3 CLASS NUCLEI SEGMENTATION FOR AUTOMATED BREAST CANCER GRADING OF H&E IMAGES: INSIGHTS AND CHALLENGES

by

Danesh Aslam

Bachelor of Engineering (Biomedical), Ryerson University, 2021

A master's research project
presented to Toronto Metropolitan University
in partial fulfillment of
the requirements for the degree of
Master of Engineering (MEng.)
in the program of
Electrical And Computer Engineering

Toronto, Ontario, Canada, 2025

Author's Declaration For Electronic Submission Of An MRP

I hereby declare that I am the sole author of this MRP. This is a true copy of the MRP, including any required final revisions.

I authorize Toronto Metropolitan University to lend this MRP to other institutions or individuals for the purpose of scholarly research.

I further authorize Toronto Metropolitan University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

3 class nuclei segmentation for automated breast cancer grading of h&e images: insights and challenges

Danesh Aslam

Master of Engineering (MEng.)

Electrical And Computer Engineering

Toronto Metropolitan University, 2025

Abstract

Nuclear grading is a critical step in the diagnosis and prognosis of breast cancer, providing valuable insights into tumor aggressiveness and treatment decisions. However, traditional manual grading is time-consuming, subjective, and prone to variability among pathologists, making automated approaches highly desirable. This study explores the application of artificial intelligence for automated nuclear grading using regions of interest in hematoxylin and eosin stained whole slide images by leveraging three class nuclei segmentation masks. The GCNN architecture for nuclei, edge, and background segmentation was modified to enhance boundary delineation and improve feature extraction. These features are combined with clinical data and processed using the Random Forest classifier to predict nuclear grade on a private breast cancer dataset achieving an accuracy of 0.706. The findings demonstrate the potential of AI driven nuclear grading to enhance the efficiency and reliability of histopathological assessment, reducing inter-observer variability and improving diagnostic reproducibility. This work highlights the transformative role of AI in digital pathology for objective breast cancer grading systems.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Breast cancer continues to be a significant public health concern with recent statistics indicating a persistent rise in incident rates. In 2024, an estimated 30,500 Canadian women are expected to be diagnosed with breast cancer, accounting for 25% of all new cancer cases among women [1]. Early and accurate histopatholgical diagnosis is critical for effective treatment planning and improved patient outcomes. Histopathology serves as the gold standard for breast cancer diagnosis, classification and prognosis [2]. It involves the microscopic examination of tissue samples stained with hematoxylin and eosin (H&E) to assess cellular and structural abnormalities. Hematoxylin stains cell nuclei a deep blue or purple, while eosin provides contrast by staining the extracellular matrix and cytoplasm in varying shades of pink. This combination of staining allows patholgoists to evaluate key histopathological features, including tissue architecture, nuclear morphology, and mitotic activity, which are all essential for grading and classification [3]. Despite its critical role for breast cancer diagnosis, histopathological assessment remains a subjective and labour intenstive process. Patologists must rely on experience and visual interpretation to identify malignant features to make a diagnosis [4]. Inter-observer discrepancies are common and studies have reported significant variability in grading assignments even among expert pathologists [5]. In addition, manual identification of nuclei is labor intensive and error prone, emphasizing the necessity for robust automated approaches to assist pathologists in accurate nuclei detection [6].

1.1   Nuclear Grading

Nuclear grading is a key histopathological parameter used in the assessment of breast cancer. It provides a quantitative and qualitative evaluation of nuclear morphology. Nuclear grading

1

is essential for determining tumor aggressiveness, predicting patient prognosis and for guiding treatment plans. The Nottingham Grading System is the most widely used grading schemes for breast cancer [7]. The grading process evaluates the degree of nuclear pleomorphism, tubule formation, and mitotic count. Each of these three features are assigned a score from 1 to 3, where the total sum determines the final histologic grade of the tumor.

Nuclear pleomorphism refers to the variability in nuclear size and shape within the tumor cells [8]. With a score of 1, well differentiated tumors, nuclei appear uniform in shape and size resembling normal breast epithelial cells. With a score of 3, poorly differentiated tumors, the nuclei exhibit significant variability. The nuclei tend to appear enlarged, have irregular shapes, and appear darker stained from increased DNA content [9].

Tubule formation assesses how much of the tumor retains normal ductal structures. A score of 1 indicates that more then 75% of the tumor consists of tubule structures and the tumor cells still form duct-like arrangements. A score of 3 is assigned when less than 10% of the tumor retains tubule formation. These cells tend to be disorganized and do not form glandular structures.

Mitotic count reflects how rapidly the tumor cells are dividing. Mitosis is the process of two genetically identical daughter cells dividing from one cell. To score this feature, pathologists count mitotic events in ten high power fields (HPFs) during visual inspection of tissue [10]. A score of 1 is assigned where few mitotic figures are observed per HPFs. A score of 3 is when many mitotic figures are identified per 10 HPF.

1.2   Whole Slide Imaging

Whole slide images are digitized images of histopathological slides produced by digital scanners. Scanners produce high-resolution images called Whole Slide Images (WSIs) by scanning glass slides at high magnifications. WSI are stored in pyramidal formats to hold multiple magnification levels and tend to be up to 10GB in size. WSI are scanned to up to 40x magnification to ensure cell morphology is captured, replicating the details that would be visible under a traditional light microscope. Using the higher magnification levels, AI based tools are developed for automated robust analysis mimicking the work of a pathologist. WSI serves as a foundational step in digital pathology workflows and allows for data sharing, image analysis and remote diagnostics [11].

Given the advantages of WSI, technical challenges appear due to the size of the data. To manage this, WSI are typically divided or tiled into smaller fixed sized image regions which can be processed. Furthermore, the quality of WSI can vary due to staining inconsistencies, scanner calibration, and tissue preparation protocols. Such issues can introduce variability in pixel intensity and texture, but can be addressed with image processing.

1.3   Research Objectives

The main objectives of this project was to improve upon current capabilities of three class segmentation model with a focus on clustered and overlapping nuclei and to perform automatic nuclear grading on whole slide images only using 10 HPFs as regions of interest (ROIs).

1.4   Contributions

The contributions of this project are listed as:

- Developed a robust metric to evaluate segmentation of clustered nuclei

- Modified the GCNN architecture to improve results of clustered nuclei segmentation

- Developed a pipeline combining 3 class segmentation with machine learning based grading using clinical and imaging features

1.5   Datasets

1.5.1   Tumor Proliferation Assessment Challenge 2016 (TUPAC)

The first dataset is obtained from the publicly available Tumor Proliferation Assessment Challenge 2016 (TUPAC) [12] dataset which consists of breast cancer histopathology whole slide images (WSIs). A total of 618 RGB images were patched to a size of 256×256 pixels at a $40\times$ magnification from WSIs. These images were segmented by a research student for 3 class segmentation consisting of nuclei, nuclei edge, and background classes. A larger subset of the TUPAC data was used to train the segmentation models and the remaining images were used for testing.

1.5.2   The Cancer Genome Atlas (TCGA)

The second dataset is from The Cancer Genome Atlas (TCGA) dataset [13] containing breast invasive carcinomas cases. From this dataset, 184 RGB images are patched to a size of 256×256 pixels at a 40× magnification from WSIs and were also segmented for 3 classes. The TCGA dataset was used only as a testing dataset for segmentation models.

1.5.3   Ontario Institute of Cancer Research Dataset (OICR)

The third dataset is a private dataset obtained from Ontario Institute of Cancer Research (OICR). Consisting of 106 WSIs derived from 103 patients, these images were collected from five distinct data centers. Each WSI was scanned at a 40× magnification level and clinical data regarding the tumor sample, patient history and any ongoing therapy was provided if the information was available. WSIs were annotated by a pathologist, who delineated 10 High Power Fields (HPFs) using bounding boxes. Additionally, the pathologist performed a comprehensive Modified Scarff-Bloom-Richarson grading of the WSIs.

Chapter 2

Literature Review

The advancement of digital pathology, computing power, and deep learning algorithms has transformed the automated analysis of histological samples in the diagnosis of breast cancer. Traditionally, nuclear segmentation and grading have been performed manually by pathologists, but with the availability of high resolution WSIs and Computer Aided Diagnosis (CAD) tools, there has been increasing interest in automating segmentation and nuclear grading to improve diagnostic accuracy and efficiency [14]. Early methods of nuclei segmentation relied on handcrafted feature based techniques. The watershed algorithm has been widely used to separate touching nuclei by detecting intensity gradients and applying morphological operations to split nuclear clusters [15]. However, this method is sensitive to over segmentation and requires post processing steps. Both simple thresholding [16] and adaptive thresholding [17, 18] distinguish foreground from background based on pixel intensity. These methods may be computationally efficient, but perform poorly in dense tumor regions with varying stain intensities [19]. The issues with these techniques is that they cannot generalize nuclei morphology [20].

Machine learning techniques have been explored for nuclei segmentation. In [21] cells are segmented based on classification of color-textures of pixel areas. Similarly, in [22] pixel are classified using linear discriminant classifier (LDA) in the color space of a reference image. The prevalent issue with machine learning techniques is their reliance on features that require extensive domain knowledge [13, 23].

Convolutional Neural Networks (CNNs), have emerged as a transformative approach in the domain of histopathology. With advancements of deep learning and CNNs, fully automated segmentation has become a major research focus. CNN-based approaches leverage spatial hierarchies and multi scale feature learning to improve segmentation performance [24]. The U-Net architec-

ture has become the standard for biomedical image segmentation by using skip connections and a symmetric encoder-decoder structure [25]. In [26] the U-Net is used for semantic segmentation of nuclei. In [27] individual nuclei are identified after segmentation using a U-Net like architecture and in [28] CNNs including the U-Net, perform segmentation with a regression task on nuclei distance maps. Such models struggle in the scenario of small overlapping nuclei, nuclei with touching boundaries and even small nuclei. Building upon the success of the U-Net, several variants have been proposed to enhance segmentation performance, particularly in challenging scenarios. In [29] a DenseRes-Unet is proposed for overlapped nuclei from multi organ histopathology images. This work uses dense blocks in the final layer of the encoder for diverse learning and distance maps for nuclei center detection. In [30] the ASPPU-Net is proposed for segmenting highly overlapping nuclei. It uses concave point detection technique after an atrous convolution bottleneck. A three stream model called Triple U-Net [31] has a stream for RGB, segmentation, and haematoxylin feature extraction. [32] proposes a lightweight model WaveSeg-UNet for touching nuclei specifically for nuclei counting. It uses a combination of atrous spatial pyramid pooling and atrous spatial channel pyramid pooling to detect small nuclei. For accurate counting, watershed is applied for post processing to refine boundaries.

Previous studies investigated ultrasound biomarkers for the diagnosis and prognosis of breast cancer [33, 34]. However, the use of histopathology images and manual analysis are considered the gold standard [35, 36, 37]. The introduction of Machine Learning (ML) algorithms has allowed for a new research focus in the assessment of cancer cases using quantitative digital histopathology. Features developed from digital histopathology are derived as pathomic features. The common CAD flowchart begins with preprocessing of a histology image, then segmentation, followed by feature extraction and dimension reduction, then classification and assessment, and finally grading results [37]. In [38], high and low grades of cancer were classified using Support Vector Machines (SVM). More than 3400 image features were initially extracted and spectral clustering was implemented to reduce the feature dimension. In [39], cell nuclei detection is performed, and four classifier types are compared: linear, quadratic, neural network, and decision tree. Feature types can even be combined as in [40] radiomic and pathomic features are used with LDA to predict recurrence of lung cancer. CNNs have also been leveraged for automating the grading of breast cancer. "Databiox" is a well annotated and standardized dataset consisting of Invasive Ductal

Carcinoma (IDC) diagnosed subjects with grade classification [41]. On this dataset, BCNet [42] uses transfer learning from a VGG16 [43] CNN without finetuning for successful grading. In [44] multistage transfer learning is performed with domain adaption from ImageNet and histopathology adaptations using the BreaKHis dataset [45]. First, 7 state of the art classifiers are pretrained on ImageNet and classify the BreaKHis dataset into 8 types of cancers. These classifiers are then used again to train on and effectively classify the Databiox dataset.

Chapter 3

Nuclei Segmentation

3.1 Introdution

This chapter presents the work to improve an existing architecture for the segmentation of nuclei in breast cancer pathology images.

3.2 Material and Methods

3.2.1 Data

For segmentation experiments the TUPAC and TCGA datasets were used. The slides are scanned at an objective magnification of 40x and are accompanied with .xml annotations of nuclei and their edges. Accompanying each training image and ground truth pair, a precomputed distance weight map and the gradient of the image are also produced. Such weight maps have been used to compensate for the different frequency of pixels in different classes and to learn separated borders between cells [25]. The weight map, w(x), is defined in Equation 3.1, for each pixel x, and where d_1 is the pixel distance to the border of the nearest cell and d_2 is the pixel distance to the border of the second nearest cell and w_o is set to 10 and σ is set to 5 pixels.

$$\text{w(x)} = w_c + w_o * exp(\frac{d_1(x) + d_2(x)^2}{2\sigma^2}) \tag{3.1}$$

The gradients of each image is obtained using the sobel filter. Example of an image, ground truth, gradient and weight map set is show in Figure 3.2. In the experiments the 618 patches from the TUPAC dataset is split into 347 training, 86 validation and 185 test images. The 184 images
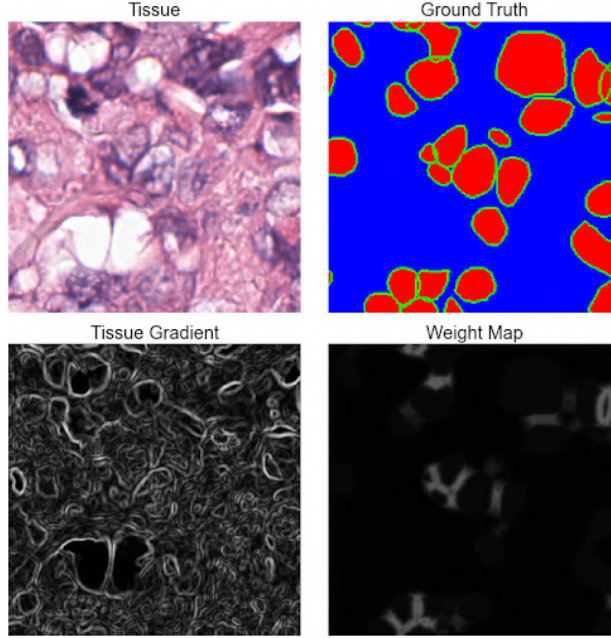
Figure 3.1: Sample Image set from TUPAC dataset. Tissue image (top left), Ground truth (top right), Tissue gradient (bottom left), Weight map (bottom right)

from the TCGA dataset are also used as a test set. The test image set is composed of the patches with the most clustered nuclei which is found using the associated weight maps. The median of the max value of each weight map is used to select a threshold to select test images. Images where the associated weight map has 100 or more pixels above this threshold are dictated as images with high overlap regions.

3.2.2 Preprocessing

Patches of size 256×256 pixels were extracted from WSI at a 40x magnification. All images were normalized using the Reinhard stain normalization technique [46] and saved offline. Reinhard stain normalization addresses the variability in staining across different slides, laboratories and imaging setups. By standardizing the color characteristics across the dataset, segmentation models perform better. Reinhard normalization uses a reference image and its statistics in the Lab color space to normalize the pixels of the target image before conversion back to RGB format. Resulting images have enhanced contrast between nuclei stained purple and cytoplasm and extracellular matrix stained pink. The segmentation models have reduced bias towards variations and have an improved generalization. Data augmentation used in this work include vertical and horizontal flips

with a probability of $p = 0.5$, and a Gaussian blur with kernel size of 5×5 and a standard deviation ranging between 0.1 and 3. Flips are introduced as a simple yet effective augmentation for nuclei segmentation. The orientation of nuclei and tissue vary greatly therefore flipping trains the model to recognize nuclei regardless of orientation. Gaussian blur improves the models ability to detect nuclei edges in challenging cases.

3.2.3 Gated CNNs

Previous work done by the IAMLAB introduced an architecture for three class nuclei segmentation called GCNN [47]. The GCNN architecture is a two stream network that integrates a regular stream and an edge stream to address the complexities of nuclei segmentation. The design of the architecture is inspired by the Gated-Shape Convolutional Neural Network (Gated-SCNN) [48], designed for semantic segmentation tasks on the Cityscapes dataset.

In the GCNN, the regular stream functions as any encoder-decoder semantic segmentation network. The edge stream is composed of residual blocks and gated convolutional layers which are connected to intermediate layers of the regular stream. This two stream design enables the model to leverage complementary information from both streams. The edge stream is specifically designed to process shape and boundary related information and operates in parallel with the regular stream. This allows the edge stream to focus on capturing and enhancing boundary details. The gated convolutional layers within this stream allows the model to adaptively emphasize relevant edge features only. In addition, gradients of the images are integrated into the edge stream as well to refine the segmentation process. The base GCNN model with a U-Net encoder decoder backbone is the baseline model used for this work.
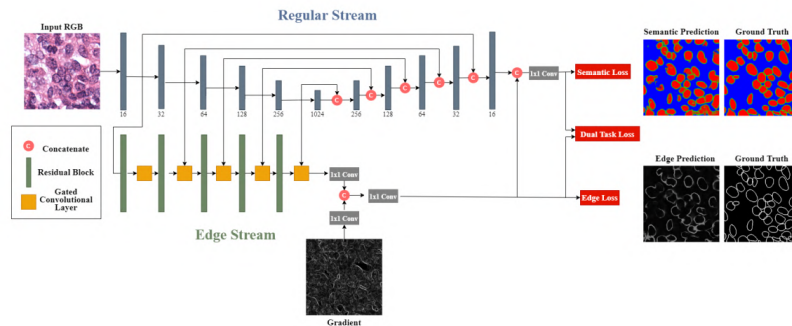

Figure 3.2: Original GCNN Architecture.

3.2.4 GCNN Modifications

Modifications of the GCNN architecture were explored to improve boundary segmentation, particularly in regions with clustered nuclei. The first modification involved replacing the bilinear interpolation blocks in the shape stream with transpose convolution blocks. Contrast to fixed ruled interpolation methods, transpose convolutions incorporate learnable parameters enabling the model to adaptively determine the optimal upsampled mapping. By increasing the number of learnable parameters, particularly in the edge stream, this modification enhances the model's capacity to segment precise boundaries. The second modification to the GCNN architecture is the removal of the fifth and final gated convolution connection between the regular stream and edge stream. Removing this last connection assists the model for boundary segmentation due to the nature of convolutional architectures. The deepest levels prioritize high level, semantic information over low level fine grained details and the deeper features no longer preserve boundary specific details. Removing this connection avoids the transfer of less relevant features. As the deepest level of the network required an upscaling operation by a factor of 16, the number of parameters required for the transpose operation would also greatly increase as shown in Table 3.1.

Table 3.1: GCNN Variations and total parameters of each combination based on number of gated connections and upsample method.

	# of Parameters	
Shape Stream Variation	**4 Gates**	**5 Gates**
Interpolation Operations	16615112	16686075
Transpose Operations	18923528	52751163

The loss function used for all training was described in 3.2. $L_{Semantic}$ is the dice loss from the regular stream output and the three class ground truth. L_{Edge} is the binary cross entropy loss between the shape stream output and the edge class ground truth. $\lambda_1 was selected to be 1 and \lambda_2 was selected to be 5 as these$

$$Loss = \lambda_1 * L_{Semantic} + \lambda_2 * L_{Edge} \tag{3.2}$$

3.2.5 Evaluation Metrics

The evaluation of segmentation performance in this work used both a proposed metric and common evaluation metrics for medical imaging. Given the nature of this three class segmentation task, the

evaluation for the edge and nuclei classes were conducted separately to provide a more detailed analysis. The primary metric used was the Dice Similarity Coefficient (DSC), a commonly used measure for assessing the spatial overlap between the predicted segmentation and corresponding ground truth. The DSC can be computed shown in Equation 3.3:

$$\text{DSC/F1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{3.3}$$

$$\text{Sensitivity/Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.4}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3.5}$$

where TP, FP, and FN represent the true positive, false positive, and false negative pixels, respectively. Sensitivity and specificity metrics are also implemented to evaluate the segmentation performance. Sensitivity, or true positive rate, measures the proportion of correctly identified positive pixels relative to the total number of actual positive pixels while specificity, or true negative rate, measures the proportion of correctly identified negative pixels relative to the total number of actual negative pixels. The equations for sensitivity and specificity are shown in Equation 3.4 and Equation 3.5 respectively.

A new metric was designed to enhance the evaluation of model performance by prioritizing regions within images that contain clustered nuclei. The precomputed weight maps are used to robustly focus on the clustered nuclei areas for any image. In equation 3.6, $w(x)$ represents a sample weight map and the threshold T, is used to isolate the clustered regions creating a new image map, $c(x)$. The threshold value of 0.015 was determined based on a visual analysis of its effect in the project's datasets. Various threshold values were tested to assess their ability to maximize focus on clustered areas and minimize the inclusion of irrelevant regions. The clustered map $c(x)$, is incorporated within the DSC calculation as shown in equation 3.7 for the edge class segmentation and in equation 3.8 for nuclei class segmentation.

$$c(x) = \begin{cases} 1, & \text{if } w(x) \geq T \\ 0, & \text{Otherwise} \end{cases} \tag{3.6}$$

$$\text{Edge15 DSC} = DSC(c(x) * y_1(x), c(x) * \hat{y_1}(x)) \tag{3.7}$$

$$\text{Nuclei15 DSC} = DSC(c(x) * y_2(x), c(x) * \hat{y_2}(x)) \tag{3.8}$$

3.3 Results and Discussion

Predictions were made on the TUPAC clustered dataset and the TCGA dataset. Table 3.2 summarizes the edge segmentation performance of the models. For the TUPAC dataset, the original GCNN achieved an edge DSC of 0.416 slightly outperforming the modified GCNN which achieved an edge DSC of 0.413. In the high clustered areas, the modified GCNN achieved an Edge15 DSC of 0.435 in comparison to the original GCNN score of 0.409. A similar trend is observed for the TCGA dataset where the overall edge DSC of the models has scores, but the Edge15 DSC shows an improvement from 0.410 to 0.433 with the proposed modifications.

Table 3.2: Edge class metrics on the test set of the original and modified GCNN models.

Model	Dataset	Edge DSC	Edge15 DSC	Sensitivity	Specificity
Original GCNN	TUPAC	0.416 (0.050)	0.409 (0.048)	0.417 (0.064)	0.954 (0.012)
Original GCNN	TCGA	0.417 (0.052)	0.410 (0.069)	0.419 (0.066)	0.953 (0.012)
Modified GCNN	TUPAC	0.413 (0.049)	0.435 (0.056)	0.436 (0.061)	0.947 (0.013)
Modified GCNN	TCGA	0.414 (0.048)	0.433 (0.054)	0.436 (0.061)	0.947 (0.013)

Figure 3.3 provides a visual comparison of edge and edge15 DSC metrics across the TUPAC dataset, highlighting the differences in performance between the original and modified GCNN models.
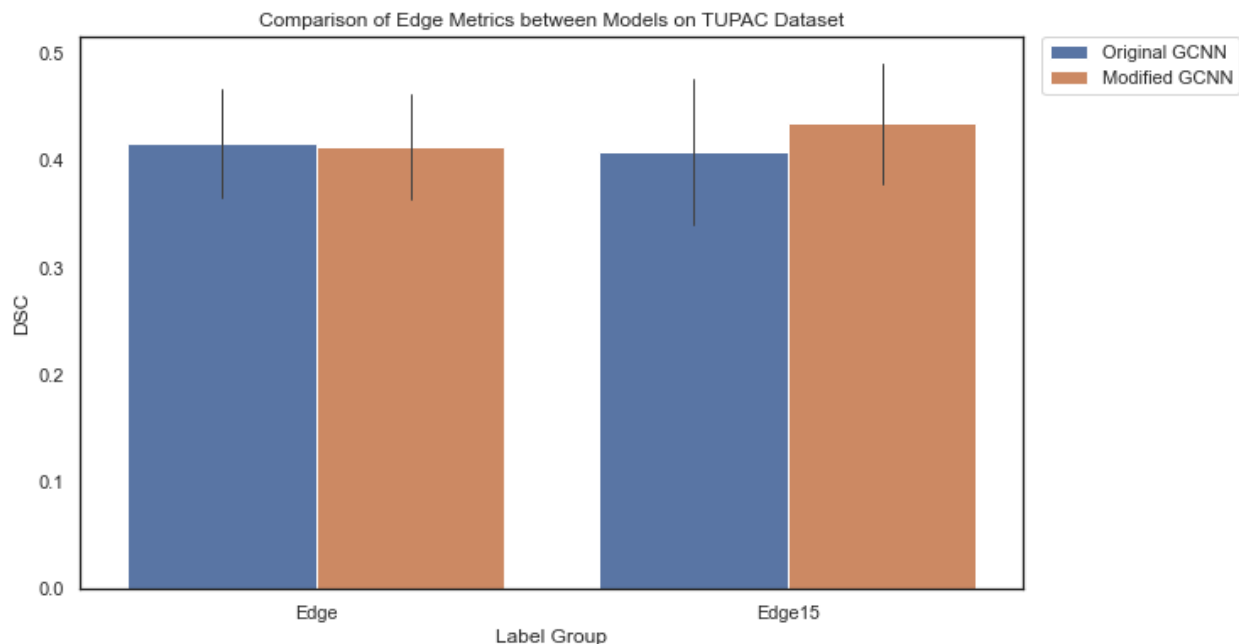
Figure 3.3: Dice score and clustered dice score results for edge class on the TUPAC dataset for Original and Modified GCNN Models.

Table 3.3 presents the results for nuclei segmentation. The nuclei DSC for the original GCNN was 0.810 for the TUPAC dataset and slightly decreased to 0.803 for the modified model. For Nuclei15 DSC, the original GCNN was 0.689 and increased to 0.693 with the modifications. Figure 3.4 displays the small differences between the models for the TUPAC dataset.

Table 3.3: Nuclei class metrics on the test set of the original and modified GCNN models.

Model	Dataset	Nuclei DSC	Nuclei15 DSC	Sensitivity	Specificity
Original GCNN	TUPAC	0.810 (0.052)	0.689 (0.060)	0.783 (0.078)	0.950 (0.022)
Original GCNN	TCGA	0.809 (0.052)	0.692 (0.053)	0.784 (0.078)	0.949 (0.021)
Modified GCNN	TUPAC	0.803 (0.049)	0.693 (0.053)	0.778 (0.073)	0.947 (0.022)
Modified GCNN	TCGA	0.802 (0.052)	0.692 (0.054)	0.777 (0.075)	0.948 (0.023)

Three class nuclei segmentation for breast cancer images is a challenging task where each model did struggle with the complexity of the job. The nuclei in tumor regions exhibit significant variability in shape, size, and texture, making it very difficult for models to generalize well across samples. In addition, many regions within tumor images contain overlapping and clustered nuclei further complicating the establishment of clear boundaries. Having a three class segmentation task introduces a large class imbalance due to the edge class only being a fraction of pixels compared
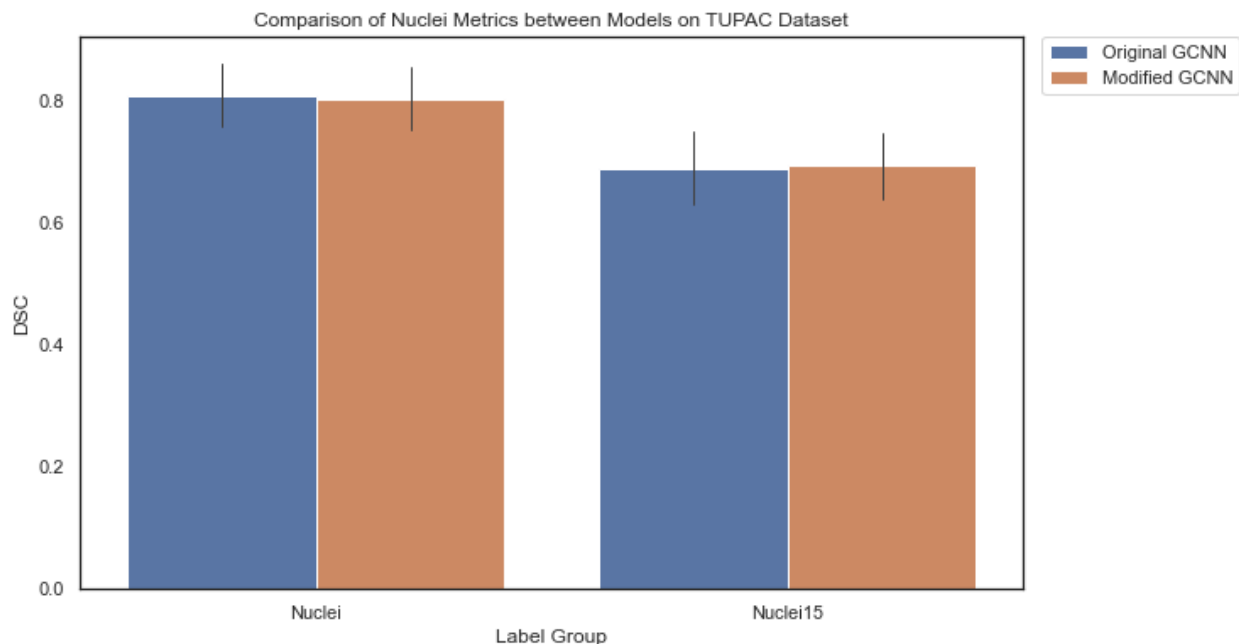
Figure 3.4: Dice score and clustered dice score results for nuclei class on the TUPAC dataset for Original and Modified GCNN Models.

to nuclei. It also brings the need for a balance in performance across all classes, ensuring that improvements in one class does not come at the cost of significant degradation of the other.

Across both datasets and all evaluation metrics, the performance remained relatively equal or showed improvement in both nuclei and edge segmentation results. The consistency of DSC, sensitivity, and specificity across the TUPAC and TCGA datasets suggests that the architectural modifications did not negatively impact overall segmentation quality. Instead, the results indicate that the boundary segmentation and nuclei delineation remained stable, with improvements observed in edge segmentation.

For nuclei segmentation, both the original and modified GCNN architectures achieved high DSC scores, demonstrating that the modifications did not degrade the model's ability to capture nuclear regions. The specificity values remained consistently high, indicating that the models effectively avoided false positives in nuclei detection. Sensitivity values, while slightly lower, highlight the challenge of capturing all nuclei, particularly in densely packed tumor regions where overlapping and irregular morphologies complicate segmentation. For edge segmentation, the Edge15 DSC suggest that the modifications influenced boundary detection, with improvements in delineating nuclei in tough areas. This is crucial for histopathology applications, where precise boundary detection plays a key role in feature extraction for grading and classification. The results confirm
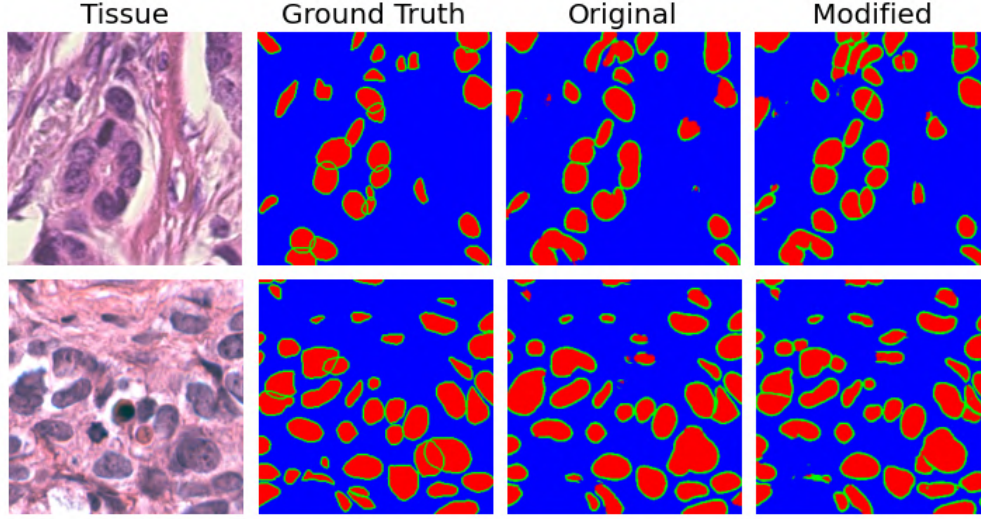
15

Figure 3.5: Sample outputs comparing original and modified GCNN models.

that the modifications did not lead to detrimental trade-offs between nuclei and edge segmentation performance, supporting the robustness of the model across different datasets. Figure 3.5 presents qualitative comparisons between ground truth segmentation masks and the outputs from both the original and modified models for two tissue regions. These examples highlight the differences in segmentation performance with respect to the delineation of nuclear boundaries. The modified GCNN has a visually improved ability to distinguish and preserve edge details, specifically in clustered areas. The modified GCNN also appears to generate smoother nuclear shapes. Neither model is able to delineate overlapping nuclei due to the difficulty of the task.

Extensive efforts have been made to enhance the performance of GCNN with the focus on clustered and overlapping nuclei segmentation. Firstly, the loss functions were targeted for improvement. The weight maps presented were incorporated into the loss function as boundary enchanting weight maps [49]. As this did improve the edge class segmentation, the nuclei class segmentation suffered harshly. For a complex task with multiple outputs, balancing a combination of multiple losses proved to be difficult.

As GCNN is compatible with any encoder decoder model, larger backbones were evaluated for the architecture design. ResUNet++ is an advanced architecture that builds upon U-Net, incorporating squeeze and excitation blocks, Atrous Spatial Pyramid Pooling and attention mechanisms for enhanced feature representation [50] and has been used for breast cancer images [51]. A GCNN with ResUNet++ backbone did not improve the test results, likely due to the data limitations as
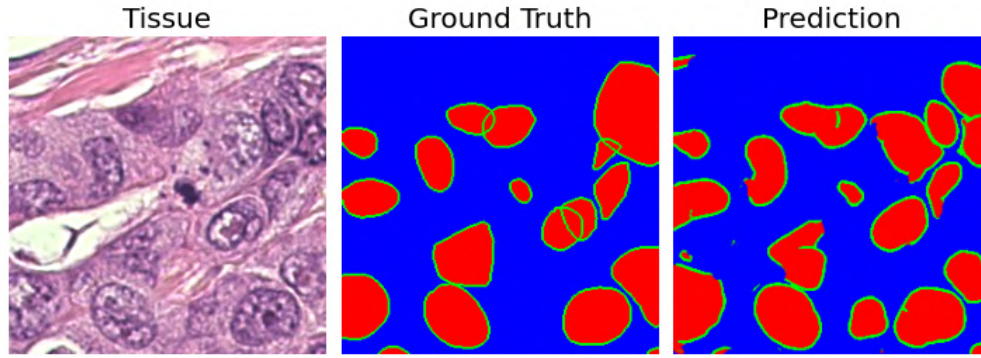
Figure 3.6: Sample testing image and output with difficult to distinguish nuclei.

large models tend to generalize and overfit to smaller datasets.

The data limitations of this study revolve around the amount of data and the annotation quality. Less than 1000 images were available for training and testing but segmentation tasks are dependent on large and diverse datasets. One of the primary observed difficult stems from subjectivity in ground truth annotations. Annotating nuclear masks requires expert pathologists to delineate individual nuclei, but high grade breast cancer tissue often has crowded, deformed or overlapping nuclei making it even more difficult to annotate. Figure 3.6 shows an example patch where nuclei are overlapping and the correct delineation is difficult to determine. In the top right of the image, the ground truth shows 1 nuclei but the model predicts 3 nuclei touching. For segmentation to improve, the subjectivity of annotating needs to be addressed, specifically for the edge class.

We observed that the GCNN shape stream output tends to have weak confidence in it's output map. To improve it's confidence, alternative training methods and architectural changes were explored. After an initial round of training, the main stream was frozen and the shape stream was fine tuned with additional data. The outputs of the shape stream improved in terms of confidence, but had no impact on the overall segmentation result. This led to the exploration of different fusion models for the two streams beside simply concatenating of the feature maps. The GSCNN for the CityScape dataset originally had an ASPP module for fusing the two streams, but when implemented for the task of nuclei segmentation, there was no difference in performance. Nuclei segmentation doesn't need as much global information as the CityScape dataset. Alterative fusion modules were also explored, but also yielded no difference.

3.4 Conclusion

This chapter of the project evaluates the GCNN model and the modifications to improve the model's ability to segment nuclei and their edges. The focus of the modifications revolved around improving segmentation performance in the toughest areas for breast cancer images, The proposed metric provides a robust way to get insight into model performance around clustered and overlapping nuclei. The architecture modifications also were able to slightly improve the edge segmentation results without impacting the overall task. This study also highlighted the benefits and challenges of working with three class masks.

Different avenues may still be explored to improve the work presented. As there was a limited amount of data with 3 class segmentation masks, more images should be carefully annotated. Current annotations should also be reviewed to ensure that nuclei delineation is accurate. The edge class masks should also be standardized to a certain pixel width [28]. Deep learning continues to improve the capabilities of computer vision tasks. State of the art architectures such as ViTs [52] should also be considered for this task.

Chapter 4

Nuclei Grading

4.1 Introdution

The purpose of this chapter was to investigate the impact of segmentation masks from the modified GCNN architecture for the task of automated grading of WSI. Figure 4.1 represents the overall pipeline.

4.2 Material and Methods

4.2.1 Data

For automatic nuclei grading experiments, the OICR dataset was used. In total there are 107 WSI and these slides are scanned at an objective magnification of 40x. A pathologist identified 10 ROIs for each WSI which were used to determine the overall Nottingham grade of the WSI is shown in Figure 4.2.
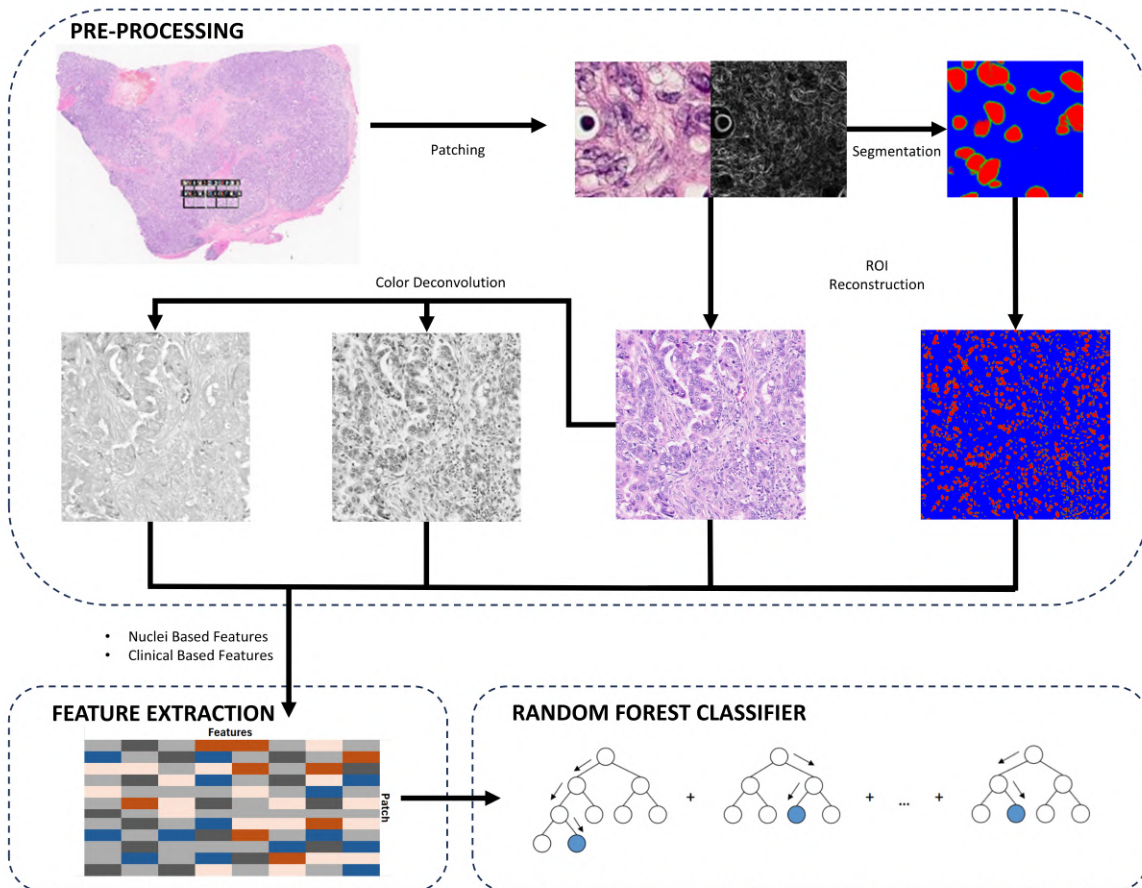
Figure 4.1: Overview of automated nuclear grading pipeline for the OICR dataset.
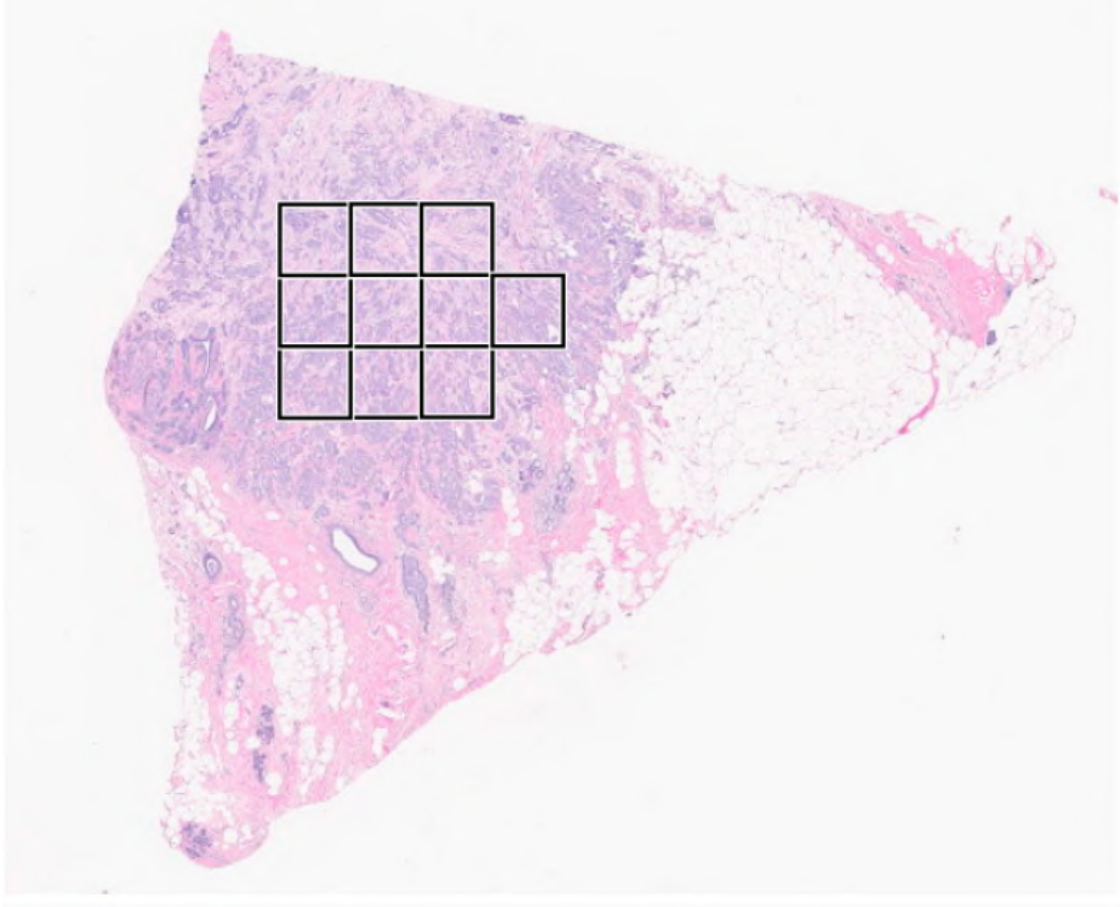
Figure 4.2: Sample WSI from OICR with 10 ROIs identified by a Pathologist.

As the Nottingham system consists of 7 classes of scores, the Nottingham grade is used instead to compact the task to a 3 class classification. A simple conversion is done as follows:

$$
\text{Nottingham Grade} = \begin{cases} 1, & \text{if } 3 \leq \text{NHS} \leq 5 \text{ (Well-Differentiated)} \\ 2, & \text{if } 6 \leq \text{NHS} \leq 7 \text{ (Moderately Differentiated)} \\ 3, & \text{if } 8 \leq \text{NHS} \leq 9 \text{ (Poorly Differentiated)} \end{cases}
$$

Figure 4.3 summarizes the distribution of the WSI using both systems. There are 21 grade 1 cases, 50 grade 2 cases, and 36 grade 3 cases. Associated with each WSI, a patient specific clinical datasheet was also available that contains information on patient's background and history.

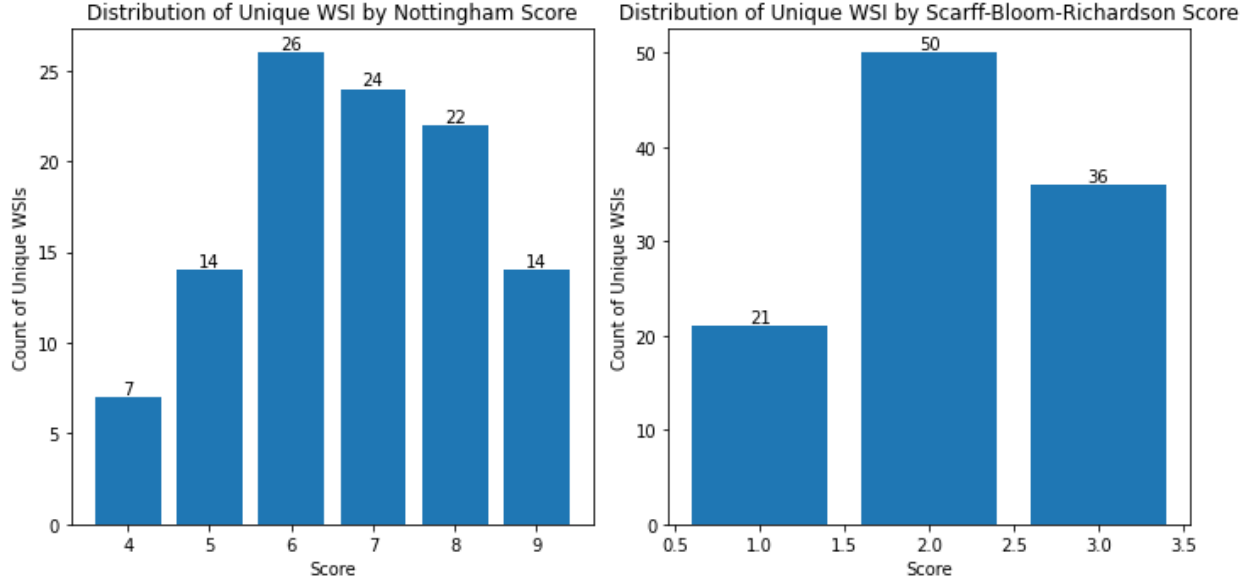Figure 4.3: Distribution of WSIs using Nottingham Scores in the OICR dataset (left) Distribution of WSIs using the Scarff-Bloom-Richardson Score (right)

4.2.2 Patching and Segmentation

The first step in this process involved patching down the WSI images. Patches of 256×256 were extracted from each ROI of the WSI. Segmenting nuclei at the border of patches presents a significant challenge as nuclei may be cut off and are misrepresented. This issue introduces artifacts in the segmentation mask and compromises the integrity of extracted features, particularly for metrics that depend on accurate size and shape of nuclei. Such errors are especially problematic for the task of automated grading as nuclei characteristics play a critical role in the prediction. To address this challenge, overlapping patches is employed. The patches are generated with a 128 pixel overlap between adjacent sections in both the x and y direction. This ensures that nuclei at the edges of one patch also appear within the boundaries of neighboring patches. This redundancy ensures that all nuclei are captured in at least one patch.

4.2.3 Segmentation and Reconstruction

Each patch is color normalized using the Reinhard stain normalization similar to the data used to train the modified GCNN. Patches are then passed through the model to generate segmentation masks. The model processes each patch predicting pixel wise labels for nuclei, edges and background. To reconstruct the original ROIs from the segmented patches, a weighted average

22

method is applied to resolve overlapping areas. In this process, each class is assigned a weight per class type and the weighted averaging is performed over the pixel-level probabilities. The weights assigned for the nuclei and edge class was 2, while the background class was 1. These weights were selected based on qualitative assessment of different weights and tend to preserves the most nuclei at bordering patches resulting in robust reconstruction of the ROIs.

4.2.4 Feature Extraction

For feature extraction, each nuclei must be identified and separated. As the segmentation masks consist of 3 classes, the actual feature extraction focuses exclusively on the nuclei class. The edge class serves to assist in determining the accurate segmentation of the nuclei. Detected nuclei with less than 50 pixels were eliminated in the generated masks as nuclei cannot include less than 50 pixels at 40x magnification [53]. The edge pixels are then removed from the masks using an XOR operation to refine the nuclei areas. The masks are then prepared for feature extraction. Features are computed per nuclei and then are averaged across the ROI.

Color deconvolution is an essential preprocessing step employed for image based feature extraction. This process separates the composite color representation of the image into stain specific channels of hematoxylin, eosin, and background, which helps isolate the structures. Color deconvolution was performed using a supervised method provided by the HistomicskTK package [54]. In this method, the tissue stain map describes the optical density (OD) vectors for the stains used and unmixes the RGB intensities in the OD space. The resulting stains are then represented as grayscale images as shown in Figure 4.4.

The HistomickTK and Pyradiomics [55] open source packages were used to extract 136 features from each ROI. The feature families extracted using the segmentation masks are morphological features, intensity features, gradient features, haralick features, and Gray-Level Co-occurrence Matrix (GLCM) features. Morpholgical features describe the size, shape, and spatial arrangement, of nuclei. Intensity features quantify the pixel intensity distribution within nuclei providing information on texture and density. Gradient features measure the intensity transitions within and around nuclei reflecting the sharpness of nuclear boundaries. GLCM features quantify the frequency of specific pixel intensity pairs within the nuclei. Haralick features are a set of textural descriptors derived from GLCM and quantify the spatial relationship between pixel intensities.
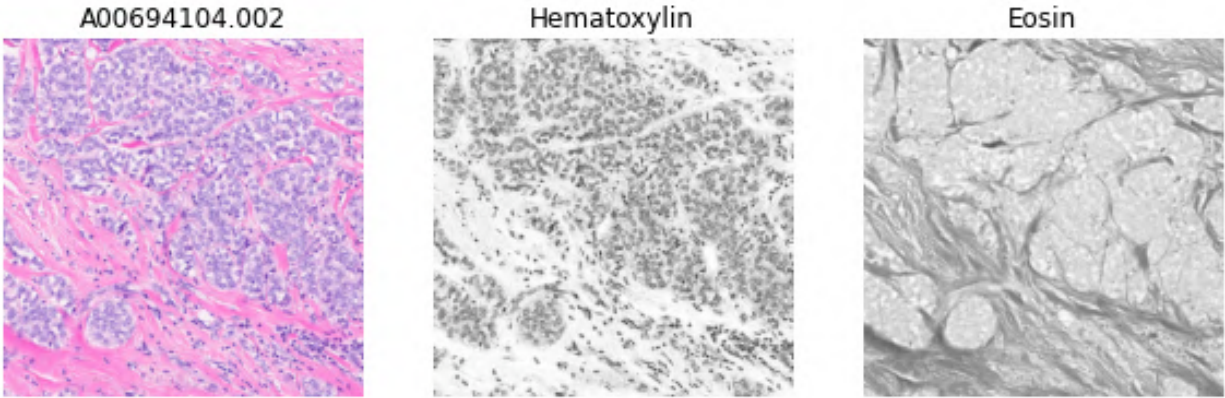
Figure 4.4: Color deconvolution performed on sample WSI ROI from the OICR Dataset. Original Image (Left), Hematoxylin Stain (Center), Eosin Stain (Right)

The OICR dataset also had accompanying clinical data sheets with patient specific information. Clinical features, when sufficiently populated with data across the patients, were selected to be included in the feature list.

4.2.5 Machine Learning

The machine learning algorithm selected for nuclear grading was the Random Forest (RF) classifier algorithm [56]. RF is commonly used for classification and regression tasks leveraging the concept of bootstrap aggregating also known as bagging. Bagging aims to improve the model performance by reducing variance and improving generalization. Each decision tree in the forest is trained on a random subset of the data and each split uses a random subset of features in it's decision. RF also is able to produce feature importance values providing insight into the grading prediction. Due to limitations in data size and imbalance in the OICR dataset, a repeated stratified train-test split was used for 20 iterations. This method better utilizes the available data and reduces the risk of bias. The data split is shown in Figure 4.5. Metrics are first computed by ROI. ROI derived from the same WSI are then aggregated for the final results.

4.2.6 Evaluation Metrics

To measure classification performance, the accuracy, precision, recall, and F1-Score, are calculated. Metrics are computed per patient by aggregating ROI predictions per patient. F1 score and recall
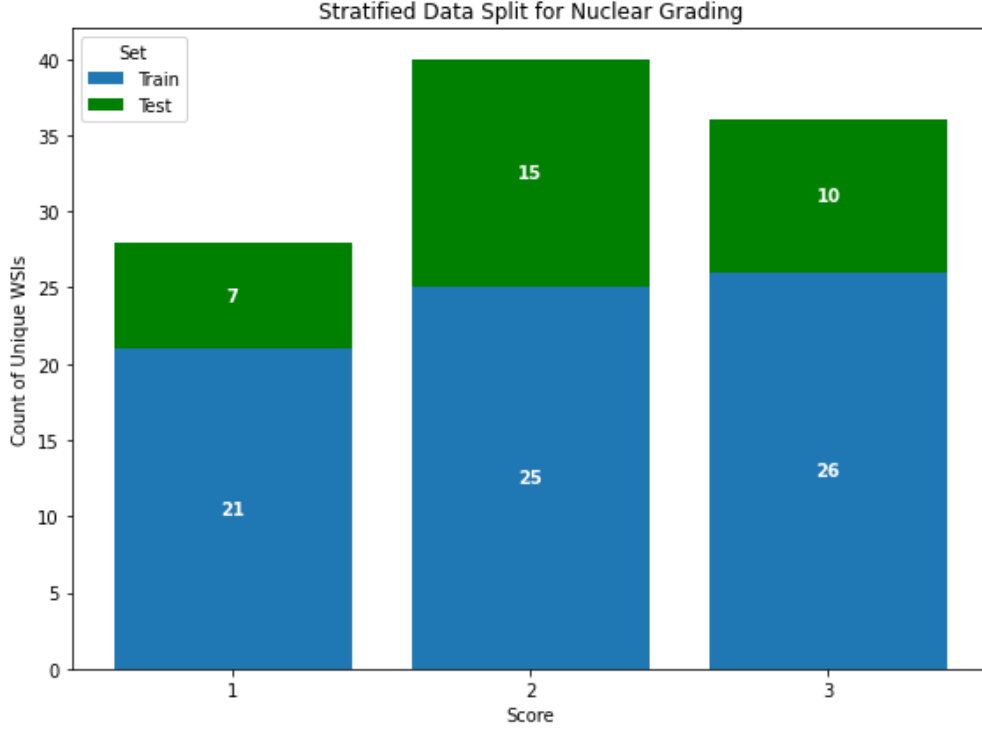
Figure 4.5: Stratified Data Split of WSI per Nuclear Grade.

are the same as the DSC and the sensitivity metrics used for the segmentation task.

$$\text{DSC/F1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{4.1}$$

$$\text{Sensitivity/Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.2}$$

Precision is the ratio of correctly predicted positive samples to all predicted positive samples as described in Equation 4.3. Accuracy is computed as the ratio of all samples correctly predicted for positive and negative cases to all predictions as described in Equation 4.4.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.3}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4.4}$$

4.3   Results and Discussion

A total of 7 experiments were conducted to predict nuclear grade using individual feature families and a combination of all of them. Table 4.1 reports the evaluation metrics along with their standard deviations calculated from 20 stratified splits of the dataset. Among individual feature families, intensity features achieve the highest performance with an accuracy of 0.596 and a F1 score of 0.588. Morphological features followed closely with an accuracy of 0.589 and an F1 score of 0.573. Gradient and clinical features demonstrated moderate performance, while GLCM and Haralick features exhibit the lowest results among individual families. The combined feature set achieved the best overall performance with an accuracy of 0.706 and an F1 score of 0.699. This demonstrates that integrating multiple feature families leads to significantly better performance than using individual feature types alone.

Table 4.1: Result of nuclear grade prediction using a random forest classifier. Results shown are aggravated for 20 stratified splits of the data.

Feaures	# of Features	Accuracy	Precison	Recall	F1-Score
Clinical Features	12	0.570 (0.078)	0.575 (0.079)	0.570 (0.076)	0.570 (0.076)
Morphological Features	32	0.589 (0.073)	0.611 (0.085)	0.589 (0.073)	0.573 (0.071)
Intensity Features	24	0.596 (0.086)	0.610 (0.092)	0.597 (0.086)	0.588 (0.089)
Gradient Features	16	0.573 (0.061)	0.575 (0.068)	0.573 (0.061)	0.561 (0.063)
Haralick Features	26	0.523 (0.085)	0.522 (0.089)	0.523 (0.084)	0.513 (0.084)
GLCM Features	24	0.519 (0.083)	0.518 (0.093)	0.519 (0.083)	0.506 (0.087)
All Features	134	**0.706 (0.081)**	0.724 (0.091)	0.706 (0.081)	0.699 (0.087)

Intensity based and morphological based features showed relatively strong performance when used in isolation. Previous studies have highlighted that intensity features correlate strongly with nuclear pleomorphism making them highly relevant for breast cancer grading [3]. Morphological features are directly associated with nuclear pleomorphism, as they quantify the size and perimeter of the nuclei shape.

Figure 4.6 and Figure 4.7 are the averaged and totaled confusion matrix for this experiment, respectively, with all features used. The labels correspond to the SBR score. Across all iterations, the model maintained consistent performance. For grade 1 cases, the results show moderate accuracy. There are a high number of misclassifications predicting grade 2 as there was an average of 30 ROI

misclassified per iteration. This is most likely the result of overlapping features between a low and intermediate grade of nuclei. The model performed best in identifying grade 2 nuclei. For grade 3, the model performed well with most of it's misclassifcations happening with grade 2 predictions.
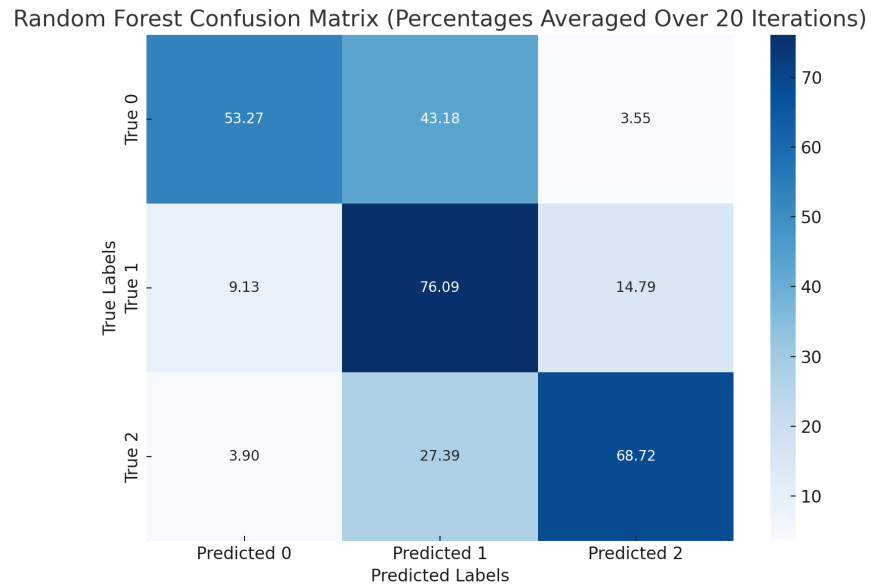

Figure 4.6: Confusion matrix averaged for 20 iterations of the Random Forest algorithm.
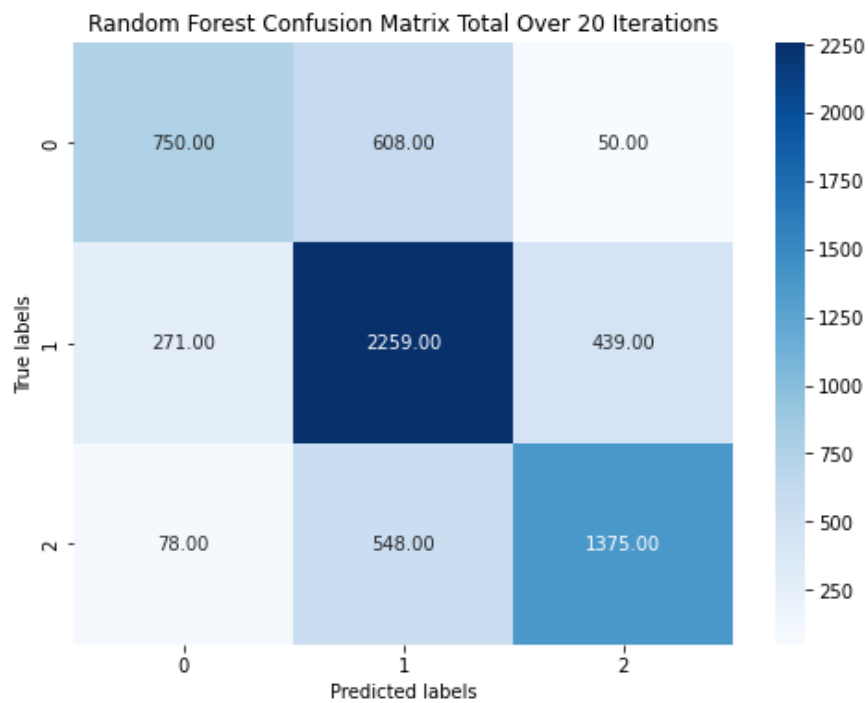

Figure 4.7: Confusion matrix summed for 20 iterations of the Random Forest algorithm.

Figure 4.8 displays the top ten features and their corresponding importance score. Each color demonstrates which family the feature is derived from. For this experiment, half of the top features are clinical features with the top 3 being tumor size, HER2 status, and ER status. Two intensity features are also in this list, the minimum intensity of nuclei, and the energy of the intensity histogram of object pixels for cytoplasm. The Fourier shape descriptor with frequency set to 6 is the only shape feature. The last of the top ten features was the sum of canny filtered gradient data for nucleus.
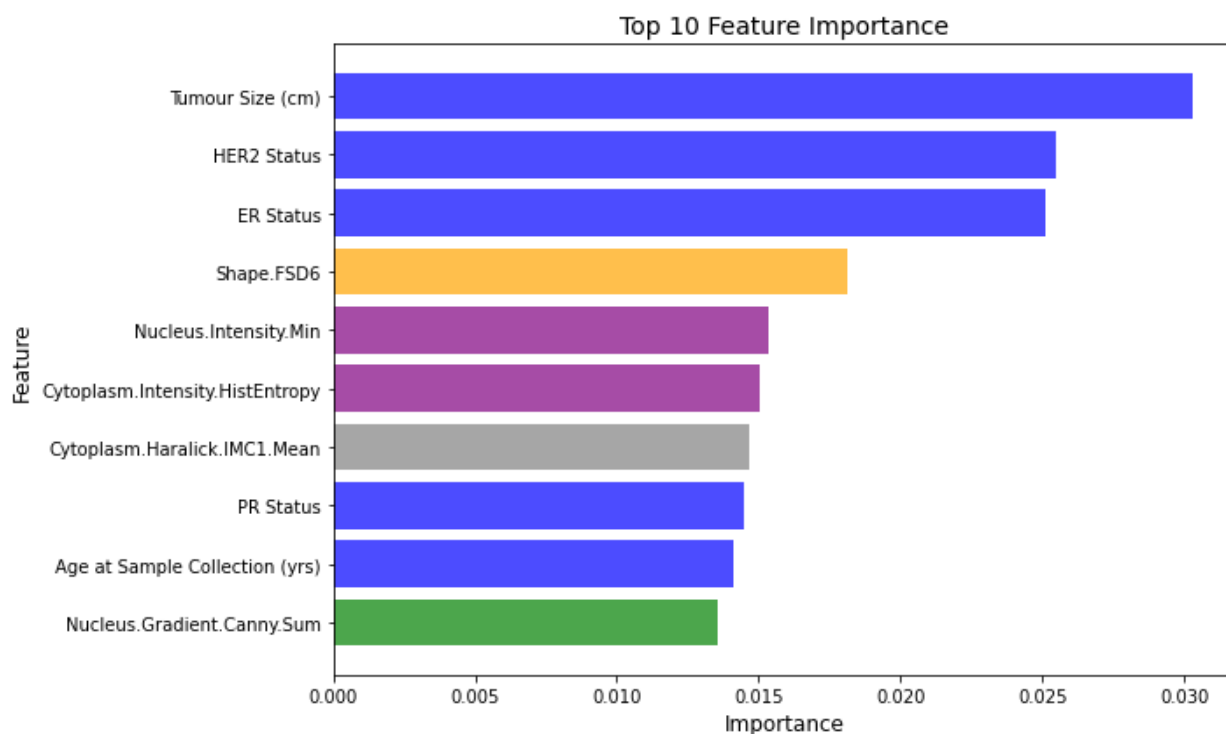
Figure 4.8: Top 10 features with the highest feature importance scores

The performance of the random forest classifier presented in this project in directly influenced by the quality of the extracted features. Inaccuracies in segmentation can propagate through the grading pipeline. As the GCNN model is not capable of robustly segmenting all nuclei for breast cancer, morphological features such as area or shape descriptors may be skewed and incorrect. The texture and intensity features may also be susceptible to errors from segmentation inaccuracies. Misidentified boundaries can lead to inclusion of non nuclei regions or exclude these areas, altering the pixel intensity distributions and texture patterns. The OICR dataset also does not have any manual segmentation results which can be used as labels to evaluate GCNN performance. For this

reason, the results of these experiments only provide a broad view of how effective this pipeline was.

A main challenge of the 3 class segmentation masks for feature extraction is separating nuclei after segmentation. Breast cancer datasets experience many cases of clustered or touching nuclei making it difficult to automate a way to accurately separate nuclei based on their edge class segmentation. Given the predictable inaccuracies in the segmentation results, the XOR method was used.

To further evaluate the impact of having 3 class segmentation masks for nuclear grading, the random forest classifier was to be compared to a deep learning method and an ensemble of machine learning and deep learning. For this task, a pretrained VGG-19 [43] and a custom DenseNet [57] were employed. These models were trained on both the smaller ROI dataset, and a dataset of patches from the ROIs. Neither model or method were able to train, due to the lack of sufficient data.

4.4   Conclusion

The objective of this chapter was to evaluate the effectiveness of three class segmentation masks in automating breast cancer grading using 10 identified ROIs extracted from WSIs. This approach aimed to assess if the segmentation masks could provide a robust foundation for nuclear feature extraction and AI based grading. The random forest classifier was employed to predict nuclear grades based on extracted features. The results demonstrated that while the combination of clinical and segmentation based features provided useful predictive signals, the model's performance is impacted by segmentation accuracy and class imbalances in the dataset. Future work should include obtaining more data to address the class imbalance and to have enough data for a deep learning approach to compare to. Oversampling methods should also be explored to further improve classification performance. A widely adopted approach is the synthetic minority oversampling technique (SMOTE) which generates new synthetic samples by interpolating between existing samples [58].

Bibliography

[1] D. R. Brenner, J. Gillis, A. A. Demers, L. F. Ellison, J.-M. Billette, S. X. Zhang, J. L. Liu, R. R. Woods, C. Finley, N. Fitzgerald *et al.*, "Projected estimates of cancer in canada in 2024," *Cmaj*, vol. 196, no. 18, pp. E615–E623, 2024.

[2] A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, "Hematoxylin and eosin staining of tissue and cell sections," *Cold spring harbor protocols*, vol. 2008, no. 5, pp. pdb–prot4986, 2008.

[3] C. W. Elston and I. O. Ellis, "Pathological prognostic factors in breast cancer. i. the value of histological grade in breast cancer: experience from a large study with long-term follow-up," *Histopathology*, vol. 19, no. 5, pp. 403–410, 1991.

[4] M. Veta, P. J. Van Diest, S. M. Willems, H. Wang, A. Madabhushi, A. Cruz-Roa, F. Gonzalez, A. B. Larsen, J. S. Vestergaard, A. B. Dahl *et al.*, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Medical image analysis*, vol. 20, no. 1, pp. 237–248, 2015.

[5] M. van Seijen, K. Jóźwiak, S. E. Pinder, A. Hall, S. Krishnamurthy, J. S. Thomas, L. C. Collins, J. Bijron, J. Bart, D. Cohen *et al.*, "Variability in grading of ductal carcinoma in situ among an international group of pathologists," *The Journal of Pathology: Clinical Research*, vol. 7, no. 3, pp. 233–242, 2021.

[6] C. Messer, R. Khurana, and S. Bedi, "Classification of uncertainty in nuclei segmentation from h&e images."

[7] J. S. Meyer, C. Alvarez, C. Milikowski, N. Olson, I. Russo, J. Russo, A. Glass, B. A. Zehnbauer, K. Lister, R. Parwaresch *et al.*, "Breast carcinoma malignancy grading by bloom–richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index," *Modern pathology*, vol. 18, no. 8, pp. 1067–1078, 2005.

[8] P. Maqlin, R. Thamburaj, J. J. Mammen, and M. T. Manipadam, "Automated nuclear pleomorphism scoring in breast cancer histopathology images using deep neural networks," in *Mining Intelligence and Knowledge Exploration: Third International Conference, MIKE 2015, Hyderabad, India, December 9-11, 2015, Proceedings 3*. Springer, 2015, pp. 269–276.

[9] E. Rakha, D. Soria, A. R. Green, C. Lemetre, D. G. Powe, C. C. Nolan, J. M. Garibaldi, G. Ball, and I. O. Ellis, "Nottingham prognostic index plus (npi+): a modern clinical decision making tool in breast cancer," *British journal of cancer*, vol. 110, no. 7, pp. 1688–1697, 2014.

[10] P. J. Van Diest, E. van der Wall, and J. P. Baak, "Prognostic value of proliferation in invasive breast cancer: a review," *Journal of clinical pathology*, vol. 57, no. 7, pp. 675–681, 2004.

[11] L. Pantanowitz, P. N. Valenstein, A. J. Evans, K. J. Kaplan, J. D. Pfeifer, D. C. Wilbur, L. C. Collins, and T. J. Colgan, "Review of the current state of whole slide imaging in pathology," *Journal of pathology informatics*, vol. 2, no. 1, p. 36, 2011.

[12] M. Veta, Y. J. Heng, N. Stathonikos, B. E. Bejnordi, F. Beca, T. Wollmann, K. Rohr, M. A. Shah, D. Wang, M. Rousson *et al.*, "Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge," *Medical image analysis*, vol. 54, pp. 111–121, 2019.

[13] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.

[14] P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman *et al.*, "Qupath: Open source software for digital pathology image analysis," *Scientific reports*, vol. 7, no. 1, pp. 1–7, 2017.

[15] N. Malpica, C. O. De Solórzano, J. J. Vaquero, A. Santos, I. Vallcorba, J. M. García-Sagredo, and F. Del Pozo, "Applying watershed algorithms to the segmentation of clustered nuclei," *Cytometry: The Journal of the International Society for Analytical Cytology*, vol. 28, no. 4, pp. 289–297, 1997.

[16] N. Otsu *et al.*, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[17] N. Phansalkar, S. More, A. Sabale, and M. Joshi, "Adaptive local thresholding for detection of nuclei in diversity stained cytology images," in *2011 International conference on communications and signal processing*. IEEE, 2011, pp. 218–220.

[18] M. Abdolhoseini, M. G. Kluge, F. R. Walker, and S. J. Johnson, "Segmentation of heavily clustered nuclei from histopathological images," *Scientific reports*, vol. 9, no. 1, p. 4551, 2019.

[19] M. Veta, P. J. Van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. Pluim, "Automatic nuclei segmentation in h&e stained breast cancer histopathology images," *PloS one*, vol. 8, no. 7, p. e70221, 2013.

[20] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 550–566, 2015.

[21] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, "Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting," *IEEE transactions on medical imaging*, vol. 30, no. 9, pp. 1661–1677, 2011.

[22] S. Kothari, J. H. Phan, R. A. Moffitt, T. H. Stokes, S. E. Hassberger, Q. Chaudry, A. N. Young, and M. D. Wang, "Automatic batch-invariant color segmentation of histological cancer images," in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2011, pp. 657–660.

[23] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen, "Deep learning for cellular image analysis," *Nature methods*, vol. 16, no. 12, pp. 1233–1246, 2019.

[24] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into imaging*, vol. 9, pp. 611–629, 2018.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *CoRR*, vol. abs/1505.04597, 2015, arXiv: 1505.04597. [Online]. Available: http://arxiv.org/abs/1505.04597

[26] R. Pandey, R. Lalchhanhima, and K. R. Singh, "Nuclei cell semantic segmentation using deep learning unet," in *2020 Advanced Communication Technologies and Signal Processing (ACTS)*. IEEE, 2020, pp. 1–6.

[27] Y. Cui, G. Zhang, Z. Liu, Z. Xiong, and J. Hu, "A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images," *Medical & biological engineering & computing*, vol. 57, pp. 2027–2043, 2019.

[28] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 448–459, 2018.

[29] I. Kiran, B. Raza, A. Ijaz, and M. A. Khan, "Denseres-unet: Segmentation of overlapped/clustered nuclei from multi organ histopathology images," *Computers in Biology and Medicine*, vol. 143, p. 105267, 2022.

[30] T. Wan, L. Zhao, H. Feng, D. Li, C. Tong, and Z. Qin, "Robust nuclei segmentation in histopathology using asppu-net and boundary refinement," *Neurocomputing*, vol. 408, pp. 144–156, 2020.

[31] B. Zhao, X. Chen, Z. Li, Z. Yu, S. Yao, L. Yan, Y. Wang, Z. Liu, C. Liang, and C. Han, "Triple u-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation," *Medical Image Analysis*, vol. 65, p. 101786, 2020.

[32] H. U. Khan, B. Raza, M. A. I. Khan, and M. Faheem, "Waveseg-unet model for overlapped nuclei segmentation from multi-organ histopathology images," *CAAI Transactions on Intelligence Technology*, 2024.

[33] O. Falou, A. Sadeghi-Naini, S. Prematilake, E. Sofroni, N. Papanicolau, S. Iradji, Z. Jahedmotlagh, S. Lemon-Wong, J.-P. Pignol, E. Rakovitch *et al.*, "Evaluation of neoadjuvant chemotherapy response in women with locally advanced breast cancer using ultrasound elastography," *Translational oncology*, vol. 6, no. 1, pp. 17–24, 2013.

[34] A. Sadeghi-Naini, N. Papanicolau, O. Falou, J. Zubovits, R. Dent, S. Verma, M. Trudeau, J. F. Boileau, J. Spayne, S. Iradji *et al.*, "Quantitative ultrasound evaluation of tumor cell death response in locally advanced breast cancer patients receiving chemotherapy," *Clinical Cancer Research*, vol. 19, no. 8, pp. 2163–2174, 2013.

[35] B. dos Anjos Pultz, F. A. C. da Luz, P. R. de Faria, A. P. L. Oliveira, R. A. de Araújo, and M. J. B. Silva, "Far beyond the usual biomarkers in breast cancer: a review," *Journal of Cancer*, vol. 5, no. 7, p. 559, 2014.

[36] F. Aeffner, M. D. Zarella, N. Buchbinder, M. M. Bui, M. R. Goodman, D. J. Hartman, G. M. Lujan, M. A. Molani, A. V. Parwani, K. Lillard *et al.*, "Introduction to digital image analysis in whole-slide imaging: a white paper from the digital pathology association," *Journal of pathology informatics*, vol. 10, no. 1, p. 9, 2019.

[37] L. He, L. R. Long, S. Antani, and G. R. Thoma, "Histology image analysis for carcinoma detection and grading," *Computer methods and programs in biomedicine*, vol. 107, no. 3, pp. 538–556, 2012.

[38] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in *2008 5th IEEE international symposium on biomedical imaging: from nano to macro*. IEEE, 2008, pp. 496–499.

[39] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren, "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer," *BMC medical imaging*, vol. 6, pp. 1–11, 2006.

[40] P. Vaidya, X. Wang, K. Bera, A. Khunger, H. Choi, P. Patil, V. Velcheti, and A. Madabhushi, "Raptomics: integrating radiomic and pathomic features for predicting recurrence in early stage lung cancer," in *Medical Imaging 2018: Digital Pathology*, vol. 10581. SPIE, 2018, pp. 172–182.

[41] H. Bolhasani, E. Amjadi, M. Tabatabaeian, and S. J. Jassbi, "A histopathological image dataset for grading breast invasive ductal carcinomas," *Informatics in Medicine Unlocked*, vol. 19, p. 100341, 2020.

[42] P. H. Zavareh, A. Safayari, and H. Bolhasani, "Bcnet: a deep convolutional neural network for breast cancer grading," *arXiv preprint arXiv:2107.05037*, 2021.

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[44] V. Mudeng, M. N. Farid, G. Ayana, and S.-w. Choe, "Domain and histopathology adaptations–based classification for malignancy grading system," *The American Journal of Pathology*, vol. 193, no. 12, pp. 2080–2098, 2023.

[45] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *Ieee transactions on biomedical engineering*, vol. 63, no. 7, pp. 1455–1462, 2015.

[46] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001.

[47] S. Beniamin, A. Khademi, and D. Androutsos, "Gated CNNs for Nuclei Segmentation in H&E Breast Images," in *Medical Imaging with Deep Learning*, 2021. [Online]. Available: https://openreview.net/forum?id=fQDGt0RJkMu

[48] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-scnn: Gated shape cnns for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 5229–5238.

[49] R. Wang and S.-i. Kamata, "Stain-refinement and boundary-enhancement weight maps for multi-organ nuclei segmentation," in *2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. IEEE, 2020, pp. 1–7.

[50] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, and H. D. Johansen, "Resunet++: An advanced architecture for medical image segmentation," in *2019 IEEE international symposium on multimedia (ISM)*. IEEE, 2019, pp. 225–2255.

[51] Y. Chen, Y. Zhou, G. Chen, Y. Guo, Y. Lv, M. Ma, Z. Pei, and Z. Sun, "Segmentation of breast tubules in h&e images based on a dks-doubleu-net model," *BioMed Research International*, vol. 2022, no. 1, p. 2961610, 2022.

[52] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[53] K. Saednia, A. Lagree, M. A. Alera, L. Fleshner, A. Shiner, E. Law, B. Law, D. W. Dodington, F.-I. Lu, W. T. Tran *et al.*, "Quantitative digital histopathology and machine learning to predict pathological complete response to chemotherapy in breast cancer patients using pre-treatment tumor biopsies," *Scientific Reports*, vol. 12, no. 1, p. 9690, 2022.

[54] D. A. Gutman, M. Khalilia, S. Lee, M. Nalisnik, Z. Mullen, J. Beezley, D. R. Chittajallu, D. Manthey, and L. A. Cooper, "The digital slide archive: a software platform for management, integration, and analysis of histology for cancer research," *Cancer research*, vol. 77, no. 21, pp. e75–e78, 2017.

[55] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.

[56] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[57] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[58] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.