialization: Begins with the labeling step, where a simple rectangle is drawn around the object of interest.
- Color Models: The foreground and background color models are constructed by using Gaussian Mixture Models (GMMs).
- Graph Construction: A graph is constructed with vertices representing pixels and edges representing relationships between pixels and source/terminal nodes.
- Edge Weights: Two types of edge weights are defined:
    - Data Term Weights: Connect pixels to the source or sink, based on their likelihood of being foreground or background.
    - Smoothness Term Weights: Connect neighboring pixels, penalizing labeling discontinuities.
- Min-Cut Optimization: The min-cut/max-flow algorithm provides the best cut of the foreground and the background, minimizing an energy function.
- Iterative Refinement: The segmentation result is refined iteratively by updating the color models and then performing graph cut computation.

2.3.6 Evaluation Method

We adopt the evaluation methods for the segmentation task as mentioned in section 1.8.7, using accuracy, recall, precision, F1 score, and Jaccard Index.

Jaccard Index or Intersection over Union (IoU) is a very important measure of the accuracy of segmentation tasks. It measures the degree of similarity between the segmentation masks generated by the model and the real segmentation masks, which directly reflects the accuracy of the model in identifying target regions. The Dice Coefficient is an alternative to the Jaccard Index and calculates twice the amount of overlap (intersection) between the predicted and the ground truth masks with respect to their total area. This metric is especially useful in medical image analysis because of its ability to detect small structures.

2.4 Treatment Response Analysis in Breast Cancer Histology

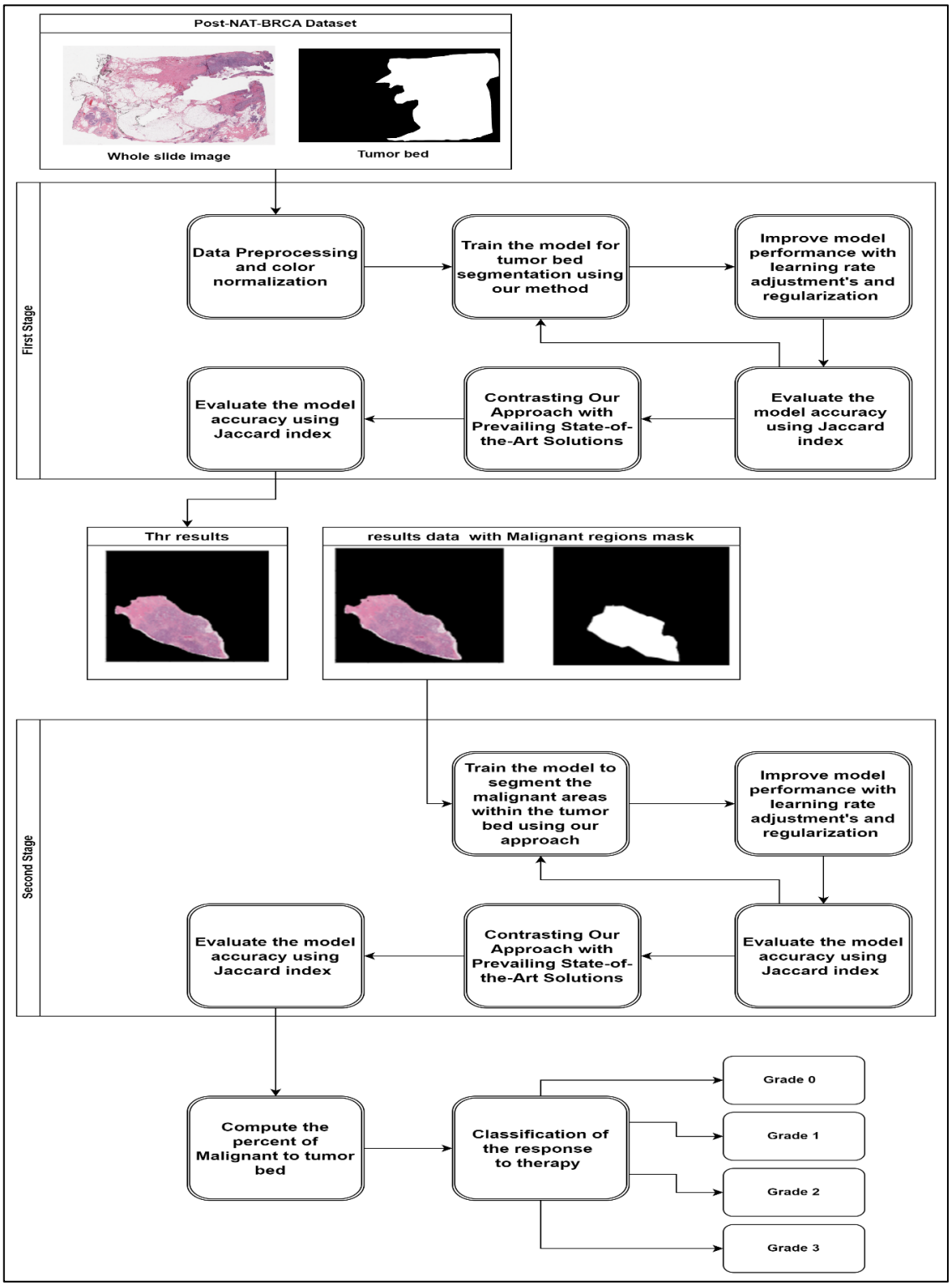
This Section presents a new approach for evaluating the treatment response in cancer treatment based on the analysis of histopathological images. The proposed framework uses three Steps for tumor bed segmentation, malignant tissue identification within the tumor bed regions, and the extent of malignancy to classify therapy responses. Refer to Figure 3.4 for the flowchart which shows the treatment response pipeline.

2.4.1 Dataset Acquisition

The research uses a Post-NAT-BRCA dataset comprising annotated whole slide images (WSIs) that show both tumor beds and malignant regions. These annotations are accompanied by data characterizing the ratio between malignant and tumor bed areas, serving for therapy response evaluation.

Figure 2.4

Flowchart of therapy response classification



2.4.2 Data Preprocessing

In our methodology for preparing histopathology images, a preprocessing routine consistent with the protocol outlined in the tumor area detection section was employed. This approach comprises three critical steps:

- Resizing: All images were resized to 512x512 pixels to maintain uniformity in the dataset for analysis.
- Normalization: Following the same normalization procedure in Section 2.1.1. Reinhard normalization was applied to standardize the color profiles of the images and to eliminate variations due to different staining procedures while maintaining the integrity of morphological features.
- Augmentation: The dataset was augmented, and the model was improved using image augmentation techniques like rotation and flipping.

The justification and specifics of these preprocessing steps are completely described in the part dedicated to tumor region detection. The use of this standard methodology guarantees a coherent and efficient histopathology image analysis framework.

2.4.3 Segmentation of Tumor Bed

As previously detailed in section 2.3, our methodology, employing a U-Net architecture with an InceptionV3 backbone, is specifically applied in this section for the precise segmentation of tumor beds within Whole Slide Images (WSIs). Enhanced by an attention mechanism, this method focuses on tumor related features, ensuring accurate segmentation. The optimization process, utilizing a combination of categorical cross-entropy and the Jaccard similarity coefficient as a loss function, meticulously improves both detection and delineation of tumor regions. This section applies the comprehensive methodological framework introduced in section 2.3 to the task of tumor bed segmentation, demonstrating its efficacy and precision in practical applications.

2.4.4 Segmentation of Malignant Regions

In the second phase, the model is isolated and retrained on the previously segmented tumor bed regions which enables it to only classify specific malignant tissue. At this point, the model fine-tuning is paramount as it moves to learn from the wide tumor bed context to the more detailed and discriminative recognition of malignant areas, which determine tumor aggressiveness and treatment resistance.

2.4.5 Therapy Response Classification (ThRC)

1. Quantitative Analysis

Post-segmentation, an automated quantitative analysis calculates the area of both the tumor bed and malignant regions. This analysis uses pixel-counting algorithms to provide accurate measurements necessary for the subsequent classification stage.

2. Ratio Calculation and ThRC

The therapeutic response is assessed by calculating the ratio of the malignant area to the total tumor bed area. Based on this ratio, the tumor's response to therapy is classified into one of the four predefined ThRC grades, with Grade I indicating a poor response to therapy (high malignancy presence) and Grade IV (63) indicating an excellent response (no malignancy detected) as following:

- Grade I: More than 50% of the tumor bed area contains viable tumor cells.
- Grade II: Between 10% and 50% of the tumor bed area consists of viable tumor cells.
- Grade III: Less than 10% of the tumor bed area has viable tumor cells.
- Grade IV: No viable tumor cells are observed in the tumor bed area.

2.4.6 Model Evaluation Metrics

The performance of the segmentation models is assessed using multi-metrics that include the Jaccard index, accuracy, Dice coefficient, sensitivity, and specificity as mentioned in section 1.8.7. These indices represent the precision of tumor bed and malignant tissue segmentation that are vital for precise ThRC.

Chapter Three
Results and Analysis

This chapter presents the results and analysis of the study, beginning with an overview of the datasets used, including the An-Najah, PanNuke, BACH Challenge, Tumor Segmentation, and Post-NAT-BRCA datasets. It then details the findings from the breast cancer classification experiments on both 20X and 40X magnification images, followed by the analysis of malignancy detection in whole slide images, highlighting data processing, deep neural network results, and evaluation. The chapter also covers tumor bed detection and therapy response classification, discussing segmentation and classification outcomes, along with the evaluation metrics used for performance comparison.

3.1 Data collection

In the area of medical imaging and pathology, a variety of datasets have been used to enhance the knowledge and skills that we have in diagnosing and treating tumors. Significantly, a new dataset gathered at An-Najah Hospital has been divided into four different classes, from which a classification system has been developed to differentiate malignant, benign, *in situ*, and normal tissue types. The original and unique origin and composition of this dataset make it possible to have a new approach in machine learning model training and diagnostic accuracy improvement.

Secondly, the segmentation of invasive malignancies has been significantly improved with the use of advanced image analysis techniques on externally supplied datasets. These applications, by clearly marking the boundaries of the invasive malignant tissues in histopathological slides, offer vital information required for proper treatment planning and cancer development.

In addition, external datasets that are obtained have played a key role in the evaluation of treatment response Classification. Based on the data, the results of the therapeutic intervention are divided into four grades, from 1 to 4, according to the effectiveness of the treatment (64). This stratification is the most beneficial for clinicians who can adjust the care of patients and develop therapeutic strategies by looking at the empirical data. Finally, this results in improving the patient's condition and improving the control over

cancer. These applications show that external data sets are essential in improving tumor pathology analysis and patient care management.

3.1.1 AN-Najah Dataset

An-Najah Dataset has provided primary data for the research carried out in this study. This dataset is collected by An-Najah National University, particularly by An-Najah Hospital Pathology Department and the dataset contains 1563 pathology images. These images are categorized into four classes: Benign, Invasive, *In situ*, and Normal. They were captured using 20x and 40x objective magnification; thus, this dataset is useful for tumor type classification and other pathological investigations.

To get detailed information about the dataset, refer to Table B.1 Appendix B which is titled “An-Najah Dataset Details”. Also, for researchers who want to use the dataset for more research, the URL of the dataset and the information on how to access it is provided in Appendix D of this thesis. This dataset is not only useful for the study of tumor pathology, but it is also an effective research material for the same.

The procedures of data collection described in Appendix C contain detailed information about how the data were obtained for this study.

3.1.2 PanNuke Dataset

The "PanNuke Dataset" is a dataset aimed at facilitating the task of nuclei instance segmentation and classification across 19 different tissue types. It is a semi-automatically generated dataset that provides exhaustive labels for 205,343 nuclei, each accompanied by an instance segmentation mask. The dataset spans 481 visual fields, 312 of which are randomly selected from over 20,000 whole slide images at various magnifications, sourced from multiple databases. From the PanNuke Dataset, only breast tissue data has been selected for this study. for detailed dataset characteristics, see Table B.2 Appendix B, "PanNuke Dataset Overview," summarizing key details (65).

3.1.3 BACH Challenge Dataset

The "BACH Challenge Dataset" is a dataset featuring four classes: Normal, Invasive, *in Situ*, and Benign. There are 400 of these pictures, all of them contain 100 images per class. The images that follow were taken at the 20x objective magnification, therefore,

they can be used for standard image analysis purposes. Specifically, they are suitable for medical research in the direction of classification of tumor type and other attributes of the data set, see Table B.3 in Appendix B for more detailed data set characteristics, “BACH Challenge Dataset” (66).

3.1.4 Tumor Segmentation Dataset

The “Tumor Segmentation Dataset” covers data from different sources such as CWRU, TCGA, CINJ, and HUP datasets. It is made up of whole slide images (WSIs) stained with Hematoxylin and Eosin, which are the most common staining in histopathology. This set of data is useful in the diagnosis and grading of tumors with a dataset of 584 images. The diversity of the dataset’s sources and application in tumor analysis and grading make it important in both medical research and clinical applications. For detailed dataset characteristics refer to Table B.4 in Appendix B “Tumor Segmentation Dataset” which includes the main specifications (67).

3.1.5 Post-NAT-BRCA Dataset

The Post-NAT-BRCA dataset comprises 96 high resolution whole slide images in. svs format, scanned using a 20x objective on an Aperio slide scanner at Sunnybrook Health Sciences Centre. It includes an excel file detailing patient clinical features such as age, treatment, and ER/PR/HER2 status, with a "Definitions" tab for column explanations. Each spreadsheet row corresponds to a slide and anonymized patient ID. Additionally, the dataset contains manual annotations of tumor cellularity and cell labels in Sedeen (.xml) files, organized into "WSI_train” and "WSI_test" folders for data analysis, For detailed dataset characteristics, see Table B.5 in Appendix B, "Post-NAT-BRCA Dataset," summarizing key details (68).

3.2 Breast Cancer Classification

For the first part of the research about classifying the breast cancer histopathology images into 4 classes (*In Situ*, Invasive, Benign, and Normal), different methods have been used on different magnifications. Each magnification has been experimented with different processes and steps.

3.2.1 20X Magnification Images

The 20x magnification histopathology images have been preprocessed, resized, augmented, and finally normalized as shown in Figure A.5 in Appendix A to a standard color before feeding them to the different deep learning models.

3.2.1.1 Data preprocessing and normalization results

The histopathology images have been prepared, normalized using Reinhard normalization, and split into 9 parts for each image to maintain the details and quality in each sample and to reach a common supported size (299,299) by different deep learning models.

After normalizing the images and splitting them, the images were transformed into 9 batches as shown in Figure A.6 in Appendix A.

3.2.1.2 Deep Neural Network Results

After preprocessing the images, and experimenting them using different deep learning models, and finally testing them using unseen data from the original dataset, in addition to another data from different and external source the Model gave an accuracy of 98.7% using InceptionV3 on the original dataset unseen data. On the other hand, the model gave an accuracy of 90% using the BACH challenge external dataset.

1. AN Najah dataset results

The results of the main dataset for each of the four breast histology classes using the sum confidences approach where the confidence of all batches for all classes are being summed and then the highest summation is chosen is 98.7% accuracy, as detailed in table 3.1 and Figure A.13 in Appendix A.

Table 3.1

X20 Breast histology classification using AN Najah dataset with summing confidences

Column Label	Precision	Recall	F1-score	Support
benign	1	0.95	0.97	20
in situ	1	1	1	13
invasive	1	1	1	28
normal	0.94	1	0.97	17

The confusion matrix for the summing confidences approach is shown in figure A.7 Appendix A.

The results show only one misclassified image which was classified as normal instead of benign. Meaning that it didn't mispredict any malignancy which could be more serious in the medical field.

On the other hand, the results of the main dataset for each of the 4 breast histology classes using the majority voting approach, where the class that appears in the most batches is selected (Sum confidences as tie-breaker) gave the same accuracy of 98.7%, as detailed in table 3.2 and Figure A.13 in Appendix A.

Table 3.2

X20 Breast histology classification using original dataset with majority voting

Column Label	Precision	Recall	F1-score	Support
benign	1	0.95	0.97	20
in situ	0.93	1	0.96	13
invasive	1	1	1	28
normal	1	1	1	17

The confusion matrix for the majority voting is shown in Figure A.8 Appendix A.

Even though the majority voting approach gave the same accuracy as the summing confidences approach, the main difference here is that the summing confidences approach misclassified a benign image into a *in situ* one, which gives the summing confidence a slight advantage over the majority voting approach according to these results.

2. Results Using the BACH Challenge Dataset for Testing

The results of the external dataset for each of the four breast histology classes, using the sum confidences approach, where the confidence of all batches for all classes are being summed, and then the highest summation is chosen, show 90% accuracy as detailed in table 3.3 and Figure A.13 in Appendix A.

Table 3.3

X20 Breast histology classification using BACH dataset for testing with summing confidences

Column Label	Precision	Recall	F1-score	Support
benign	0.85	0.90	0.87	100
in situ	0.93	0.86	0.90	100
invasive	0.91	0.93	0.92	100
normal	0.91	0.91	0.91	100

The confusion matrix for the summing confidences approach is shown in Figure A.9 in Appendix A.

The results show that most of the misclassified images (9) were normal images which were predicted as benign. Also, there are no normal images that were predicted as *in Situ* or invasive.

On the other hand, the results of the BACH challenge dataset for each of the four breast histology classes using the majority voting approach, where the class that appears in the most batches is selected (Sum confidences as tie-breaker), gave the same accuracy of 89.25% as detailed in table 3.4 and Figure A.13 in Appendix A.

Table 3.4

X20 Breast histology classification using BACH dataset for testing with majority voting

Column Label	Precision	Recall	F1-score	Support
benign	0.83	0.90	0.86	100
in Situ	0.93	0.84	0.88	100
invasive	0.91	0.93	0.92	100
normal	0.91	0.90	0.90	100

The confusion matrix for the majority voting approach is shown in Figure A.10 in Appendix A.

Like the insights from the summing confidence approach, most of the misclassified images (10) where normal images predicted as benign, also there were no normal images predicted as *in Situ* nor invasive.

3- Results on the BACH Challenge Dataset: Training and Testing

The results of the BACH challenge dataset for each of the four breast histology classes using both approaches, sum confidences and majority voting, show 96.6% accuracy, as detailed in table 3.5 and Figure A.13 in Appendix A.

Table 3.5

X20 Breast histology classification using BACH dataset with summing confidences

Column Label	Precision	Recall	F1-score	Support
benign	0.94	1	0.97	15
in Situ	0.94	1	0.97	15
invasive	1	0.87	0.93	15
normal	1	1	1	15

The confusion matrix for the summing confidences approach is shown in Figure A.11 in Appendix A.

The results show that there are two misclassified images, where invasive image was predicted as benign, and another invasive image predicted as *in situ*.

3.2.1.3 Evaluation and comparison

1- A Comparison with Different CNN Architecture

After preparing the images, they have been trained using multiple deep learning algorithms, including ResNet50, VGG16, EffiecentNetB0, and InceptionV3. Each deep learning algorithm took the batches themselves as an input rather than the full image.

The results of running different models using the summing confidences approach with the AN Najah dataset (using weighted average) are as shown in table 3.6 and Figure A.14 in Appendix A.

Table 3.6

X20 Breast histology classification models comparison using summing confidences

Column Label	Accuracy	Precision	Recall	F1-score	Support
ResNet50	0.76	0.77	0.75	0.75	78
VGG16	0.55	0.51	0.51	0.55	78
EffiecentNetB0	0.51	0.46	0.43	0.60	78
InceptionV3	0.987	0.99	0.99	0.99	78

And the results of running different models using the majority voting approach with the AN Najah dataset (using weighted average) are shown in table 3.7 and Figure A.15 in Appendix A.

Table 3.7

X20 Breast histology classification models comparison using majority voting

Column Label	Accuracy	Precision	Recall	F1-score	Support
ResNet50	0.76	0.78	0.76	0.75	78
VGG16	0.53	0.49	0.49	0.56	78
EffiecentNetB0	0.53	0.48	0.45	0.63	78
InceptionV3	0.987	0.99	0.99	0.99	78

The InceptionV3 model outperformed all other models with a consistently high accuracy of 98.7% in both approaches and 99% for recall, f1-score, and precision. On the other hand, RestNet50 showed an average accuracy of 76% while the other models performed poorly.

2- A Comparison with Existing Approaches using "Bach challenge dataset."

In Table 3.8, we compare the performance metrics of our approach with those reported in other studies, focusing on accuracy as the primary metric. The datasets used for these comparisons are from the BACH Challenge dataset.

Table 3.8

A Comparative Analysis with Existing Approaches

Column Label	Accuracy
Our approach	0.966
(39)	0.875
(44)	0.83
(45)	0.92
(46)	0.84
(48)	0.91

The comparative analysis presented in Table 3.8 highlights the effectiveness of our approach against several existing methods in terms of accuracy where the same dataset was trained and tested using different models and algorithms. Table 3.8 shows that the 96.6% testing accuracy of our approach significantly outperformed the four studies mentioned in this research (39), (44), (45), (46), (48), which achieve accuracies of 0.875, 0.83, 0.92, 0.84, and 0.91 respectively.

3.2.2 40X Magnification Images

The 40x magnification histopathology images have been preprocessed, resized, augmented, and normalized, then passed to a pretrained nuclei type segmentation model. Finally, multiple features have been extracted for each segmented nuclei type before training them using multiple ensemble techniques.

3.2.2.1 Data preprocessing and normalization results

The histopathology images have been prepared, normalized using Reinhard normalization, and split into 16 parts for each image to maintain the details and quality in each sample and to reach a (256,256) size which is needed for the segmentation model.

After normalizing and splitting the images, they and the annotations were transformed as shown in Figure A.18 in Appendix A.

The Figure A.18 in Appendix A shows batches of the Pannuke data which has the required annotations to detect different nucleus types, and the last row is a representation of the available annotations which shows the areas of different nuclei with different color for

each nuclei type (Epithelial, Inflammatory, Neoplastic, Dead, and Connective Tissue).

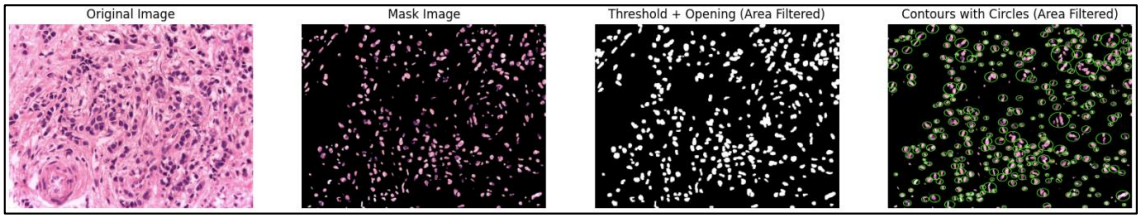
3.2.2.2 Nuclei Segmentation in 40X Magnification Images

After training the data for 100 epochs using the U-Net segmentation model which uses the InceptionV3 encoder, the model resulted in an f1-score of 0.6 and an iou-score of 0.5, Figure A.19 in Appendix A shows an example result.

Figure 3.1 shows the original mask and image on the left, and the predicted overall mask on the right. This and most of the images contained a lot of noise that was handled in post-processing before extracting the features. The noise was removed by applying a morphological opening filter.

Figure 3.1:

Nuclei segmentation post processing



The third and fourth images show the nuclei considered for features extraction which doesn't include a lot of noise resulting from the segmentation model.

3.2.2.3 Feature extraction

After preparing the segmented nuclei, a set of features has been extracted for all images, including the area, circularity, circumference, intensity, texture, and number of nuclei for each segmented type. The operation resulted in a dataset where each row has an image id and a set of features for each nuclei type and finally, a label which is invasive, *in Situ*, benign, and normal, as shown in Figure A.20 in Appendix A.

3.2.2.4 Features analysis

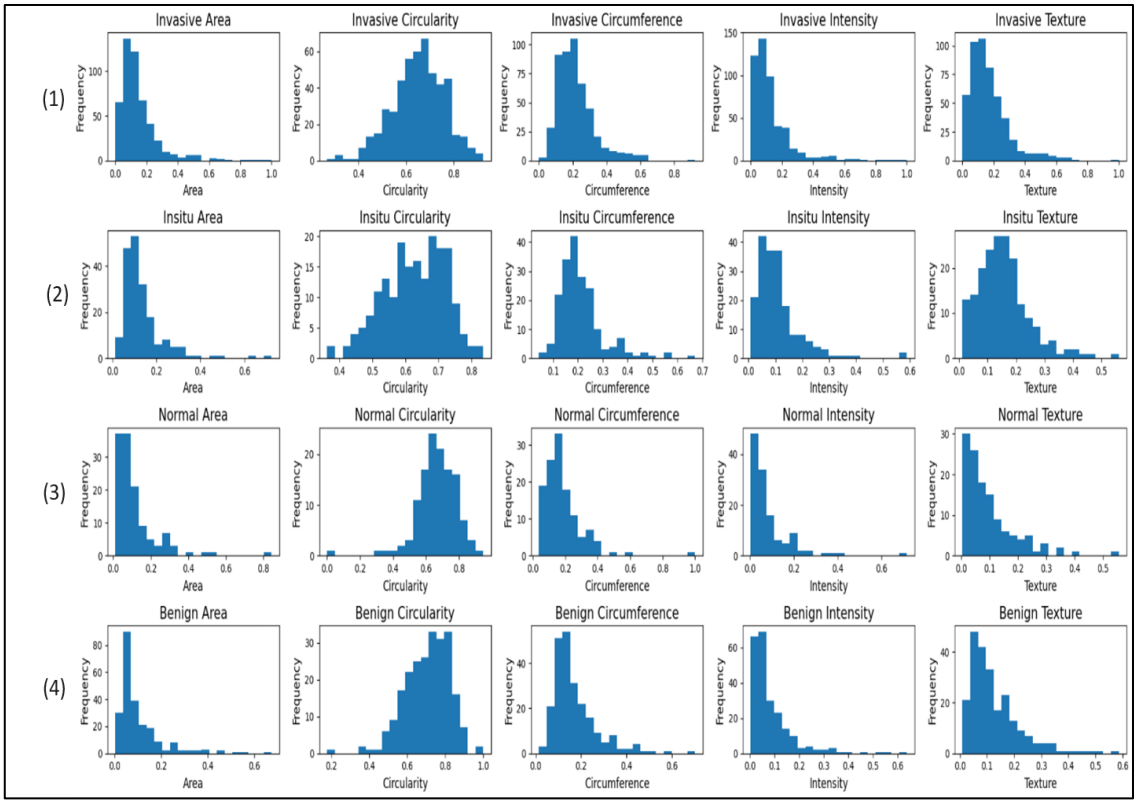
The resulting features have been analyzed to further understand the class differences on the extracted features.

The most class in the dataset is invasive with 495 samples as shown in Figure A.21 in Appendix A, then benign with 232 samples, then *in situ* with 190 samples, and finally

normal with 126 samples. The imbalance in data is because of the variance in the available samples from the collected source where the normal samples for example were very rare with most of the available samples were invasive.

Figure 3.2

Features distribution for data labels



In Figure 3.2, it appears that the Benign and Normal area histogram shows a significant skew towards the left with a high frequency on the smallest area range, while *in situ* shows a similar skew to the left with fewer peaks. On the other hand, the invasive histogram shows the most event distribution among the small areas, but it still indicates a skew towards the small values. Normal nuclei have the most circularity with a normal distribution shape on the right side of the histogram, like the benign nuclei which are focused on the large value size. However, Invasive and *in Situ* nuclei circularity has a wide range that covers most of the values from low to high circularity.

All classes show a skew towards the lower circumference values. Benign has a very high frequency for lower values indicating large number of small nuclei. Regarding the intensity, all nucleus types have a clear skew to the lower values with invasive having

more values in the middle. Finally, the texture is also mostly skewed to the left towards lower values with the lowest values in normal and benign nuclei, and a wide range of low values for both invasive and *in situ*.

Figure A.22 in Appendix A describes a clustering for 4 different clusters using K-means. The results show a very clear separation of clusters 0 and 1 from the rest, with a partial overlap between clusters 2 and 3, which indicates a strong possibility of separating the data into different classes. PCA1 (Principal Component 1) captures the highest variance in the dataset, while PCA2 (Principal Component 2) captures the second highest variance orthogonal to PCA1. This dimensionality reduction helps to visualize the data and the clustering results effectively.

3.2.2.5 Evaluation and comparison

The data frame has been trained on multiple models, the classes have been balanced, and the data has been tested using cross validation with 5 folds. For each model, the hyperparameters were tuned to find the best parameters that bring the highest accuracy, and weighted average metrics, as shown in table B.6 in Appendix B and Figure A.16 in Appendix A.

The results show high and consistent accuracy across all models with an advantage for the gradient boosting model and support vector classifier with an overall accuracy of 98% across all test folds. This could be explained by the nature of gradient boosting which is known for the ability of handling varied data types while reducing bias and variance, while SVC showed efficiency in a relatively high dimensional space.

On the other hand, the random forest has slightly lower metrics when compared to the ensemble Extra tree classifier. Extra trees might be performing slightly better due to the way of randomly choosing split points which could lead to a better generalization. Finally, the KNN model high performance indicates the classes were well-clustered in the dataset.

The main difference between the two models is the accuracy of classifying the benign class. Gradient boosting managed to identify only 92.7% of the benign classes correctly compared to 97.5% on SVC. On the other hand, the Gradient boosting outperformed SVC on *in situ* and normal classes slightly with a 100% accuracy of classifying the normal class compared to 98% in SVC and an accuracy of 99% of classifying the in-situ class

compared to 98% in SVC which both models have the same accuracy of 99% in classifying invasive.

In Figure A.23 in Appendix A, the ROC curve for both models shows an area of approximately 1 for all classes under the curve, with a slight advantage in the shape of Benign curve which is explained in the confusion matrix above.

While both models showed an excellent result, the SVC has more consistently high results across all classes with fewer false positives in benign.

3.3 Whole Slide Image Detect Malignancy and Analysis

For the second part of the research about detecting malignancy in histopathology images and computing the percentage of malignancy in the slice, multiple experiments have been done using different segmentation models and architectures.

3.3.1 Data processing and normalization results

The histopathology images have been prepared, resized to (512,512), and then normalized using Reinhard normalization, and finally, the train data was augmented by rotating the samples before feeding them into different models.

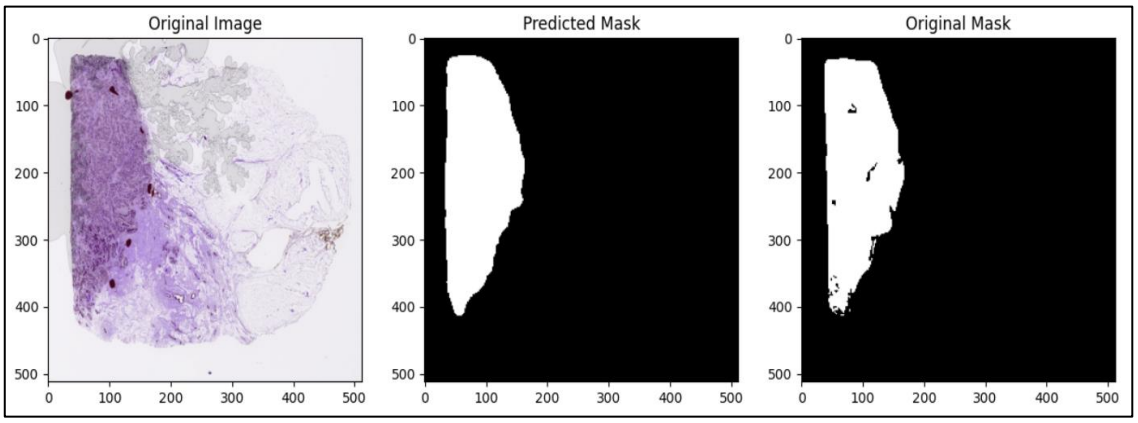
After normalizing and resizing the images, they were transformed as shown in Figure A.24 in Appendix A.

3.3.2 Deep neural network results

After preprocessing the images, and experimenting them using different deep learning models, and finally testing them using unseen data, the best Model (Customized U-Net - Inception with attention gates) segmented the tumor area with an 84% Jaccard index, 91% dice coefficient 96.5% accuracy, 93.2% sensitivity and 97.3% specificity. An example illustrating the segmentation result is shown in Figure 3.3.

Figure 3.3:

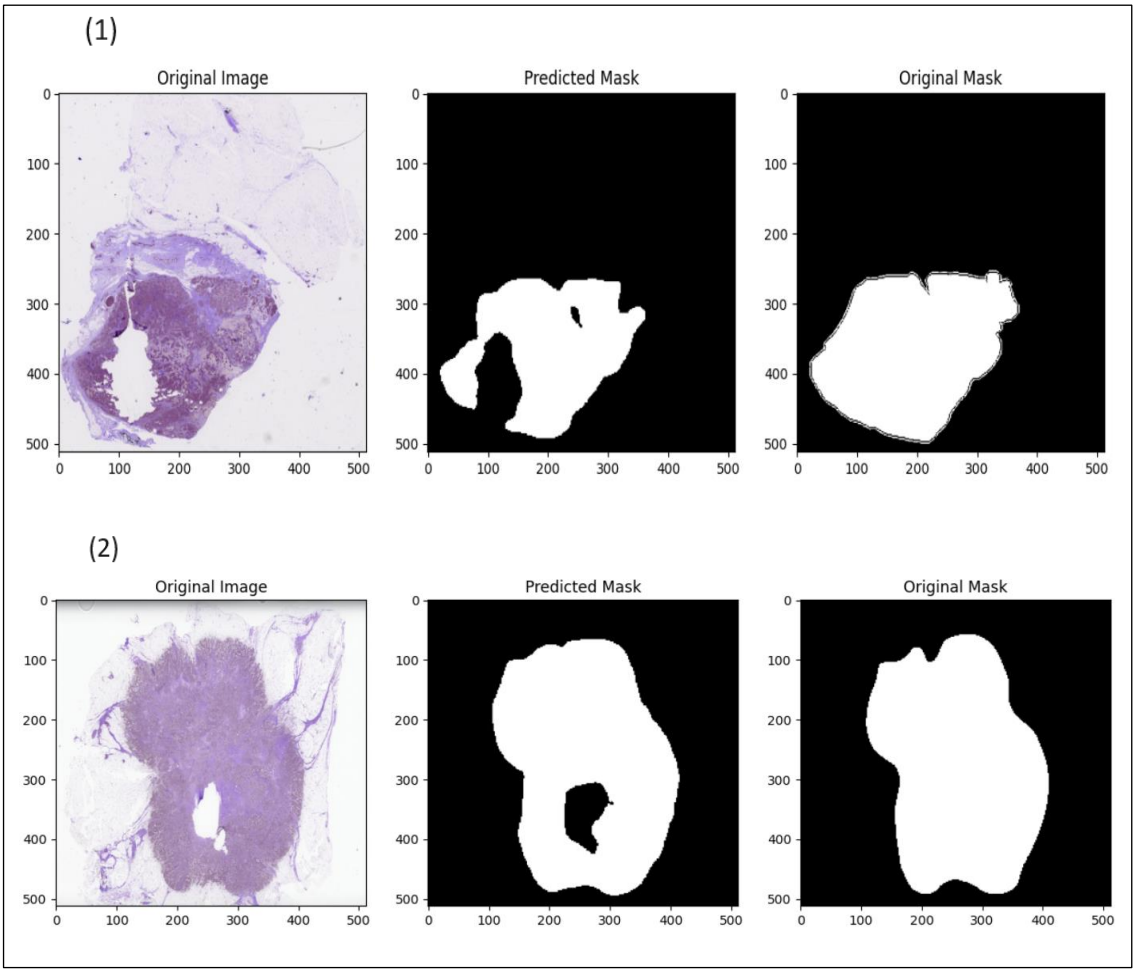
Tumor area segmentation



It is also worth mentioning that the masks in the dataset used were not very accurate in some cases where the trained model gave a more reasonable segmentation, as shown in Figure 3.4.

Figure 3.4:

Tumor area segmentation with original mask is not accurate



3.3.3 Evaluation and comparison

The dataset has been trained and experimented with using 6 different models of architectures including U-Net with 3 different backbones (Customized InceptionV3 with attention mechanism, VGG16), ENet, and PSPNet. All the models used a learning rate scheduler that reduced learning rate over epochs and used a custom loss function that combines both the Jaccard index and cross entropy.

The metrics of training the models are shown in Table B.7 in Appendix B and Figure A.17 in Appendix A.

In this comparative analysis of various deep learning segmentation models, particular focus was placed on the Jaccard index and dice coefficient as main metrics for comparison between models. The Jaccard index measures the intersection over union between the

predicted mask and the original mask which provides a fundamental measurement of a segmentation model accuracy. The dice coefficient, similarly, assesses the model performance by evaluating the balance between precision and recall, which can give sufficient insights about both sensitivity and specificity in a single metric.

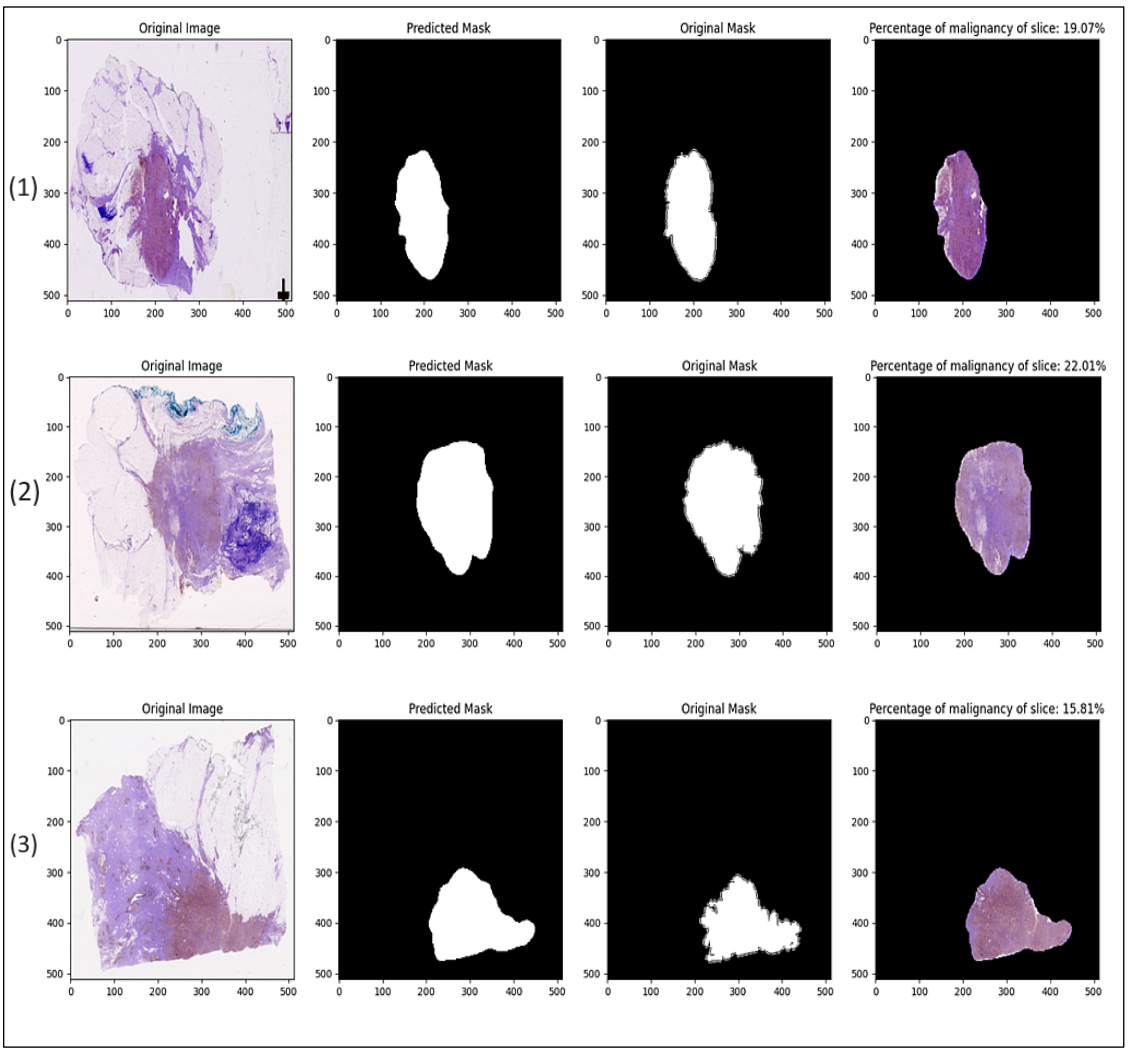
The results showed that U-Net with InceptionV3 resulted in the best overall performance, with a Jaccard index of (0.84) and dice coefficient of (0.91). On the other hand, PSPNet and U-Net with vgg16 backbone resulted in a very similar overall performance with Jaccard index (0.77) and dice coefficient of (0.86). In contrast, the ENet model, despite its high specificity, exhibited the lowest scores in both the Jaccard index (0.41) and the Dice coefficient (0.516). An example showing the difference between algorithms is provided in Figure A.25 in Appendix A.

3.3.4 Area Calculating

The last step after segmenting the tumor area in a histopathology slice is calculating the percentage of tumor in the given slice. This has been done using an iterative grabcut algorithm to remove the background from the original slice and only focus on the slice itself to calculate the percentage of the segmented mask from the slice.

Figure 3.5

Tumor area size calculation



In Figure 3.5, both the original mask and the predicted mask, and finally the size of tumor area in the slice after 5 iterations of grab cutting the foreground (slice) from the background.

3.4 Tumor bed detection and Therapy Response Classification

The treatment response classification procedure was done on several steps that include preprocessing the histopathology images, segment tumor bed, then convert the sample to only include the tumor bed, then segment the tumor area in the bed, and finally calculate tumor percentage in the bed and compare it against the dataset actual values to predict whether there is a full response, mid response, or no response.

3.4.1 Data processing and normalization results

Similar to the preprocessing done on tumor area detection, the histopathology images have been resized to (512,512), normalized using Reinhard normalization, and augmented.

After normalizing and resizing the images, they were transformed as shown in Figure A.26 in Appendix A.

3.4.2 Segmentation results

For the therapy response part, the images went through a multiple phases process, so after preprocessing the images, the images were segmented to extract the tumor bed first, then the resulted mask was multiplied by the original image, and the results were segmented to further detect the malignant area.

3.4.2.1 Stage one: Tumor bed detection

After preprocessing the images and training them using U-Net InceptionV3 segmentation model, which gave the best promising results for a similar task in tumor area segmentation, the unseen test data resulted with an 82.5% Jaccard index, 90.1% dice coefficient, 94% accuracy, 86.9% sensitivity, and 97.01% specificity. An example showing the sample of results is provided in Figure A.28 in Appendix A.

The ROI as shown in Figure A.28 in Appendix A is the last result of this phase, which takes the tumor bed area and ignores all other data from the slice. This process has been done on all dataset which generated another dataset with tumor beds only which will make it easier for the next step in the flow to focus on specific details in the histopathology images.

3.4.2.2 Stage two: Malignant detection and size calculation.

After preparing the data to contain tumor beds only, the second segmentation phase starts which focuses on extracting the malignant parts of the image, as shown in Figure A.29 in Appendix A.

After using the same U-Net InceptionV3 architecture to train a new model to segment malignancy, the results on test data were 63.3% Jaccard, 73.6% dice coefficient, 97.6% accuracy, 79.6% sensitivity, and 98.5% specificity.

The segmentation captured the main area of malignancy, and the final output of the level two segmentation was the percentage of malignancy in the tumor bed area, prepared the data for the last step which is to classify the number into a treatment level.

3.4.3 Classification results

The last part is straightforward as the predicted and actual malignancy percentages are available.

So, after comparing the malignancy percentage, the regression result was:

- Root means square error = 6.335
- R2 score = 0.9357

These results indicate a solid overall accuracy for the whole therapy response pipeline. Finally, the classification result for 4 classes (Grades 1 – 4) was 100%, as shown in Figure A.27 in Appendix A.

Chapter Four
Discussion and Conclusion

4.1 Conclusion

The thesis presents a comprehensive exploration of breast cancer classification, whole slide image analysis for malignancy detection, and therapy response classification using advanced deep learning models across different magnifications and imaging techniques. The results indicate a significant success in utilizing these models for precise cancer detection and classification.

The method used in this study, where two different magnification levels (20X and 40X) are applied to the classification of breast cancer, illustrates the subtlety that is necessary to achieve the best result in medical imaging. The purposeful incorporation of image preprocessing techniques such as normalization and segmentation is one of the aspects that improve the overall performance of the entire analytical framework. This research emphasizes the acute necessity of selecting a method appropriate for the specific requirements of image analysis. The higher results of InceptionV3 and Gradient Boosting models in various cases prove that it is crucial to adhere to a proper approach. In detail, the InceptionV3 model with summing of confidences has been successful in coming up with an accuracy of 98.7%, and when data from an external source was used it achieved 90% accuracy. Likewise, Gradient Boosting with 40X magnification achieved an accuracy of 98%, which proves the effectiveness of the mentioned methodologies in enhancing diagnostic performance.

Additionally, the analysis of whole slide images for malignancy detection, accompanied by the quantification of tumor areas, demonstrates the methodology complexity needed not only to detect cancer presence with high precision but also to measure the volume of tumor infiltration. This dual role is critical to the creation of successful treatment plans and careful monitoring of disease development, demonstrating methodological progress that underlies these accomplishments. Of significance, the use of the Jaccard index which returned an 84% score, emphasizes the accuracy in measuring tumor areas.

This research also looks into methodological advancements that can be useful in the detection of tumor beds and the categorization of therapy response, thereby contributing to personalized therapy. The possibility of the therapy results' prediction depending on a

detailed tumor description is a significant part of the strong methodological base for patient care enhancement.

This thesis emphasizes the importance of advanced methods in medical imaging particularly in changing breast cancer detection, classification, and treatment response analysis. The study also shows that these methods are effective in a clinical setting since the methods consistently produce high accuracy, sensitivity, and specificity in numerous models and imaging techniques. This can be considered as the future vision when the mentioned advanced approaches will become a part of the clinical practice and help to make personalized and precision medicine more accessible and efficient, leading to the further development of cancer care. Therefore, the studies should be continued in this area with more data sets, different magnification levels, and advanced models to provide better diagnostic and therapeutic approaches for breast cancer. This research also promotes the collaboration between computer science, medical imaging, and cancer treatment to develop new and improved ways of diagnosing and treating patients.

4.2 Limitation

One of the major limitations experienced in our study was the limited availability of medical scanning devices required to digitize histopathology slides. These instruments are critical in scanning histopathological slides to high-definition digital pictures which are suitable for analysis by artificial intelligence (AI). The unavailability of such scanning machines in our research setting, which is mainly due to their cost and the special technical knowledge required to operate them, greatly hindered the creation of a strong and internally produced dataset, and this limitation necessitated the use of the external dataset in some parts of the work.

One of the difficulties in the development of our methodology was not sharing a common language with the medical professionals. The problems we faced as an AI specialist were to link the technical aspect of AI with the clinical observations given by the medical team. This interdisciplinary project required a lot of work in coming to a common understanding of the aims, the terminology, and the interpretations. These issues are tried to be solved by continuous discussions and consultations, however, the different knowledge bases and points of view between AI and medical fields resulted in some cooperation challenges.

4.3 Recommendations

- Enhance Image Preprocessing Techniques: Further improve the accuracy and efficiency of deep learning models by continuing to improve and develop advanced image preprocessing techniques, normalization, and segmentation among others.
- Model and Approach Selection: Promote continued study on the best deep learning models and methods for different types of breast cancer image analysis, leveraging the abilities of models such as InceptionV3 in diverse situations.
- Quantitative Analysis Capabilities: Progress the performance of deep learning models for quantitative assessment of the tumor extent in whole slide images that will enable more detailed planning and monitoring.
- Interdisciplinary Collaboration: Encourage interdisciplinary collaboration between computer science, medical imaging, and oncology to utilize the combined knowledge in improving AI models for breast cancer diagnostics and treatment.
- Investment in Advanced Machine Learning Research: Champion more funding and resources for research in advanced machine learning techniques, in particular deep learning, to stimulate the development of improved medical imaging for breast cancer.
- Clinical Implementation and Evaluation: Support pilot studies and clinical trials to assess the implementation of AI models in living clinical environments, concentrating on their integration into diagnostic workflows and their effect on patient outcomes.
- Global Data Sharing and Collaboration: Promote global data sharing and collaboration initiatives to train AI models on diverse datasets, improving their generalizability and effectiveness across various populations and healthcare systems.
- Overcome Scanning Device Accessibility in Palestine: Promote the production and dissemination of cheaper scanning devices used for digitizing histopathology slides, maybe by partnering with technology companies and healthcare providers.
- Cross-Disciplinary Training Programs: Create cross-discipline training programs and workshops to foster effective communication and cooperation among AI researchers and medical professionals, so that they can understand each other’s terminologies and goals.

- Standardize Image Processing Methods: Standardize image processing and feature extraction methods to ensure consistency and reliability in various studies thus solving the problem of data preprocessing inconsistency.

4.4 Future works

- Enhance Model Validation on Diverse Datasets: Promote AI model validation on various uncorrelated datasets to measure generalization and performance in different clinical settings. This effort aims to address the issue of dataset specificity, with plans to expand validation to additional datasets in the future.

- Extended Classification: In future work, the classification model could be expanded to handle other magnifications, such as X10 or X100, allowing for a more comprehensive analysis across different levels of detail. This would improve the model's generalization capabilities across varied clinical settings.

- Integration with Prognostic Tools: Develop methods to integrate the classification model with prognostic tools to evaluate tumor progression and therapy response.

List of Abbreviations

Abbreviation	Meaning
WSI	Whole Slide Image
CNN	Conventional Neural Network
SVM	Support Vector Machines
LDA	Linear Discriminant Analysis
K-NN	K-Nearest Neighbors
RF	Random Forest
NB	Naive Bayes
ANNs	Artificial Neural Networks
IDC	Invasive Ductal Carcinoma
ThRC	Therapy Response Classification
SVC	Support Vector Classifier
ROC	Receiver Operating Characteristic curve

References

1. W. H. O. Brest Cancer. WHO [Internet]. 2021; Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

2. Health Annual Report. Palest Heal Inf Cent. 2021.

3. Abu Seir R, Najjar O. Laboratory Medicine in Palestine. EJIFCC. 2018 Dec;29(4):248–52.

4. Dettmeyer RB. Histopathology. In: Payne-James J, Byard RW, editors. Encyclopedia of Forensic and Legal Medicine (Second Edition) [Internet]. Second Edi. Oxford: Elsevier; 2016. p. 35–46. Available from: https://www.sciencedirect.com/science/article/pii/B9780128000342002020.

5. Zeiser FA, da Costa CA, Roehe AV, da Rosa Righi R, Marques NMC. Breast cancer intelligent analysis of histopathological data: A systematic review. Appl Soft Comput [Internet]. 2021;113:107886. Available from: https://www.sciencedirect.com/science/article/pii/S1568494621008085.

6. Shinohara S, Bychkov A, Munkhdelger J, Kuroda K, Yoon HS, Fujimura S, et al. Substantial improvement of histopathological diagnosis by whole-slide image-based remote consultation. Virchows Arch. 2022 Aug;481(2):295–305.

7. Kiran N, Sapna F, Kiran F, Kumar D, Raja F, Shiwlani S, et al. Digital Pathology: Transforming Diagnosis in the Digital Age. Cureus. 2023 Sep;15(9):e44620.

8. Shafi S, Parwani A V. Artificial intelligence in diagnostic pathology. Diagn Pathol. 2023 Oct;18(1):109.

9. Atallah NM, Toss MS, Verrill C, Salto-Tellez M, Snead D, Rakha EA. Potential quality pitfalls of digitalized whole slide image of breast pathology in routine practice. Mod Pathol an Off J United States Can Acad Pathol Inc. 2022 Jul;35(7):903–10.

10. Rawson R V, Adhikari C, Bierman C, Lo SN, Shklovskaya E, Rozeman EA, et al. Pathological response and tumour bed histopathological features correlate with

survival following neoadjuvant immunotherapy in stage III melanoma. Ann Oncol [Internet]. 2021;32(6):766–77. Available from: https://www.sciencedirect.com/science/article/pii/S092375342100884X.

11. Hu Y, Loizou P. Techniques for estimating the ideal binary mask. 2009.

12. Mannu GS, Wang Z, Dodwell D, Broggio J, Charman J, Darby SC. Invasive breast cancer and breast cancer death after non-screen detected ductal carcinoma in situ from 1990 to 2018 in England: population based cohort study. BMJ [Internet]. 2024;384. Available from: https://www.bmj.com/content/384/bmj-2023-075498.

13. Heller SL, Plaunova A, Gao Y. Ductal Carcinoma In Situ and Progression to Invasive Cancer: A Review of the Evidence. J Breast Imaging [Internet]. 2021;3(2):135–43. Available from: https://doi.org/10.1093/jbi/wbaa119.

14. Reinhard E, Ashikhmin M, Gooch B, Shirley P. Color Transfer between Images. IEEE Comput Graph Appl. 2001;21:34–41.

15. Cubuk ED, Zoph B, Mane D, Vasudevan V, Le Q V. AutoAugment: Learning Augmentation Policies from Data. 2019.

16. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper with Convolutions. 2014.

17. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015.

18. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015.

19. Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2020.

20. Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. Ann Stat [Internet]. 2001 [cited 2024 Mar 22];29(5):1189–232. Available from: http://www.jstor.org/stable/2699986.

21. Louppe G. Understanding Random Forests: From Theory to Practice. 2015.

22. Evgeniou T, Pontil M. Support Vector Machines: Theory and Applications. In 2001. p. 249–57.

23. Geurts P, Ernst D, Wehenkel L. Extremely Randomized Trees. Mach Learn. 2006;63:3–42.

24. Cunningham P, Delany S. k-Nearest neighbour classifiers. Mult Classif Syst. 2007;54.

25. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. CoRR [Internet]. 2015;abs/1505.0. Available from: http://arxiv.org/abs/1505.04597.

26. Oktay O, Schlemper J, Folgoc L Le, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: Learning Where to Look for the Pancreas. 2018.

27. Canziani A, Paszke A, Culurciello E. An Analysis of Deep Neural Network Models for Practical Applications. CoRR [Internet]. 2016;abs/1605.0. Available from: http://arxiv.org/abs/1605.07678.

28. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid Scene Parsing Network. CoRR [Internet]. 2016;abs/1612.0. Available from: http://arxiv.org/abs/1612.01105.

29. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings. In 2018. p. 3–11.

30. Li Y, Zhang J, Gao P, Jiang L, Chen M. Grab Cut Image Segmentation Based on Image Region. In: 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). 2018. p. 311–5.

31. Tharwat A. Classification assessment methods: a detailed tutorial. 2018.

32. Siegel AF. Chapter 4 - Landmark Summaries: Interpreting Typical Values and Percentiles. In: Siegel AF, editor. Practical Business Statistics (Sixth Edition) [Internet]. Sixth Edition. Boston: Academic Press; 2012. p. 65–93. Available from:

https://www.sciencedirect.com/science/article/pii/B9780123852083000043.

33. Hancock J. Jaccard Distance (Jaccard Index, Jaccard Similarity Coefficient). In 2004.

34. Raina V, Molchanova N, Graziani M, Malinin A, Muller H, Cuadra MB, et al. Tackling Bias in the Dice Similarity Coefficient: Introducing nDSC for White Matter Lesion Segmentation. 2023.

35. Hussain L, Aziz W, Saeed S, Rathore S, Rafique M. Automated Breast Cancer Detection Using Machine Learning Techniques by Extracting Different Feature Extracting Strategies. Proc - 17th IEEE Int Conf Trust Secur Priv Comput Commun 12th IEEE Int Conf Big Data Sci Eng Trust 2018. 2018;327–31.

36. Sharma S, Aggarwal A, Choudhury T. Breast Cancer Detection Using Machine Learning Algorithms. In: 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS). 2018. p. 114–8.

37. Amrane M, Oukid S, Gagaoua I, Ensarİ T. Breast cancer classification using machine learning. In: 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT). 2018. p. 1–4.

38. Mohammed SA, Darrab S, Noaman SA, Saake G. Analysis of Breast Cancer Detection Using Different Machine Learning Techniques. In: Tan Y, Shi Y, Tuba M, editors. Data Mining and Big Data. Singapore: Springer Singapore; 2020. p. 108–17.

39. Guo Y, Dong H, Song F, Zhu C, Liu J. Breast Cancer Histology Image Classification Based on Deep Neural Networks [Internet]. Vol. 10882 LNCS, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer International Publishing; 2018. 827–836 p. Available from: http://dx.doi.org/10.1007/978-3-319-93000-8_94.

40. Madduri A, Adusumalli SS, Katragadda HS, Dontireddy MKR, Suhasini PS. Classification of Breast Cancer Histopathological Images using Convolutional Neural Networks. Proc 8th Int Conf Signal Process Integr Networks, SPIN 2021. 2021;755–9.

41. Nawaz W, Ahmed S, Tahir A, Khan HA. Classification Of Breast Cancer Histology

Images Using ALEXNET. In: Campilho A, Karray F, ter Haar Romeny B, editors. Image Analysis and Recognition. Cham: Springer International Publishing; 2018. p. 869–76.

42. Vesal S, Ravikumar N, Davari A, Ellmann S, Maier A. Classification of Breast Cancer Histology Images Using Transfer Learning. In: Campilho A, Karray F, ter Haar Romeny B, editors. Image Analysis and Recognition. Cham: Springer International Publishing; 2018. p. 812–9.

43. Vaka AR, Soni B, K. SR. Breast cancer detection by leveraging Machine Learning. ICT Express [Internet]. 2020;6(4):320–4. Available from: https://doi.org/10.1016/j.icte.2020.04.009.

44. Awan R, Koohbanani N, Shaban M, Lisowska A, Rajpoot N. Context-Aware Learning Using Transferable Features for Classification of Breast Cancer Histology Images. In 2018. p. 788–95.

45. Yao H, Zhang X, Zhou X, Liu S. Parallel Structure Deep Neural Network Using CNN and RNN with an Attention Mechanism for Breast Cancer Histology Image Classification. Cancers (Basel) [Internet]. 2019;11(12). Available from: https://www.mdpi.com/2072-6694/11/12/1901.

46. Sethy PK, Behera SK. Automatic classification with concatenation of deep and handcrafted features of histological images for breast carcinoma diagnosis. Multimed Tools Appl [Internet]. 2022;81(7):9631–43. Available from: https://doi.org/10.1007/s11042-021-11756-5.

47. Simonyan EO, Badejo JA, Weijin JS. Histopathological breast cancer classification using CNN. Mater Today Proc [Internet]. 2023; Available from: https://www.sciencedirect.com/science/article/pii/S2214785323050605.

48. Liu Y, Liu X, Qi Y. Adaptive Threshold Learning in Frequency Domain for Classification of Breast Cancer Histopathological Images. Int J Intell Syst. 2024;2024(1):9199410.

49. Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NNC, et al.

Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. Sci Rep. 2017 Apr;7:46450.

50. Celik Y, Talo M, Yildirim O, Karabatak M, Acharya UR. Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. Pattern Recognit Lett [Internet]. 2020;133:232–9. Available from: https://www.sciencedirect.com/science/article/pii/S0167865520300891.

51. Mridha K, Kumbhani S, Jha S, Joshi D, Ghosh A, Shaw RN. Deep Learning Algorithms are used to Automatically Detection Invasive Ducal Carcinoma in Whole Slide Images. In: 2021 IEEE 6th International Conference on Computing, Communication and Automation (ICCCA). 2021. p. 123–9.

52. Peyret R, Pozin N, Sockeel S, Kammerer-Jacquet SF, Adam J, Bocciarelli C, et al. Multicenter automatic detection of invasive carcinoma on breast whole slide images. PLOS Digit Heal. 2023 Feb;2(2):e0000091.

53. Saednia K, Lagree A, Alera MA, Fleshner L, Shiner A, Law E, 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 [Internet]. 2022;12(1):9690. Available from: https://doi.org/10.1038/s41598-022-13917-4.

54. Agarap AF. Deep Learning using Rectified Linear Units (ReLU). 2019.

55. Lin M, Chen Q, Yan S. Network In Network. 2014.

56. Al-Kababji A, Bensaali F, Dakua S. Scheduling Techniques for Liver Segmentation: ReduceLROnPlateau vs OneCycleLR. In 2022. p. 204–12.

57. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2017.

58. Thiran JP, Macq B. Morphological Feature Extraction for the Classification of Digital Images of Cancerous Tissues. IEEE Trans Biomed Eng. 1996;43:1011–20.

59. Haralick R, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Trans Syst Man Cybern. 1973 Jan 1;SMC-3:610–21.

60. Belete D, D H M. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. Int J Comput Appl. 2021;44:1–12.

61. Kim C, Kim S, Kim J, Lee D, Kim S. Automated Learning Rate Scheduler for Large-batch Training. 2021.

62. Kukačka J, Golkov V, Cremers D. Regularization for Deep Learning: A Taxonomy. 2017.

63. Hagi T, Makino T, Yamasaki M, Yamashita K, Tanaka K, Saito T, et al. Pathological Regression of Lymph Nodes Better Predicts Long-term Survival in Esophageal Cancer Patients Undergoing Neoadjuvant Chemotherapy Followed by Surgery. Ann Surg. 2022 Jun;275(6):1121–9.

64. Zhou Y, Zhang C, Gao S. Breast Cancer Classification From Histopathological Images Using Resolution Adaptive Network. IEEE Access. 2022;10:35977–91.

65. Gamper J, Koohbanani NA, Benes K, Graham S, Jahanifar M, Khurram SA, et al. PanNuke Dataset Extension, Insights and Baselines. 2020.

66. Polónia A, Eloy C, Aguiar P. BACH Dataset : Grand Challenge on Breast Cancer Histology images [Internet]. Zenodo; 2020. Available from: https://doi.org/10.5281/zenodo.3632035

67. Cruz-Roa A et al. High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: application to invasive breast cancer detection [Dataset]. 2018; Available from: https://doi.org/10.5061/dryad.1g2nt41.

68. Martel, A. L., Nofech-Mozes, S., Salama, S., Akbar, S., & Peikari M. Assessment of Residual Breast Cancer Cellularity after Neoadjuvant Chemotherapy using Digital Pathology [Data set]. 2019; Available from: https://doi.org/10.7937/TCIA.2019.4YIBTJNO.

Appendices

Appendix A

Figures

Figure A.1

A percentage of cancer according to the site in woman.

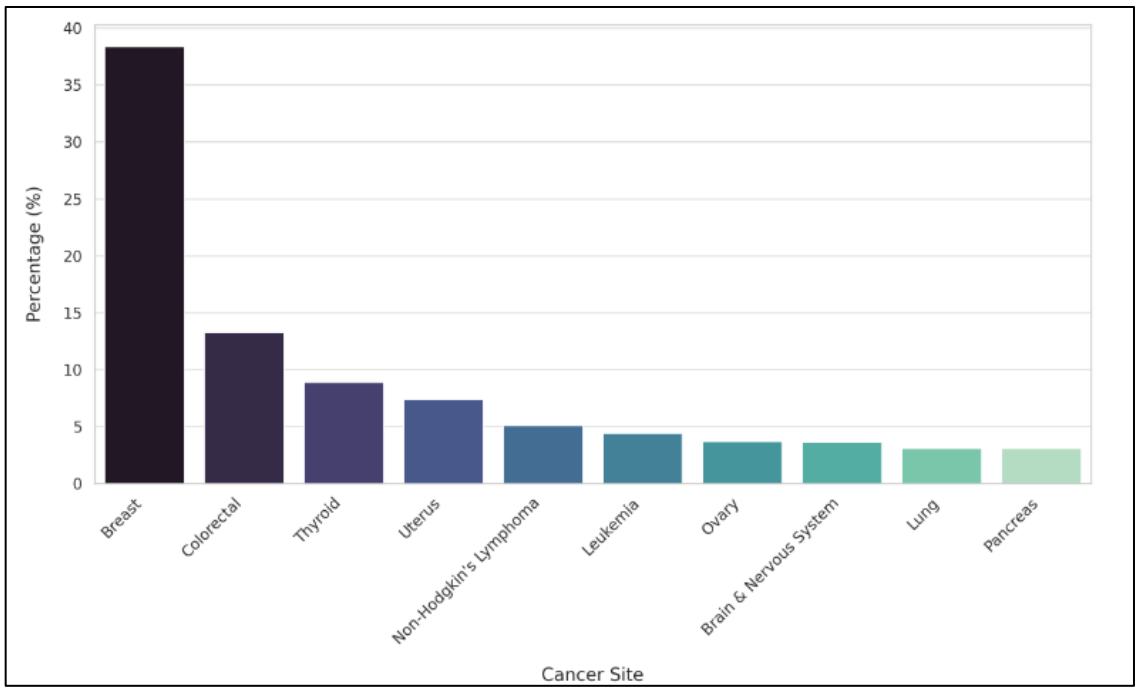


Figure A.2

Whole Tissue Slice and Corresponding Region of Interest Mask

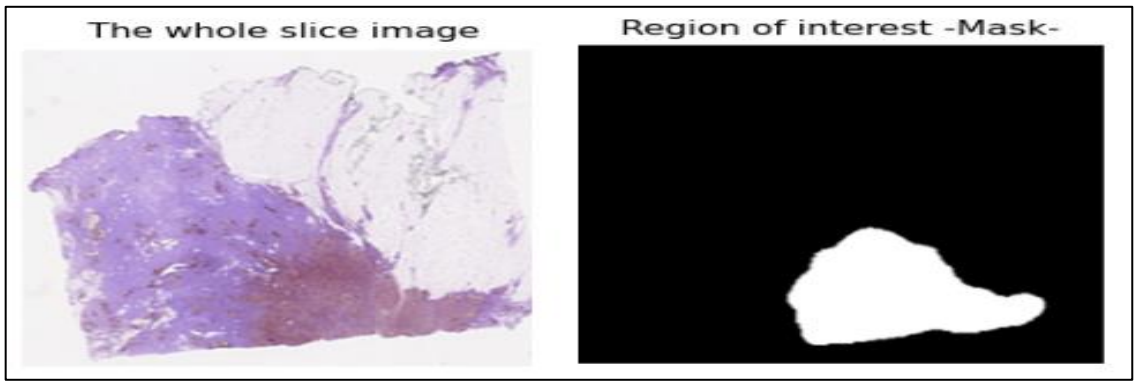


Figure A.3

Inception-V3 CNN Architecture

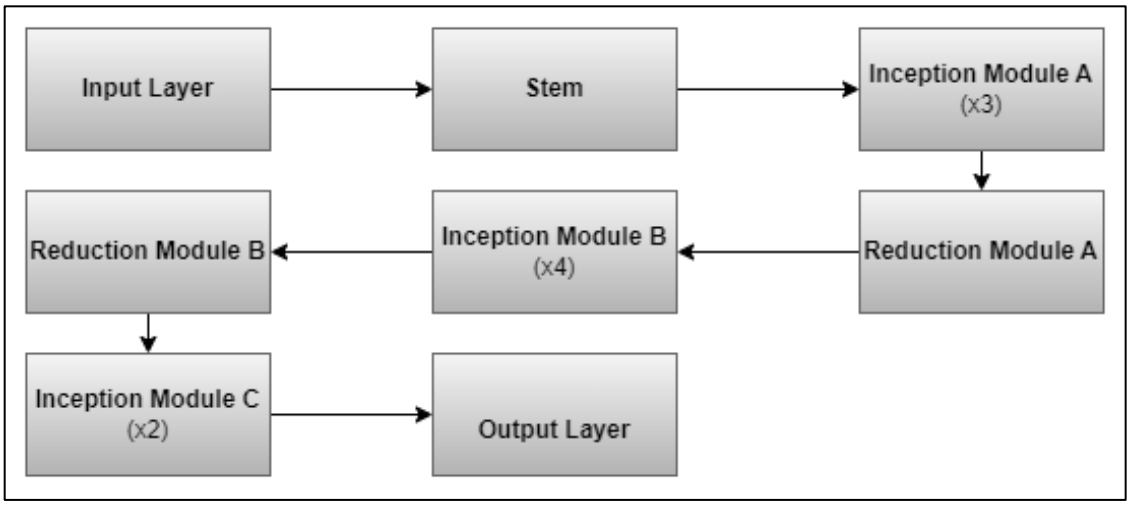


Figure A.4

U-Net Architecture

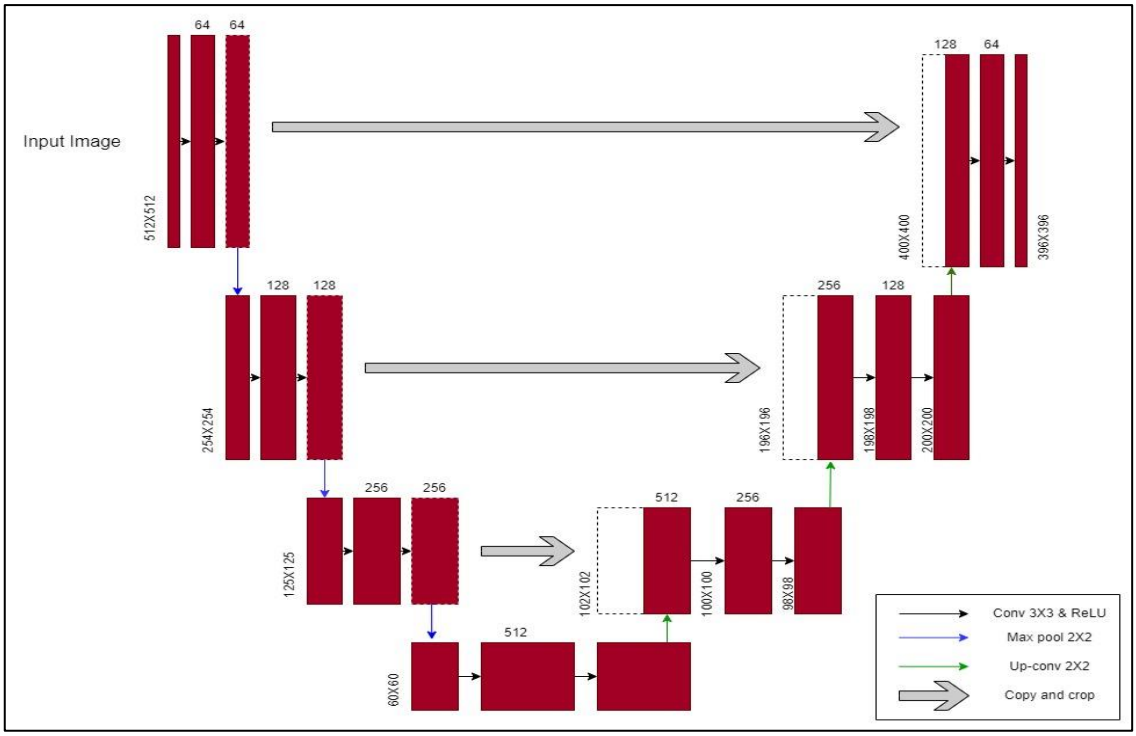


Figure A.5

Color Normalization

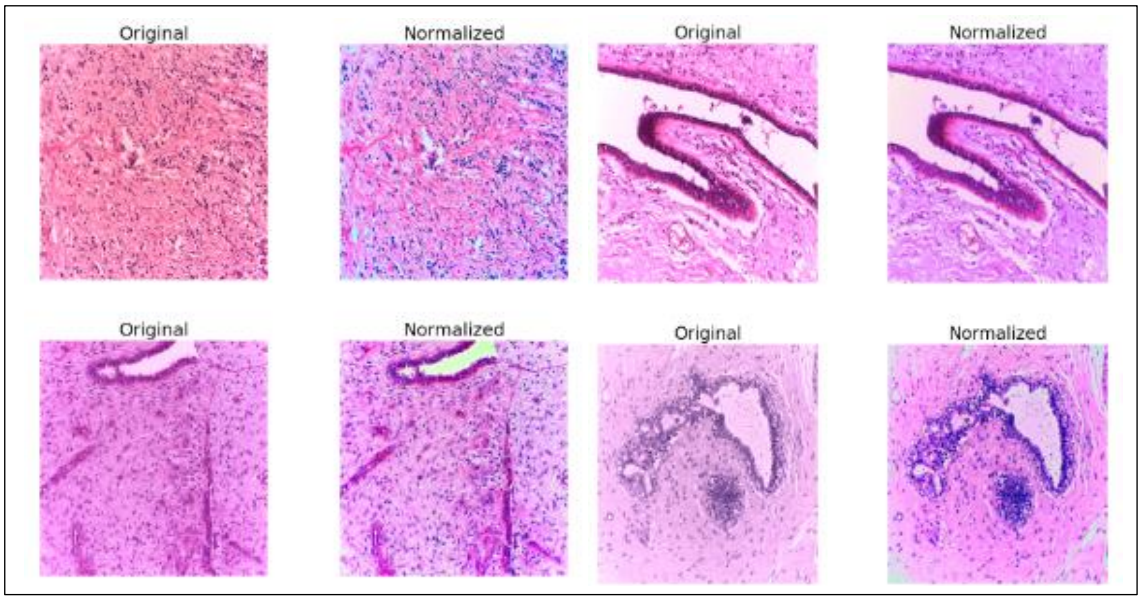


Figure A.6

X20 data pre-processing

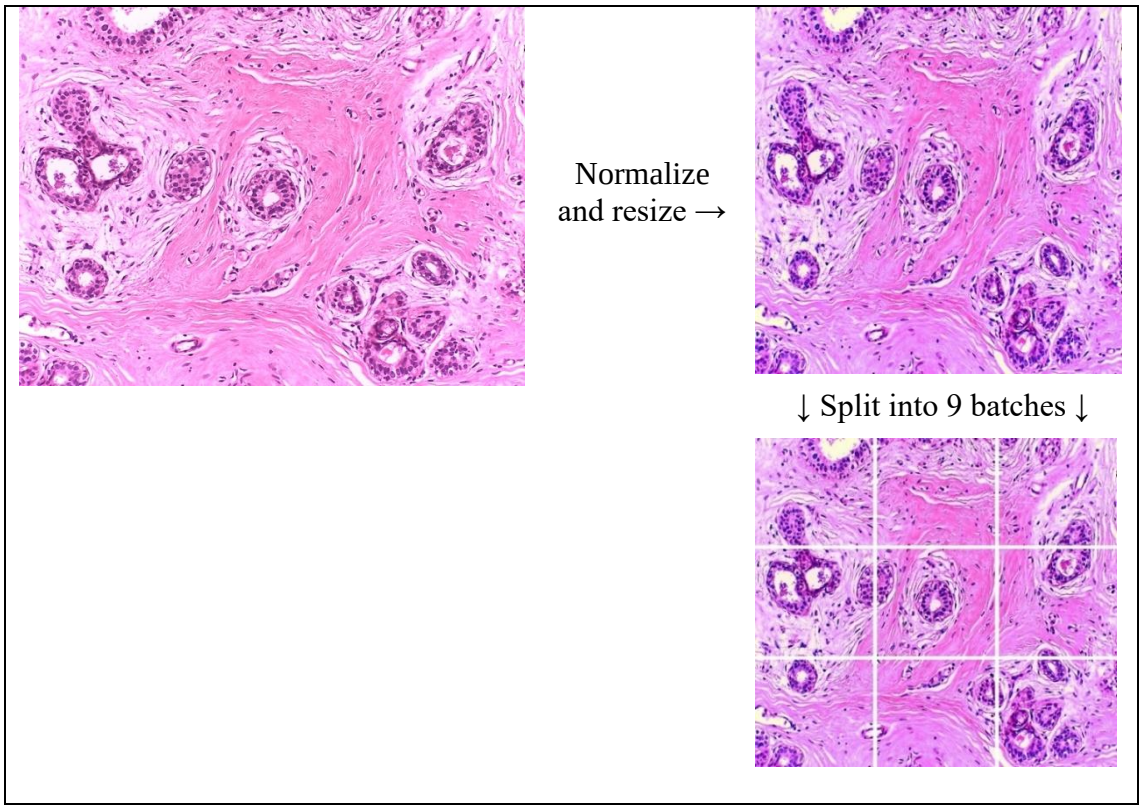


Figure A.7

X20 Breast histology classification confusion matrix using AN Najah dataset with summing confidences

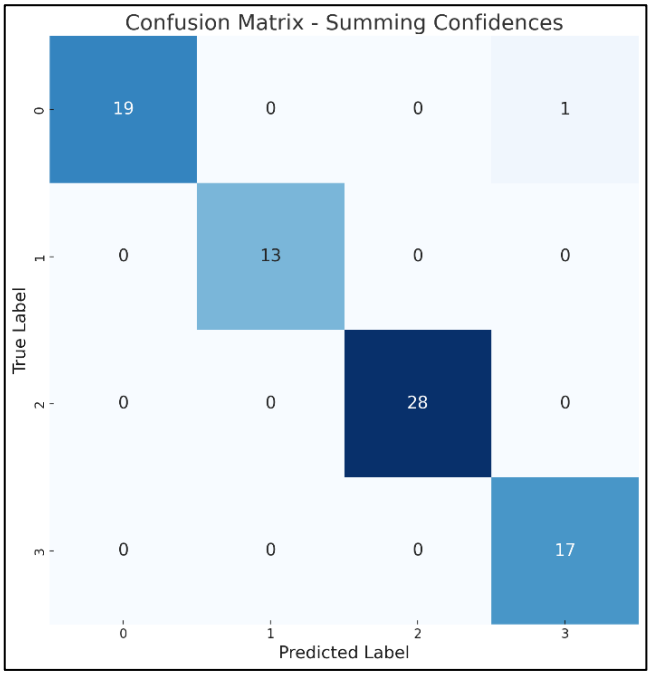


Figure A.8:

X20 Breast histology classification confusion matrix using AN Najah dataset with majority voting

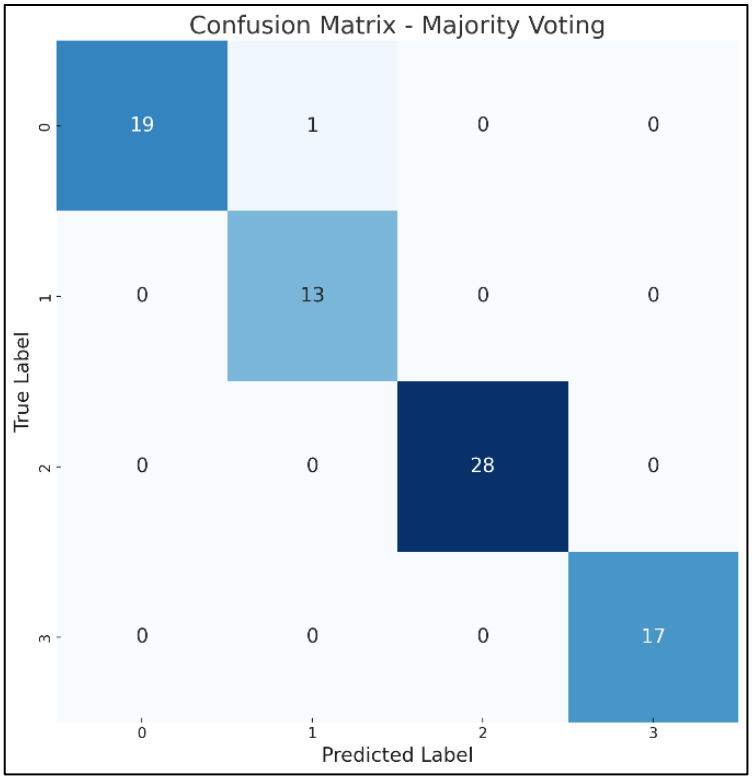


Figure A.9

X20 Breast histology classification confusion matrix using BACH dataset with summing confidences

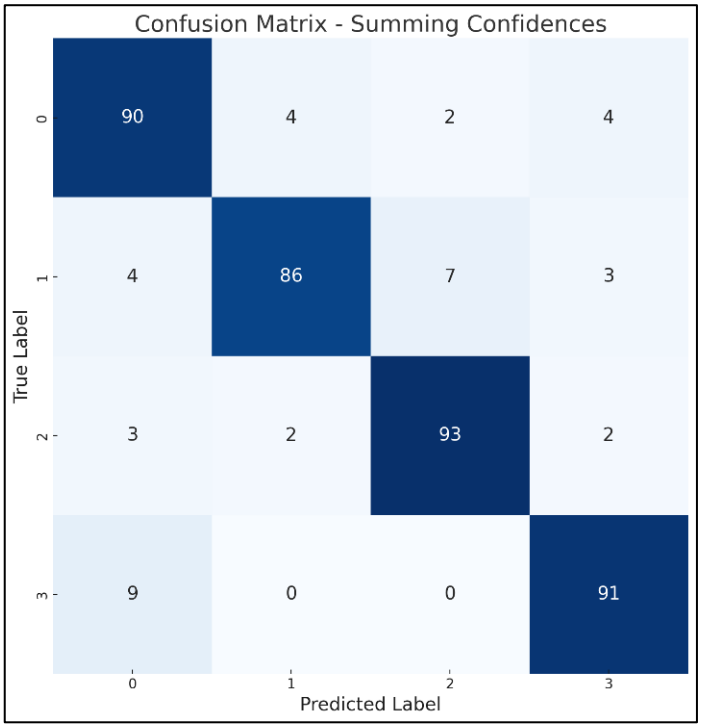


Figure A.10:

X20 nuclei classification confusion matrix using BACH dataset with majority voting

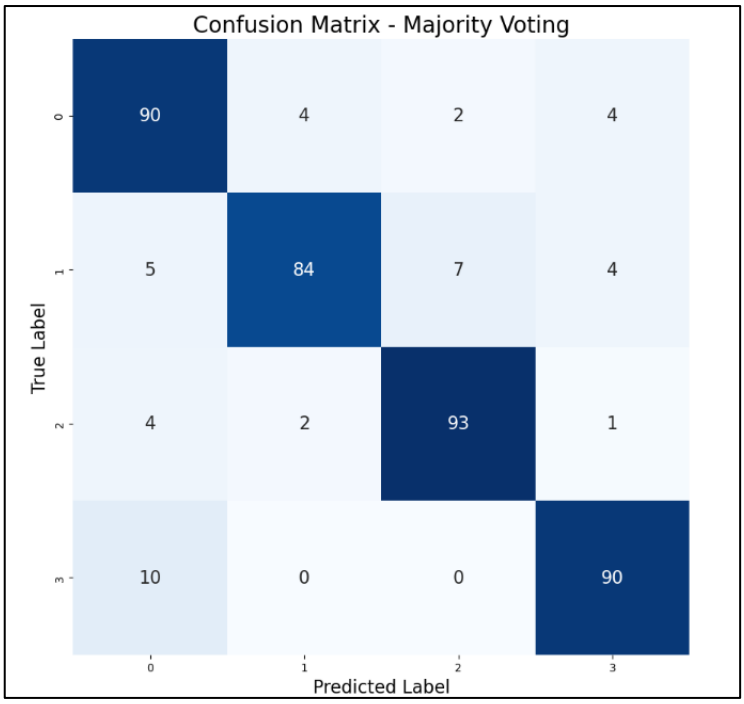


Figure A.11

X20 Breast histology classification confusion matrix using BACH dataset with summing confidences

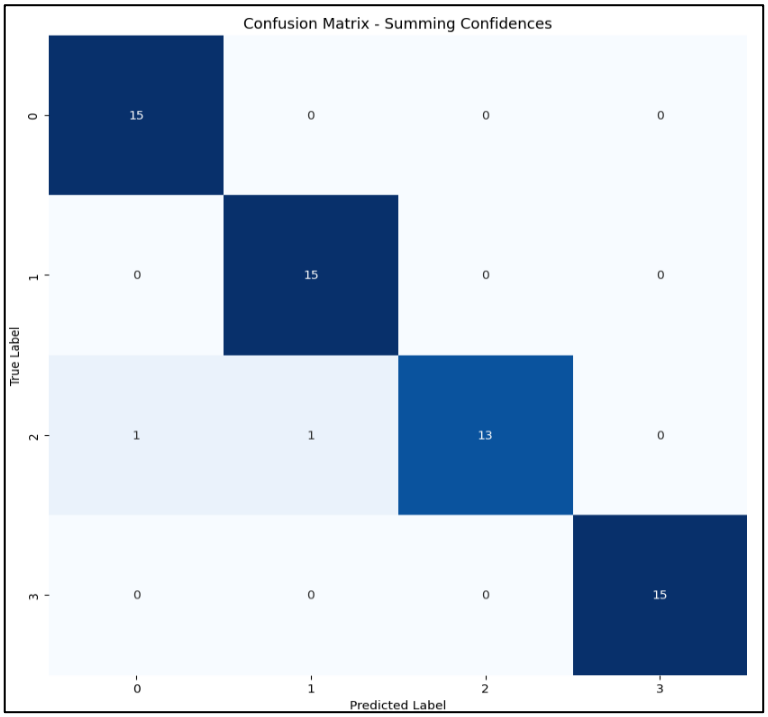


Figure A.12

X20 nuclei classification confusion matrix using BACH dataset with majority voting

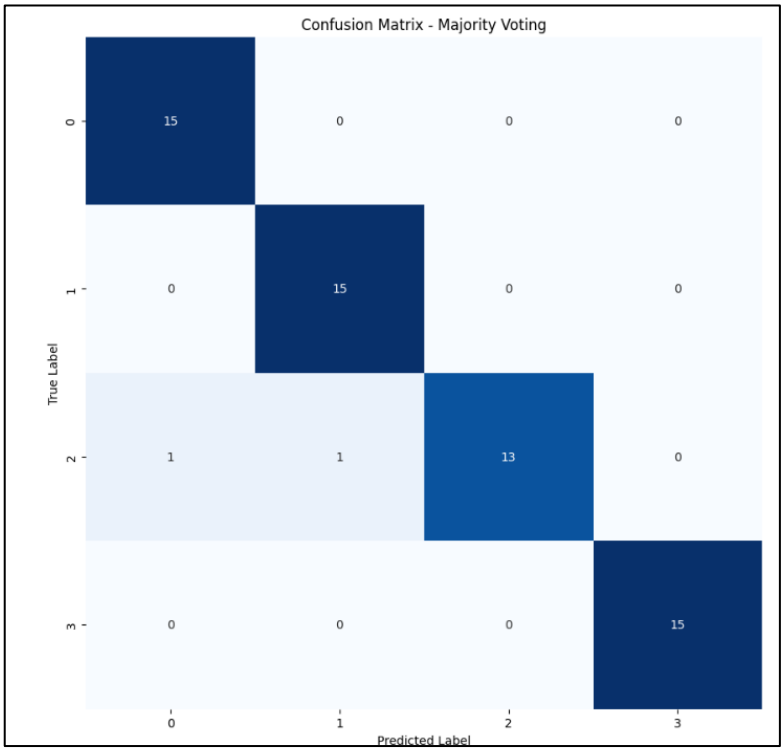


Figure A.13

X20 Breast histology classification evaluation matrix using BACH dataset and AN Najah dataset

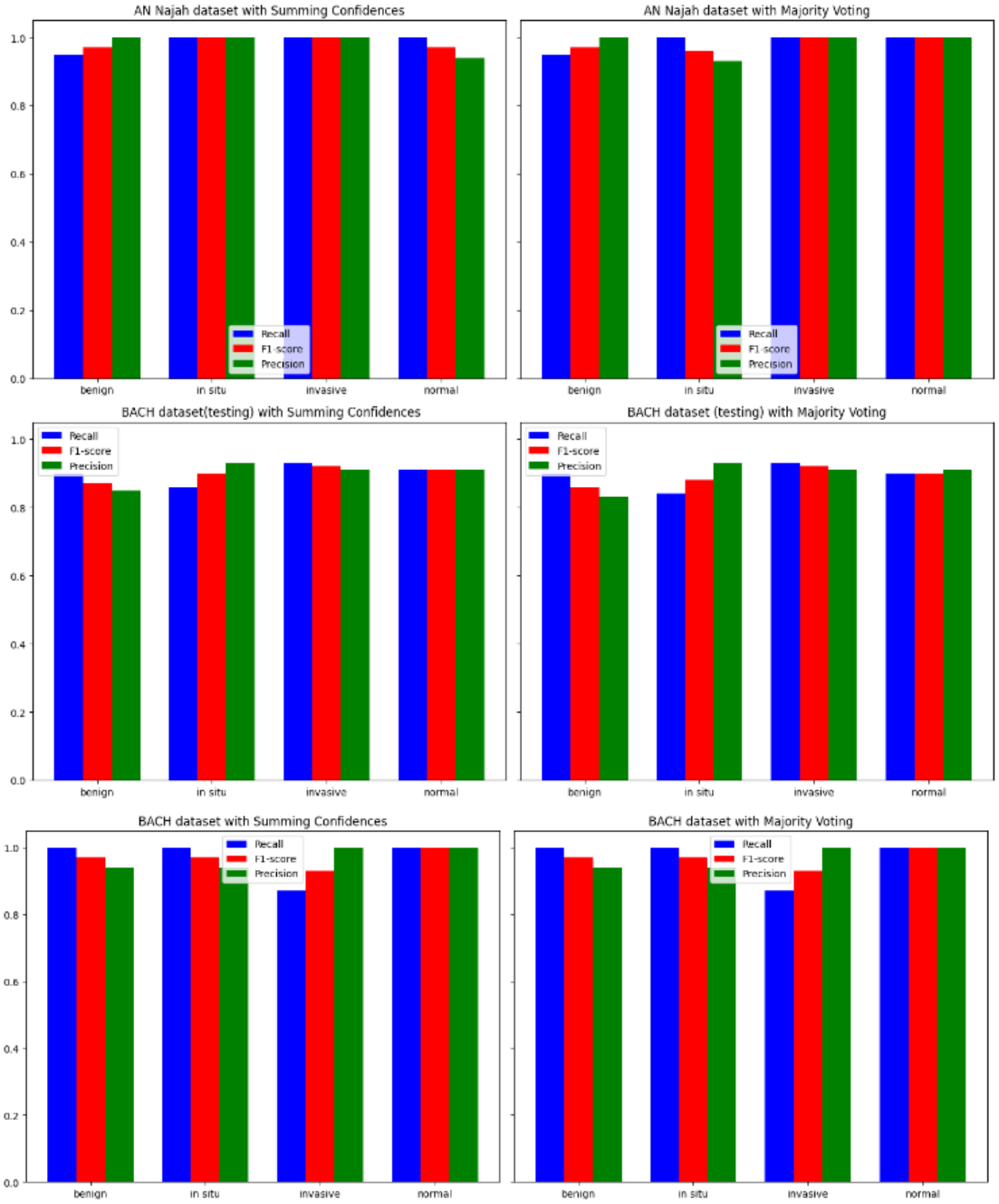


Figure A.14

Classification Metrics by deep learning Models

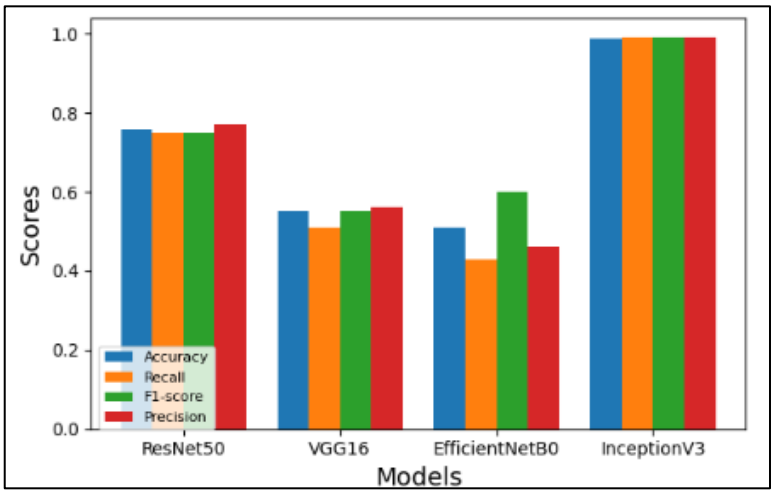


Figure A.15

Classification Metrics by deep learning Models

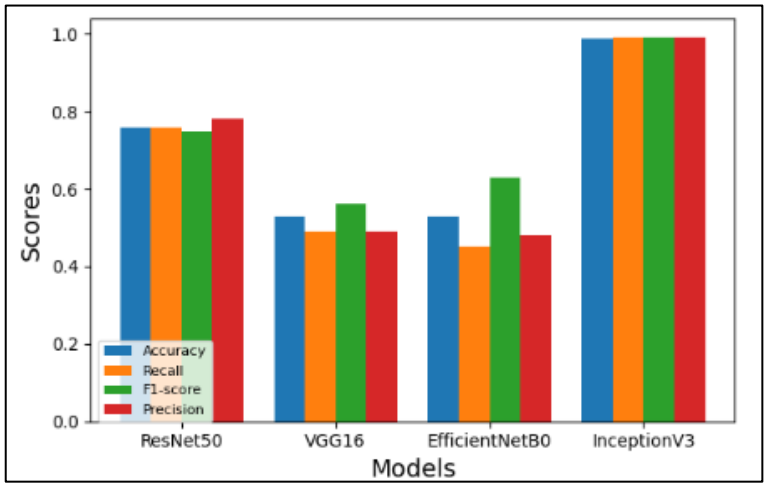


Figure A.16

Classification Metrics by Model

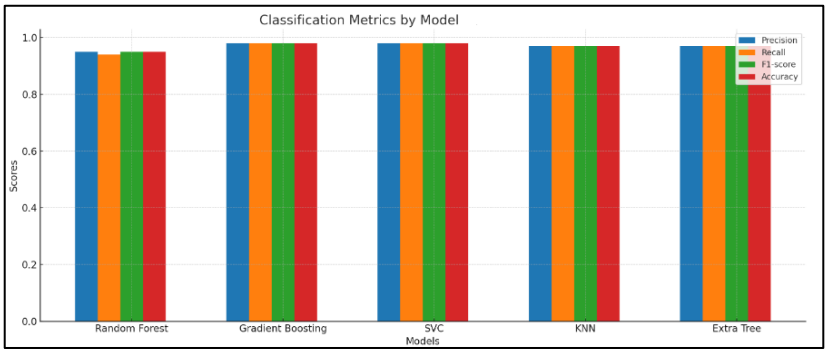


Figure A.17

Segmentation Metrics by Model

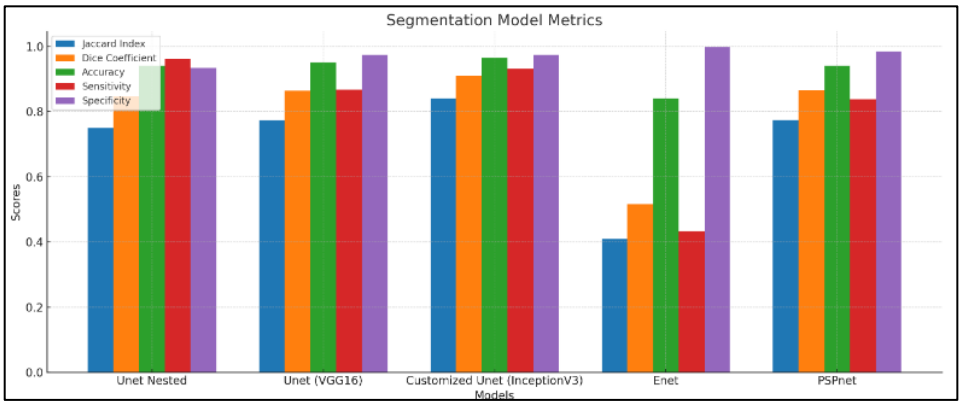


Figure A.18

Pannuke nuclei segmentation data and annotations

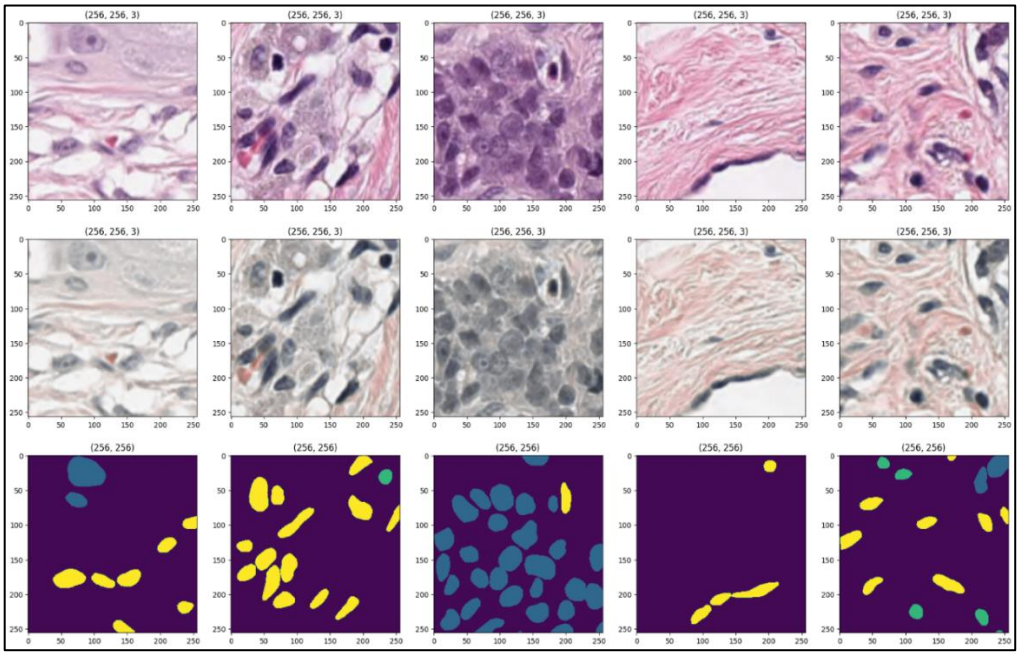


Figure A.19

Nuclei segmentation test example

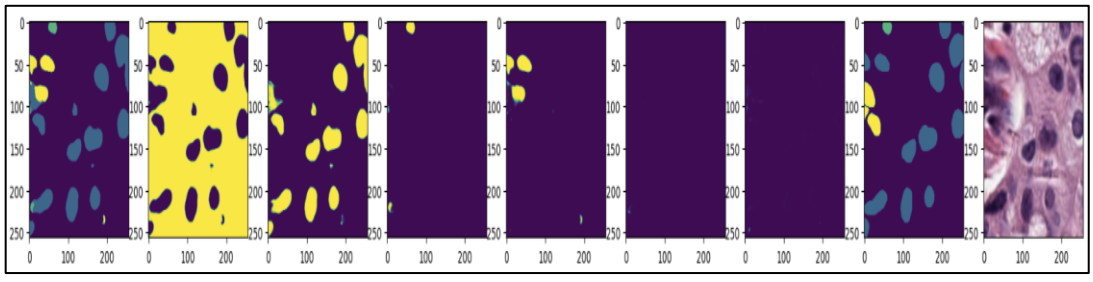


Figure A.20

Dataframe for nuclei type classification

	Id	Label	Type1 Area	Type1 Circularity	Type1 Circumference	Type1 Intensity	Type1 Texture	Type1 Count
245	BC_p1813_X40_14.jpg	Invasive	0.14231284820596682	0.8023767490928629	0.26690440881709043	0.19168467967597339	0.2295163393609544	112.0
246	BC_p1813_X40_15.jpg	Invasive	0.2118801095644742 3	0.7906566483010289	0.33599217249931107	0.29612561695450285	0.33353887692422796	152.0
247	BC_p1813_X40_16.jpg	Invasive	0.15587660287916846	0.7603217667117862	0.293876271387097 7	0.22361699834472568	0.29164502465587383	109.0
248	BC_p1813_X40_17.jpg	Invasive	0.1485868938284751	0.7792602501761595	0.2775618626830918	0.20264354322855188	0.24818567773570893	102.0
249	BC_p1813_X40_18.jpg	Invasive	0.15274635076977683	0.7716003499745615	0.2863516855979737	0.20829023125082707	0.25405174665600166	96.0
250	BC_p1813_X40_19.jpg	Invasive	0.1904960694392346 4	0.8163003872816728	0.31306440934421453	0.273804803609543 9	0.32188741059483134	132.0
251	BC_p1813_X40_20.jpg	Invasive	0.14925086372027616	0.8006463212942876	0.2714071901469942	0.2039346538798436	0.24990099471283617	115.0
252	BC_p1813_X40_21.jpg	Invasive	0.09303361353141734	0.7786123803524363	0.22052304381083085	0.14363536907067628	0.23609564156491125	104.0
253	BC_p1813_X40_22.jpg	Invasive	0.11963102339782761	0.7665488989903807	0.24901346188247267	0.16550911629166354	0.225427852514854	66.0
254	BC_p1813_X40_5.jpg	Invasive	0.14568081991215226	0.7843870670751497	0.28198786343316856	0.19723186776933418	0.2503662263670477	36.0
255	BC_p1813_X40_6.jpg	Invasive	0.13025842924895095	0.7719908701854432	0.26593181051399645	0.1744414485124961	0.20830713448771815	76.0
256	BC_p1813_X40_7.jpg	Invasive	0.13661653134566754	0.730652532363434	0.2873822366992763	0.1802839425404972	0.22937790888025505	25.0
257	BC_p1813_X40_8.jpg	Invasive	0.15454101779138382	0.807719673438059	0.2819240754860107	0.21514807545706818	0.2710178217099678	60.0

Figure A.21

X40 data set classes distribution

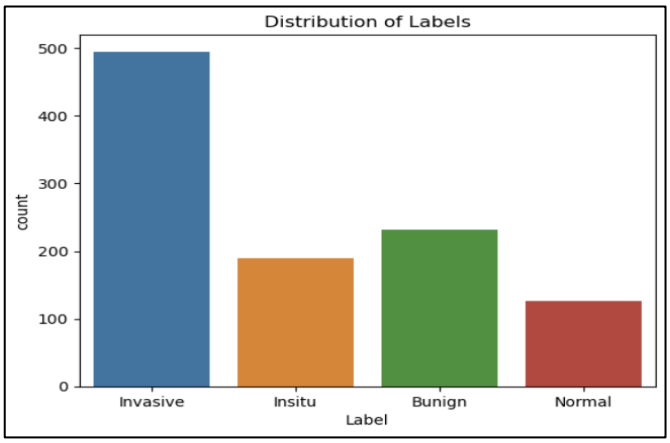


Figure A.22

Clustering the X40 image into 4 clusters

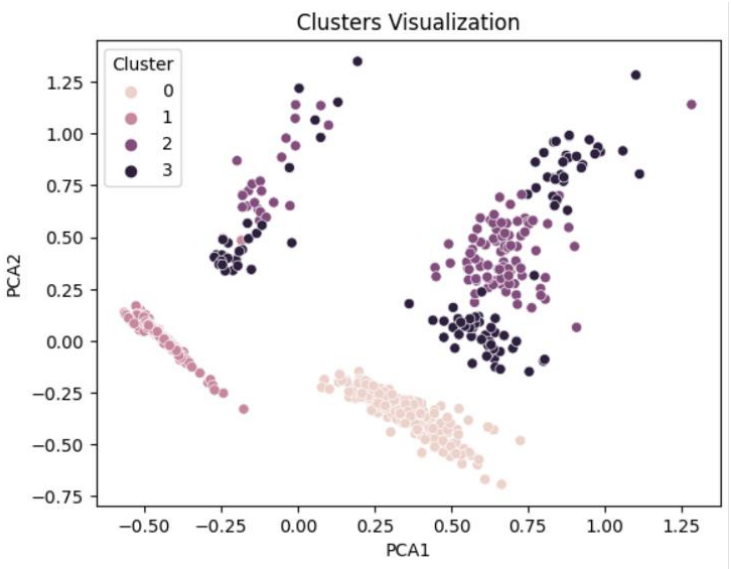


Figure A.23

Gradient boosting vs SVC classifier for classifying nuclei type

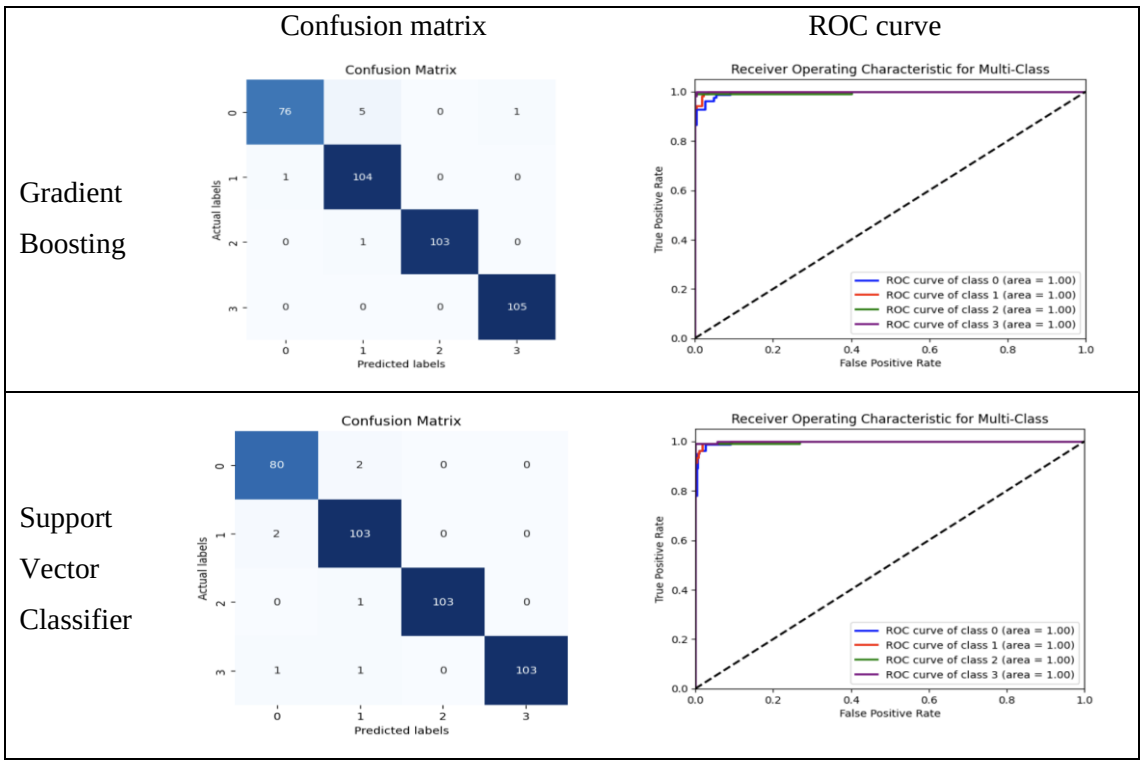


Figure A.24

Tumor area detection preprocessing

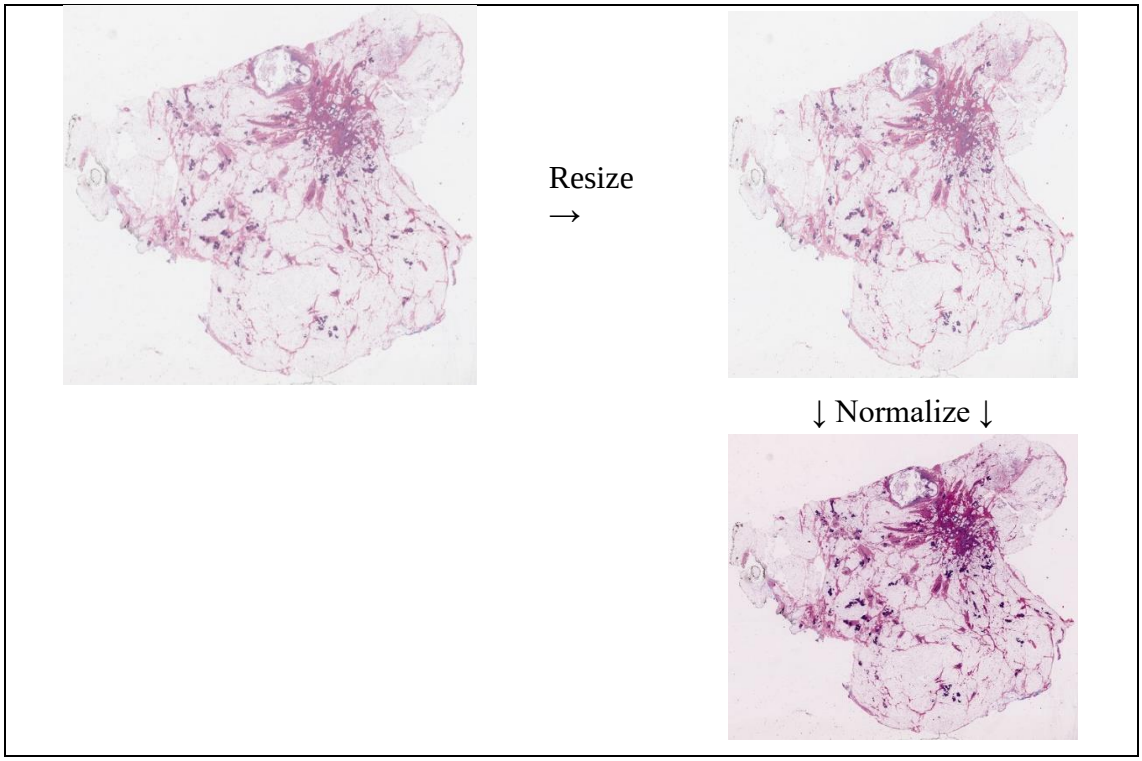


Figure A.25:

Tumor area segmentation models result comparison

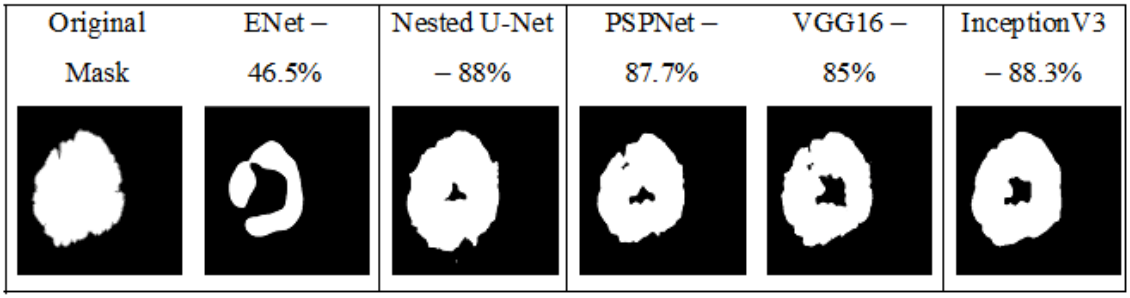


Figure A.26

Treatment response data preprocessing

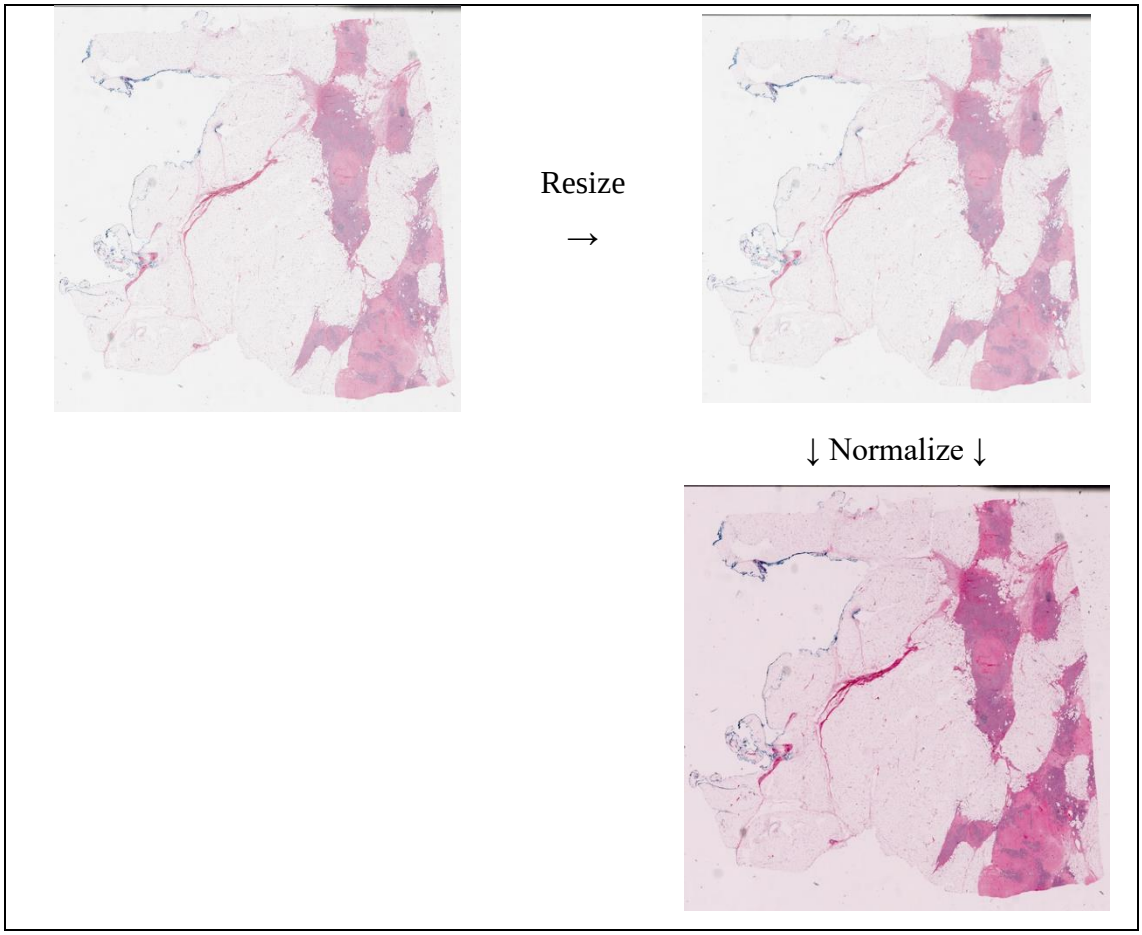


Figure A.27

Therapy response classification results

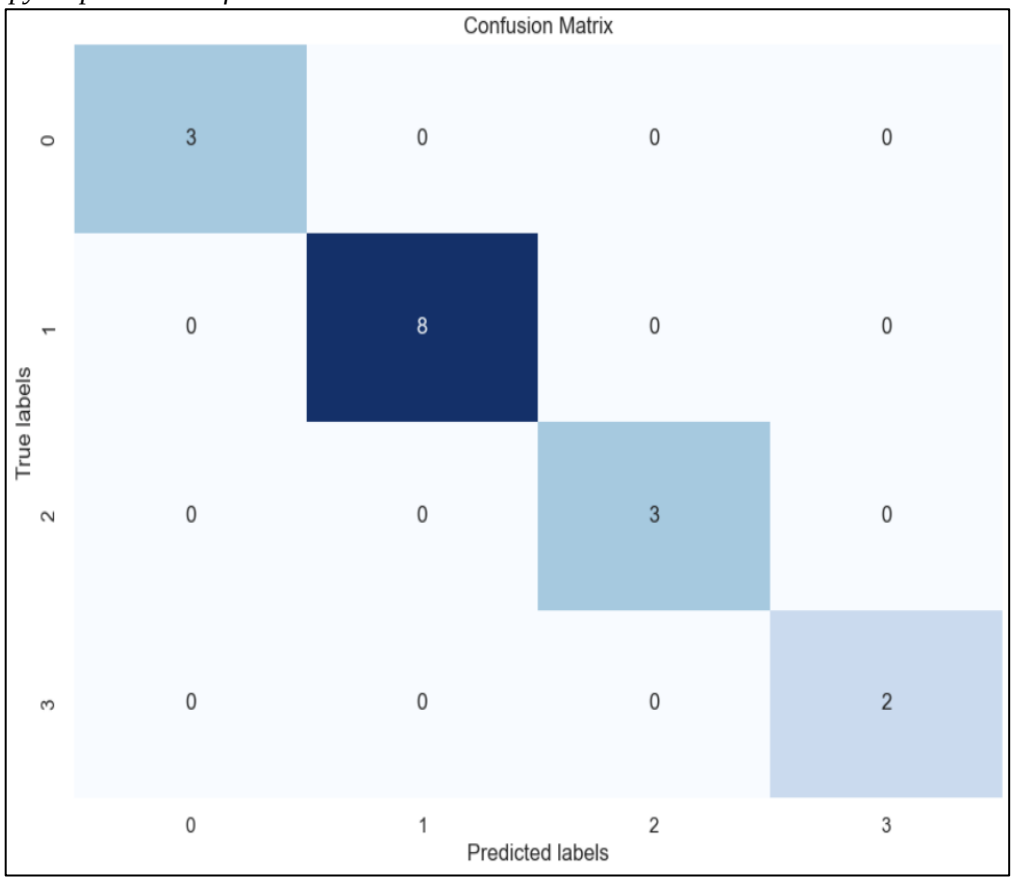


Figure A.28

Tumor bed segmentation

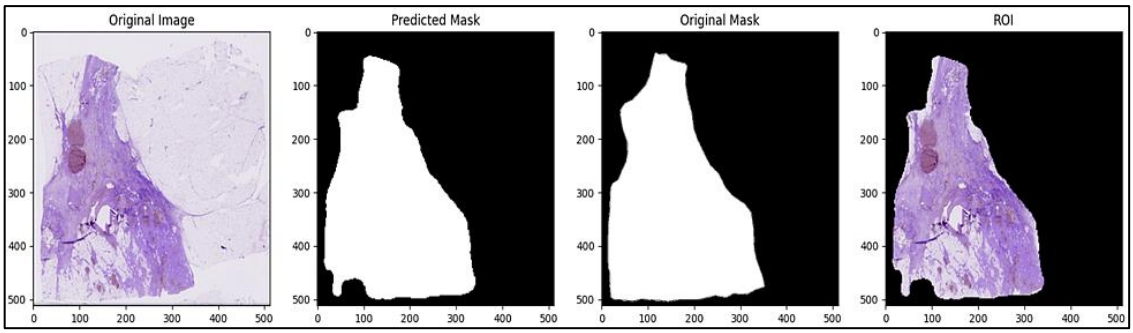
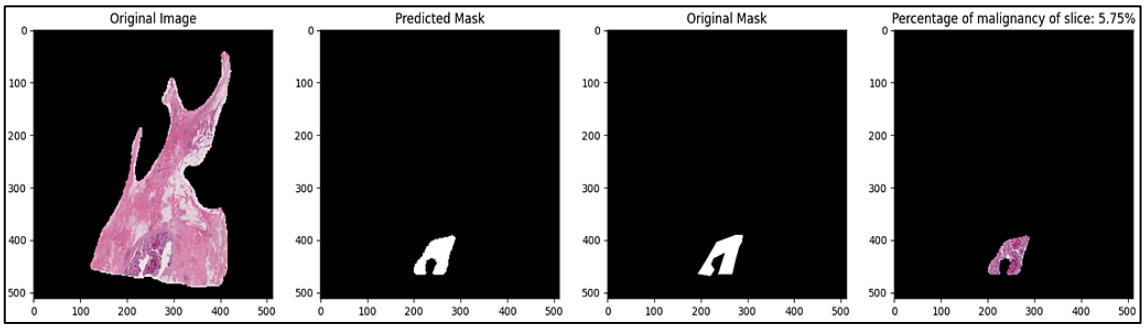


Figure A.29

Level two malignant segmentation and size calculation



Appendix B

Tables

Table B.1

An-Najah Dataset Details

Attribute	Details
Name	An-Najah Dataset
Origin	An-Najah University
Classes	4 (Normal, Invasive, In situ, Benign)
Image Count	1563 of 20x and 40x
Magnification	20x/40x objective

Table B.2

PanNuke Dataset

Attribute	Details
Name	Pan Nuke
Tissue Types	19
Total Visual Fields	481
Sampled Visual Fields	312 (randomly sampled)
Total Labeled Nuclei	205,343
Data Source	Multiple, over 20K whole slide images
Magnification	Various magnifications

Table B.3

BACH Challenge Dataset

Attribute	Details
Name	BACH Challenge
Classes	4 (Normal, Invasive, In situ, Benign)
Image Count	Normal: 100, In situ: 100, Invasive: 100, Benign: 100
Magnification	20x objective

Table B.4

Tumor Segmentation Dataset

Attribute	Details
Name	CWRU, TCGA, CINJ, HUP Datasets
Image Type	WSIs (Histopathological slide images)
Staining	Hematoxylin and Eosin
Usage	Diagnosing and grading tumors
Sources	CWRU, TCGA, CINJ, HUP
Images	584

Table B.5

Post-NAT-BRCA Dataset

Attribute	Details
Name	Post-NAT-BRCA
Image Count	96 slides
Image Size	Stored in uncompressed. svs format
Magnification	20x objective
Scanner	Aperio slide scanner
Location	Sunnybrook Health Sciences Centre
Clinical Data	Included in Excel (.xlsx) format
Patient Anonymization	Yes (anonymized IDs)
Annotation Files	Sedeen (.xml) format
Annotation Directories	"WSI_train" and "WSI_test"

Table B.6:

Classifiers comparison for x40 images

Column Label	Precision	Recall	F1-score	Support	Accuracy
Random Forest	0.95	0.94	0.95	396	0.95
Gradient Boosting	0.98	0.98	0.98	396	0.98
SVC	0.98	0.98	0.98	396	0.98
KNN	0.97	0.97	0.97	396	0.97
Extra Tree	0.97	0.97	0.97	396	0.97

Table B.7

Malignancy area segmentation - models comparison

Column Label	Jaccard index	Dice Coeff	Accuracy	Sensitivity	Specificity
U-Net Nested	0.75	0.847	0.94	0.962	0.934
U-Net (VGG16)	0.773	0.864	0.95	0.867	0.973
Customized U-Net (InceptionV3)	0.84	0.91	0.965	0.932	0.973
ENet	0.41	0.516	0.84	0.432	0.997
PSPNet	0.774	0.866	0.94	0.838	0.984

Appendix C

Data Collection producers

Producer of Data Collection

Type of Data Collected: The data comprises the histology images of breast cancer and they were classified into *in situ*, invasive, normal, and benign types.

Magnification: All histology images were taken at two different magnifications of X20 and X40.

Method of Collecting Data:

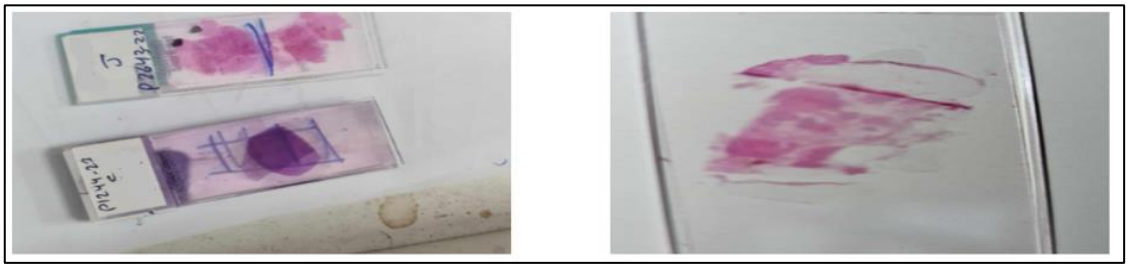
Initial Data Search: A patient record file system with all the patient records and documentation was employed in the study. The first process was to search through the file system to obtain requisite cases regarding breast cancer histology images.

Sample Identification and Extraction: After identifying the cases relevant to the study, the next step was to identify the slices of those cases that contained the samples. Every single sample was assigned a unique number to Assist in identification of the results obtained from each sample. The following slices were then selected from the batch stored for the study.

Microscopic Imaging: These samples were then collected back to the laboratory which has a microscope that is linked with a computer. The slices were examined at the appropriate magnification of X20 and X40 using the microscope. This scanning process enables the acquisition of high-quality images that are to be used for analysis in the next step.

Figure C.1:

Histopathology Slides



Microscope and Camera Configuration: The configuration of the microscope's camera, as depicted in the attached image, is detailed below:

- **Camera Model:** ICC50 HD - 38142026
- **Configuration:** Last Used
- **Captured Format:** 2048 x 1536 pixels
- **Live Format:** 640 x 480 pixels
- **Exposure Adjustments:**
    - Brightness: Set to 67%
    - Gamma: Set to 0.90
    - Saturation: Set to 103%
- **Processing Settings:**
    - Shading: None
    - Sharpening: Hard
    - Image Flip: Horizontal
    - Line Orientation: Vertical

Data Management: The scanned images were then sorted into various folders and saved in a database for further use. Furthermore, each image was marked with more detailed characteristics like the magnification scale, the number of the given sample, and histological identification.

Approval and Timing: Accomplishing the relevant permissions to start with data collection was a process that took about 5 months. This was done to ensure that all ethical and procedural requirements in the case of conducting the study were met during this period. After that, the actual data collection took approximately 3 to 4 months depending on the availability of the images and the time needed for categorizing and labeling the histology images.

Appendix D
URLS

- Source Code:

  https://github.com/omarodeh1717/Interpretation-of-Breast-Histopathology

- AN Najah dataset:

  https://drive.google.com/drive/folders/1n1McPtgRMNUcgmxf_2r4iBXuUBuav84S?usp=sharing

- Inception-v3 (section 2.1.2):

  https://drive.google.com/file/d/1lZ9tQ0scciwsjTHp37O0T27SkyUscPT5/view?usp=sharing

- U-Net: InceptionV3 Backbone with attention mechanism (section 2.3.2):

  https://drive.google.com/file/d/1rWMbBLHOcIOtqxqc9wsnkzkS6P5l0UX6/view?usp=sharing

جامعة النجاح الوطنية
كليَّة الدّراسات العليا

مناهج رؤية الحاسوب والتعلم الآلي لتفسير التشريح المرضي للثدي

إعداد
عمر حسني عوده

إشراف
د. انس طعمه

قدمت هذه الرسالة استكمالا لمتطلبات الحصول على درجة الماجستير في الذكاء الاصطناعي،
من كلية الدراسات العليا، في جامعة النجاح الوطنية، نابلس–فلسطين.

2024

مناهج رؤية الحاسوب والتعلم
الآلي لتفسير التشريح المرضي للثدي

إعداد

عمر حسني عوده

إشراف

د. انس طعمه

الملخص

سرطان الثدي، إلى جانب أنواع السرطان الأخرى، يمثل مشكلة صحية كبرى على مستوى العالم، وخصوصًا في المناطق التي تفتقر إلى الموارد الطبية، مثل فلسطين حيث يصعب تشخيص وعلاج مرضى السرطان. تعتمد الأساليب التشخيصية التقليدية بشكل أساسي على الفحص اليدوي لشرائح الأنسجة في المختبر من قبل الخبراء، وهي عملية تستغرق وقتًا طويلًا وتتطلب مهارات عالية، مما يؤدي غالبًا إلى تأخير في التشخيص والعلاج. تهدف هذه الدراسة إلى الاستفادة من تصوير الشريحة بالكامل وتقنيات معالجة الصور المتقدمة، بما في ذلك تعلم الآلة والتعلم العميق، لتحسين دقة وكفاءة تشخيص سرطان الثدي في فلسطين. تم استخدام مجموعة من الصور النسيجية من حالات سرطان الثدي، مقدمة من مستشفى النجاح في نابلس، فلسطين في هذه الدراسة. تم تطوير نموذج تصنيف آلي، قادر على تحديد الأنسجة الطبيعية والحميدة والموجودة في الموقع والمنتشرة لسرطان الثدي عند تكبيرات X20 و40X والذي تم تعزيزه بتقنيات معالجة متقدمة لتوحيد ألوان وأحجام البيانات المدربة. أسفر النموذج التصنيفي عن نتائج عالية الدقة لنماذج InceptionV3 وGradient Boosting بدقة تصل إلى 98.7% و98% على التوالي في إعدادات مختلفة. كما تم تطبيق نفس الخوارزمية على مجموعة بيانات BACH، والذي نتج عنه دقة 96.7٪، متجاوزة نتائج الدراسات الأخرى. كما تغطي الدراسة تحليل صورة الشريحة بالكامل، مع التأكيد على النهج المعقد المطلوب للكشف الصحيح عن مكان الورم وتقدير حجمه باستخدام نهج تجزئة مخصص أسفر عن درجة مؤشر Jaccardبنسبة 84%. بالإضافة إلى ذلك، تناقش الدراسة النهج الجديدة في تحديد مواقع أسرة الأورام وتقييم استجابات

العلاج، بهدف توفير أساس منهجي لتحسين نتائج رعاية المرضى في بيئات صعبة مثل فلسطين.

الكلمات المفتاحية: تشخيص سرطان الثدي، تحليل الصور الآلي؛ التصوير بالشرائح الكاملة؛ التعلم العميق في علم الأمراض؛ الرعاية الصحية في فلسطين.