Marquette University
e-Publications@Marquette

Master's Theses (2009 -)

Dissertations, Theses, and Professional Projects

Deep Learning Classification of Deep Ultraviolet Fluorescence Images for Margin Assessment During Breast Cancer Surgery

Tyrell To
Marquette University

Follow this and additional works at: https://epublications.marquette.edu/theses_open

 Part of the Engineering Commons

Recommended Citation

To, Tyrell, "Deep Learning Classification of Deep Ultraviolet Fluorescence Images for Margin Assessment During Breast Cancer Surgery" (2023). *Master's Theses (2009 -)*. 768.
https://epublications.marquette.edu/theses_open/768

DEEP LEARNING CLASSIFICATION OF DEEP ULTRAVIOLET FLUORESCENCE IMAGES FOR MARGIN ASSESSMENT DURING BREAST CANCER SURGERY

by

Tyrell To

A Thesis submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin

August 2023

ABSTRACT
DEEP LEARNING CLASSIFICATION OF DEEP ULTRAVIOLET FLUORESCENCE IMAGES FOR MARGIN ASSESSMENT DURING BREAST CANCER SURGERY

Tyrell To, M.S.

Marquette University 2023

Breast-conserving surgery (BCS) is a widely used treatment for breast cancer, but ensuring the complete removal of cancer cells from the surgical margins remains a challenge. Deep ultraviolet (DUV) fluorescence scanning microscopy offers a potential solution by providing real-time whole-surface imaging of resected tissues during BCS. However, interpreting DUV images for margin assessment requires an automated classification method. This dissertation addresses this need by proposing a deep learning-based classification approach for DUV fluorescence images in intra-operative margin assessment of breast cancer.

To overcome the limited availability of DUV image datasets and potential overfitting, the study combines patch-level classification using transfer learning with regional importance maps generated through the Grad-CAM++ algorithm. The proposed methodology involves dividing DUV whole-slide images into smaller patches, converting them to grayscale, and analyzing pixel values to identify valid patches. A pre-trained ResNet50 network and an XGBoost classifier are utilized for patch-level classification, while Grad-CAM++ generates regional importance maps for the entire DUV image.

The decision fusion method combines patch-level classification labels and regional importance maps to determine the whole-slide image (WSI)-level classification label by calculating the total number of malignant patches and comparing it to a threshold percentage of total foreground patches. A binary classification is obtained for the entire WSI.

The proposed methodology is implemented using PyTorch and a dataset consisting of 60 DUV images of breast tissue samples. The DUV images were obtained using a custom DUV-Fluorescence Scanning Microscopy system, which provided high-resolution images with fluorescence staining for accurate tissue classification.

The results of this study contribute to the field of intra-operative margin assessment in breast cancer by demonstrating the effectiveness of deep learning-based classification of DUV images. The combination of transfer learning, regional importance maps, and decision fusion provides a robust approach for accurately classifying breast tissue as malignant or normal/benign. This research opens new avenues for utilizing deep learning techniques in DUV fluorescence imaging and has the potential to improve surgical outcomes in breast-conserving surgery.

ACKNOWLEDGMENTS

Tyrell To

I thank Dr. Dong Hye Ye and Dr. Richard Povinelli for their invaluable guidance and support throughout my journey. Their critical nature and rigorous approach have constantly challenged me to strive for excellence and have undoubtedly shaped me into a better student and researcher. I am genuinely grateful for their mentorship and the opportunities they have provided me.

I am also indebted to David Helmaniak for his unwavering assistance with hardware and software issues. His technical expertise and willingness to help have been instrumental in overcoming various challenges, ensuring the smooth progress of my research. I am grateful for his timely and practical support.

Furthermore, I would like to sincerely thank Dr. Bing Yu for his invaluable contributions to breast cancer research. His insightful feedback and constructive criticism have significantly improved the quality of my writing in my publications. I am truly fortunate to have had the opportunity to collaborate with Dr. Yu.

In addition, I would like to thank Dr. James Richie and Dr. Frederick Frigo for their guidance and advice during my undergraduate studies. Their expertise and wisdom have been invaluable in shaping my academic foundation and instilling in me a passion for research. I am thankful for their mentorship and the knowledge they imparted to me.

Lastly, I express my deepest gratitude to my parents for their unwavering support and love throughout this journey. Their constant encouragement, belief in my abilities, and sacrifices have driven my accomplishments. I am forever grateful for their unconditional love and unwavering belief in my potential.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Motivation

Breast cancer remains a significant health concern, affecting approximately 1 in 8 women in the United States during their lifetime [1]. It is the most common cancer among women worldwide, accounting for approximately 30% of all cancer cases reported among women [2]. The disease's incidence and mortality rates vary across different demographic groups, highlighting the complex interplay of genetic, environmental, and lifestyle factors in its etiology. Despite the advancements in breast cancer detection and treatment, significant challenges persist, particularly in surgical intervention.

Breast-conserving surgery (BCS), also known as lumpectomy, has emerged as a critical tool in the fight against breast cancer. More than half of women diagnosed with breast cancer undergo this procedure [3, 4]. This surgical approach aims to excise the tumor while preserving as much healthy breast tissue as possible, offering patients a less invasive alternative to mastectomy and improving post-operative quality of life. However, the success of BCS is heavily contingent upon the complete removal of the tumor, including any cancer cells that may reside at the edges of the excised tissue, known as the margin. If these cells are not entirely detected and removed, patients may experience cancer recurrence and require additional surgery [5].

Incomplete excision of the tumor margin is a significant issue in BCS, leading to a higher risk of local recurrence and the need for additional surgery. Current methods for intraoperative margin assessment have limitations, including time consumption, resource intensity, and variable accuracy. Therefore, there is a pressing need for improved tech-

niques for intraoperative margin assessment to ensure the complete removal of the tumor and reduce the need for reoperation. The intra-operative margin evaluation is a crucial step in the surgical process. It is typically performed through radiographic examination, confirmed by post-operative histological analysis of hematoxylin and eosin (H&E) images. However, this process can be time-consuming and may not provide real-time feedback during surgery, which is crucial for making immediate surgical decisions.

Developing and validating such techniques could significantly improve BCS outcomes, reducing the local recurrence rate and the associated physical, emotional, and financial burdens on patients. Furthermore, advancements in this area could contribute to the broader goal of personalized breast cancer treatment, tailoring the surgical intervention to the individual characteristics of the tumor and the patient.

An approach to real-time margin assessment involves the use of deep ultraviolet fluorescence scanning microscopy (DUV-FSM) for real-time whole-surface imaging (Whole Surface Images) of resected tissues during BCS [5, 6]. DUV-FSM provides high-quality images of specimens without destructive techniques or unnecessary optical sectioning. It offers excellent color and texture contrast between cancer and normal/benign breast tissue, which aids in detecting cancer cells at the edge of the surgical specimen. This real-time imaging capability can significantly enhance the precision of intra-operative margin assessment.

While previous research has utilized texture analysis (TA) methods for intra-operative margin assessment with the DUV dataset, these methods have limitations, which include potential bias and reduced diagnostic accuracy for less common tissue types [6]. Deep learning models present a promising alternative for automated breast cancer classification

of DUV images. These models can automatically learn complex patterns and hierarchical features directly from the data, offering improved generalization and adaptability across diverse tissue types. This adaptability could lead to more accurate and robust classification results, overcoming the limitations of texture analysis methods.

The motivation behind this study is to leverage the power of deep learning techniques to locate cancerous regions in DUV Whole Surface Images accurately. Doing so aims to enhance the precision of intra-operative margin assessment, thereby reducing breast cancer re-excision rates following BCS. This approach could ultimately improve patient outcomes and the overall effectiveness of breast cancer treatment. By providing real-time feedback during surgery, surgeons can make more informed decisions, ensuring the complete removal of cancer cells and reducing the likelihood of cancer recurrence. This study represents a significant step forward in applying deep learning techniques in medical imaging, particularly in the context of breast cancer surgery.

1.2 Objectives

The primary objective of this study is to develop an automated method for breast cancer classification in DUV Whole Surface Images. This objective addresses the critical need for accurate intra-operative margin assessment during breast cancer surgery. Margin assessment is crucial in breast-conserving surgery (BCS), as it helps determine whether all cancerous cells have been removed. Incomplete removal of cancerous cells can lead to recurrence of the disease and necessitate further surgery, known as re-excision. Therefore, enhancing the accuracy of margin assessment is paramount in reducing breast cancer re-excision rates following BCS.

The proposed method aims to provide surgeons with a more precise, real-time tool

during operations. The ability to accurately classify and localize cancerous regions in DUV Whole Surface Images in real-time could significantly improve surgical decision-making. Surgeons could be more confident in their decisions regarding the extent of tissue resection, potentially reducing the likelihood of leaving behind cancerous cells. This improved precision leads to better patient outcomes, decreasing the need for subsequent surgeries and the associated physical and psychological stress for patients.

Furthermore, the study aims to contribute to the broader medical imaging and machine learning field by demonstrating the potential of deep learning techniques in improving disease detection and classification. The proposed method leverages advanced deep learning techniques, including transfer learning and Grad-CAM++, to extract meaningful features from DUV Whole Surface Images and make accurate classification decisions. By demonstrating the effectiveness of these techniques in a challenging real-world application, the study could inspire further research and development in this field.

In summary, this study aims to develop a more accurate and efficient method for breast cancer classification in DUV Whole Surface Images and demonstrate the potential of deep learning techniques in enhancing clinical decision-making and patient outcomes. The ultimate goal is to provide a tool that can offer real-time, accurate feedback during surgery, thereby improving the precision of intra-operative margin assessment, reducing the need for re-excision, and potentially improving the overall effectiveness of breast cancer treatment.

1.3 Contributions

This research makes several significant contributions to the field of breast cancer classification using deep learning techniques:

1. **Novel Method:** This study introduces a novel method for breast cancer classification in DUV Whole Surface Images. The proposed method combines patch-level classification using transfer learning with regional importance maps generated through Grad-CAM++. This unique approach allows for accurate localization and classification of cancerous regions within DUV Whole Surface Images. The methodology is designed to enhance the precision of intra-operative margin assessment during breast cancer surgery, potentially reducing breast cancer re-excision rates following BCS.

2. **Enhanced Accuracy:** The proposed method demonstrates high classification accuracy in determining each Whole Surface Image's malignant or normal/benign labels. The method achieved an impressive classification accuracy of 95.0%, indicating its potential for accurate intra-operative margin assessment during breast cancer surgery. This enhanced accuracy is crucial for reducing breast cancer re-excision rates following BCS, ultimately improving patient outcomes and the overall effectiveness of breast cancer treatment.

3. **Robustness with Limited Data:** Despite the limited dataset size, the proposed method exhibits robustness. It uses transfer learning and a gradient-boosting algorithm to minimize overfitting and maximize the utility of the available data. The method's robustness, even with a limited dataset, offers advantages over other deep learning techniques, showcasing the potential of ensemble learning-based approaches in medical imaging.

4. **Potential Clinical Applications:** The proposed method has potential clinical applications in real-time assessment of margin status during breast cancer surgery. It can provide real-time guidance to surgeons during operations, ensuring they make better decisions about resection margins. Furthermore, it can help develop personalized treatment

plans, tailoring the surgical approach to the patient's needs. The proposed method could be integrated into surgical navigation systems to improve surgical interventions' precision and minimize damage to healthy tissues.

1.4 Overview of the Proposed Method

The proposed method in this study presents a robust and innovative solution for breast cancer classification in DUV Whole Surface Images. The novel method combines patch-level classification using transfer learning with regional importance maps generated through Grad-CAM++. This combination allows the method to focus on highly significant regions within the images, enhancing the robustness of the classification process and improving the accuracy of cancer detection.

In the initial stage of the process, DUV Whole Surface Images are divided into patches. This division is a strategic move to handle the images' high-resolution nature and focus on localized features indicative of cancerous tissues. These patches are then used to extract convolutional features using a pre-trained ResNet50 model. ResNet50, a deep convolutional neural network known for its residual learning framework, is particularly effective at handling complex image classification tasks. By leveraging a pre-trained ResNet50 model, the method can effectively utilize the power of transfer learning, taking advantage of the model's ability to extract high-level features from the DUV Whole Surface Image patches.

These extracted features are subsequently used to train an XGBoost classifier for patch-level classification. XGBoost, a gradient-boosting framework, is known for its efficiency and performance. It builds an ensemble of weak prediction models, typically decision trees, in a stage-wise fashion. It generalizes the model by allowing the optimization of

an arbitrary differentiable loss function, making it a powerful tool for classification tasks.

Concurrently, regional importance maps for the DUV Whole Surface Images are calculated using Grad-CAM++ on a pre-trained DenseNet169 model. DenseNet169, another deep convolutional neural network, is known for its densely connected layers, significantly reducing the number of parameters while maintaining performance. The Grad-CAM++ algorithm, applied to the DenseNet169 model, generates heatmaps that highlight the regions in the image that are most important for classification. These maps help to localize crucial areas in the Whole Surface Image for margin assessment, providing a visual explanation of the model's decision-making process. The final stage involves a decision fusion method, merging the patch-level classification results with the regional importance map. This fusion enhances the classification accuracy at the Whole Surface Image level, assigning a malignant or normal/benign label to each Whole Surface Image with increased confidence. This fusion process ensures the final classification decision considers the localized patch-level classifications and the overall regional importance within the image.

The potential of the proposed method extends beyond breast cancer surgery. It could be applied to a real-time assessment of margin status during other types of cancer surgeries. Providing accurate and immediate feedback on the presence of cancerous cells at the surgical margins could significantly improve surgical outcomes and reduce the need for subsequent surgeries. By integrating this method into the surgical workflow, surgeons could make more informed decisions during operations, potentially reducing the likelihood of leaving behind cancerous tissues and improving patient outcomes.

1.5 Overview of the Thesis Structure

This thesis is organized into five chapters, each providing a detailed exploration of the research process, from the initial literature review to the conclusions.

Chapter 2: Literature Review and Prior Methods: This chapter comprehensively reviews the existing literature and prior methods relevant to the study. It is divided into two sections:

- 2.1 Literature Review: This section reviews previous studies and methodologies on breast cancer classification using machine learning techniques.
- 2.2 Background: This section provides essential information on breast cancer, deep learning techniques, and the importance of accurate margin assessment during breast cancer surgery.

Chapter 3: Methods: This chapter outlines the methodology used in the study, including:

- 3.1 Data Collection: This section describes the data collection process and the preprocessing steps for the DUV Whole Surface Images for analysis.
- 3.2 Method: This section details the proposed method for breast cancer classification, including transfer learning for patch-level classification and Grad-CAM++ for generating regional importance maps.

Chapter 4: Experiment and Results: This chapter presents the experimental setup and results of the study, including:

- 4.1 Experimental Setup: This section outlines the experimental design, including the hardware and software used and the evaluation metrics.
- 4.2 Visual Inspection: This section analyzes the DUV Whole Surface Images and the corresponding regional importance maps.
- 4.3 Classification Performance: This section presents the quantitative results of the study, including the accuracy, sensitivity, and specificity of the proposed method.
- 4.4 Misclassified Samples: This section discusses the samples misclassified by the proposed method and provides possible explanations.

Chapter 5: Discussion: This chapter discusses the results of the study, its limitations, and potential future research directions:

- 5.1 Methodology and Results: This section provides a detailed discussion of the proposed method and its performance.
- 5.2 Limitations: This section acknowledges the study's limitations and suggests areas for improvement.

Chapter 6: Conclusion: This chapter concludes the thesis by summarizing the findings and contributions, discussing their implications for clinical practice, suggesting future work, and providing final remarks:

- 6.1 Summary of Findings and Contributions: This section summarizes the main findings of the study and its contributions to the field of breast cancer classification.

- 6.2 Implications for Clinical Practice: This section discusses how the study's findings could impact clinical practice, particularly in the context of breast cancer surgery.
- 6.3 Future Work: This section suggests potential future research directions based on the study's findings.
- 6.4 Final Remarks: This section provides concluding remarks, reflecting on the study's significance and potential impact.

CHAPTER 2
LITERATURE REVIEW AND BACKGROUND

2.1 Literature Review

The application of deep learning methods in the medical field has seen significant advancements in recent years, particularly in disease detection. Image classification algorithms have proven to be instrumental in detecting and localizing disease regions of interest (ROI), with breast cancer classification and margin assessment emerging as significant research areas [7, 8].

Several studies have proposed various methods for breast cancer classification. For instance, a study by Zhang et al. combined a convolutional graph network with a convolutional neural network (CNN) to classify breast cancer [9]. This approach leverages the strengths of both graph networks and CNNs, with the graph network capturing the global structure of the data and the CNN capturing local features. This combination resulted in higher classification accuracy, demonstrating the potential of hybrid models in medical image analysis.

Another study by Singh et al. utilized a patch-level CNN for classification [10]. This approach involves dividing the image into smaller, non-overlapping patches and extracting spatial information from each patch using integrated CNN and filter algorithms. This method allows for more detailed image analysis and can capture local variations that might be missed when analyzing the image as a whole.

Paul Gamble et al. segmented tumor and non-tumor areas with a patch-level CNN for tumor localization, enabling deeper tissue analysis [11]. This approach allows for identifying specific regions within the image that are likely to be cancerous, providing valuable

information for diagnosis and treatment planning.

Jaber et al. adopted a similar patch-level CNN approach to classify breast cancer subtypes [12]. By focusing on specific regions within the image, this approach can identify subtle differences between different breast cancer subtypes, providing more detailed and accurate classification results.

Li et al. used a patch screening method with a CNN and clustering methods for different tumor types [13]. This approach combines the strengths of CNNs and clustering algorithms, allowing for identifying different tumor types within the same image.

Farahmand et al. and Couture et al. used transfer learning approaches to determine breast cancer grades and predict specific tumor grades, respectively [14, 15]. Transfer learning involves using a pre-trained model on a new problem, allowing for faster training and potentially better performance, especially when the available data is limited.

Ye et al. incorporated a transfer learning method to address data imbalance, improving classification results using a VGG-19 model pre-trained on ImageNet weights [16]. This approach allows for using a large, diverse dataset to pre-train the model, which can then be fine-tuned for the specific task of breast cancer classification.

However, these studies primarily trained and tested their networks on publicly available H&E image datasets. While these datasets are valuable for developing and testing new methods, they may generalize poorly to other data types, such as DUV images. DUV images present unique challenges due to their relative novelty and the limited number of subjects available for study.

To overcome this challenge, this study proposes a novel approach that combines patch-level classification using transfer learning with regional importance maps generated

through Grad-CAM++. By dividing the DUV Whole Surface Images into small, non-overlapping patches, the network can learn better generalizations, boosting the training data at the patch level and helping localize cancer regions in DUV Whole Surface Images for margin evaluation. The performance accuracy of this approach is tested by classifying 60 images at the Whole Surface Image level. This approach aims to leverage the strengths of deep learning and transfer learning to improve the accuracy and interpretability of breast cancer classification in DUV images.

2.2 Background

Convolutional Neural Networks (CNNs) are a specialized kind of neural network for processing data with a grid-like topology, such as an image that can be thought of as a 2D grid of pixels. CNNs have succeeded incredibly in fields such as image and video recognition, recommender systems, and natural language processing [17].

The critical aspect of CNNs is their ability to exploit spatial correlations within the input data. They achieve this by applying filters and pooling layers, which systematically reduce the dimensionality of the input data while preserving its hierarchical structure. This makes CNNs particularly effective for object detection and recognition within images, where the object of interest can be anywhere in the image.

A typical CNN consists of an input layer, multiple hidden layers, and an output layer. The hidden layers often include a series of convolutional layers, pooling layers, fully connected layers, and normalization layers.

Convolutional Layer

This is the core building block of a CNN. The layer's parameters consist of a set of learnable filters (or kernels), which have a small receptive field, but extend through the

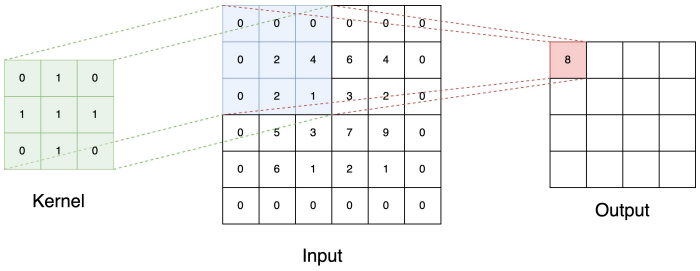

Figure 2.1: An example of how a convolution layer works with a 3x3 kernel that outputs a new feature map

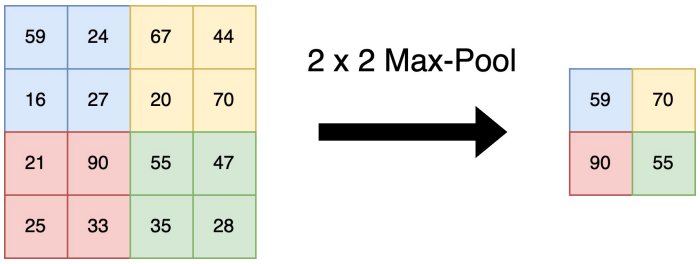

Figure 2.2: An example of how max pooling works with an input feature map that has four sub-regions

full depth of the input volume. During the forward pass, each filter is convolved across the width and height of the input volume, computing the dot product between the filter entries and the input and producing a 2-dimensional activation map of that filter. As a result, the network learns filters that activate when it detects some specific type of feature at some spatial position in the input.

Pooling Layer

The layers of pooling (or subsampling) reduce the input volume's spatial dimensions (width and height). They are used to decrease the computational complexity, control overfitting, and make the network invariant to small translations. A commonly used pooling strategy is max pooling, which extracts subregions of the input (e.g., 2x2 pixels) and keeps only the maximum value.

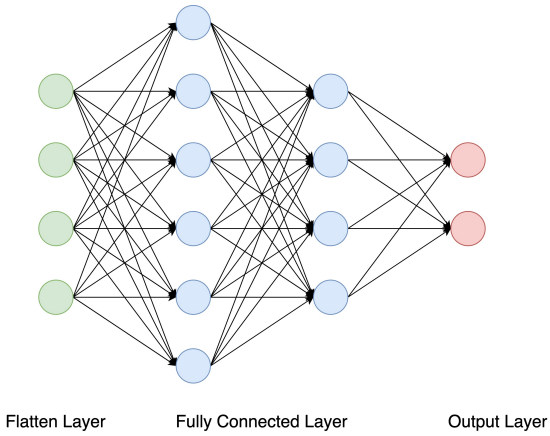

Figure 2.3: Simple visualization of a Fully Connected Layer in a convolutional neural network

Fully Connected Layer

Fully connected layers connect every neuron in one layer to every neuron in another layer. It is the same as the traditional Multi-Layer Perceptron (MLP). The fully connected layer is usually the final layer that uses the previous layers' high-level features to perform classification.

Normalization Layer

Normalization layers, such as Batch Normalization, are used to normalize the activations of a given input volume before passing it to the next layer. It helps improve the speed, performance, and stability of the network.

ResNet and DenseNet

Two specific types of CNNs, ResNet, and DenseNet, are utilized in this study. ResNet, or Residual Network, introduces the concept of skip connections, or shortcuts, to enable the training of much deeper networks [18]. The core idea of ResNet is to introduce these shortcuts that skip one or more layers. This allows the model to learn an identity function that ensures the higher layer will perform at least as well as the lower layer and

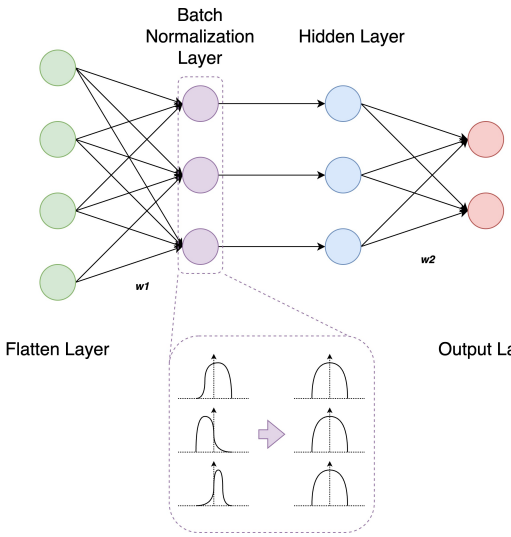

Figure 2.4: A simplified workflow of how batch normalization is applied to inputs

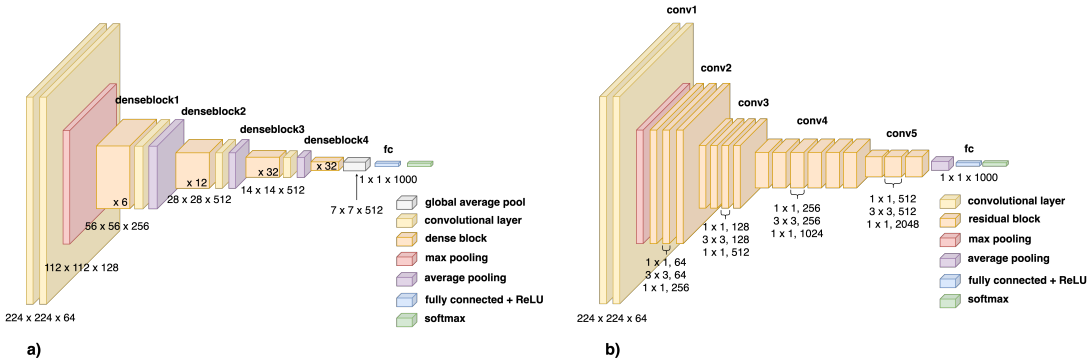

Figure 2.5: Top level block diagrams of a) ResNet50, and b) DenseNet169 models used in this study

not worse. This helps to combat the problem of vanishing gradients, a common issue in deep neural networks where the gradients of the loss function approach zero, making the network hard to train.

DenseNet, or Densely Connected Convolutional Networks, extends the concept of ResNet [19]. In DenseNet, each layer is connected directly to every other layer in a feed-forward fashion. This dense connectivity pattern encourages feature reuse throughout the network and significantly reduces the number of parameters, making the network more efficient. In this study, DenseNet169, a variant of DenseNet, is used to generate regional

importance maps for the DUV Whole Surface Images.

2.2.1 Microscopy with UV Surface Excitation for Breast Cancer Classification

Microscopy with UV surface excitation (MUSE) is a cutting-edge, non-destructive imaging technology that has demonstrated potential in medical imaging. MUSE employs multiple fluorescence dyes to stain fresh, unfixed tissue, generating fluorescence images with microscopic resolution, sharpness, and contrast [20]. However, earlier research indicated that further development was necessary for MUSE to apply in breast surgical procedures.

MUSE has been advanced to include a fully automated three-dimensional (3D) sample translation, facilitating the imaging of fresh tissue at a rate of 5 minutes per square centimeter. This advancement also incorporated an algorithm capable of creating a fluorescent analog of conventional H&E images [21]. This progress demonstrated that MUSE could provide microscopic visualization of breast margin surfaces at speeds suitable for intraoperative use.

A study explored the potential of MUSE as an intraoperative tool for margin assessment during BCS [22]. The aim was to identify the features of MUSE images that could distinguish fresh, unprocessed malignant from normal/benign breast tissues, assess the technology's accuracy, and determine the speed at which a breast tissue sample could be surveyed. To achieve adequate accuracy and time efficiency required in a clinical setting for margin assessment, a combination of deep ultraviolet (DUV) excitation and low magnification with a slightly reduced spatial resolution was utilized to achieve a faster imaging speed.

The findings demonstrated that a low-cost deep-ultraviolet fluorescence scanning

microscope (DUV-FSM) could rapidly image fresh breast tissues with excellent contrast at a resolution sufficient to resolve cells of different tissue types. Visual interpretation of MUSE images could achieve excellent sensitivity and specificity in distinguishing invasive cancer from normal breast tissue. Furthermore, the nuclear–cytoplasm ratio (N/C) could be used for quantitative assessment to differentiate cancer from normal breast tissue.

The images generated by MUSE can be utilized for breast cancer classification. These images' high contrast and resolution allow for accurately identifying and localizing cancerous cells within the tissue. This capability, combined with the speed at which MUSE can image a tissue sample, makes it a promising tool for real-time intraoperative margin assessment during BCS. This could reduce the need for additional surgeries and improve patient outcomes. Thus, automated breast cancer classification can aid in the real-time detection of breast cancer tissues during BCS.

2.2.2 Transfer Learning

Transfer learning is a machine learning method using a pre-trained model on a new problem [23]. It is a popular approach in deep learning where pre-trained models are used as the starting point for computer vision and natural language processing tasks. Transfer learning has the benefit of decreasing the training time for a neural network model and can result in lower generalization errors.

This study uses transfer learning to extract convolutional features from DUV Whole Surface Images' patches using a pre-trained ResNet50 model with ImageNet weights. ImageNet is an extensive database of annotated images designed for use in visual object recognition research. Using a pre-trained model on ImageNet, the study can leverage the model's ability to recognize low-level features (like edges and textures) and high-level features (like

shapes or objects) from the images. These extracted features are then used to train an XGBoost classifier for patch-level classification.

2.2.3 XGBoost

XGBoost, short for eXtreme Gradient Boosting, efficiently implements the gradient boosting framework. Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees[24]. XGBoost improves upon the base Gradient Boosting Machines (GBM) framework by adding a regularization term to the loss function to control the model's complexity, making the model less likely to overfit.

XGBoost is known for its execution speed and model performance. Its features are designed to make it efficient and flexible, such as parallelizable tree building, handling sparse data and missing values, and the ability to customize the optimization objective and evaluation metrics. This study uses XGBoost to classify the extracted features into malignant or benign classes.

2.2.4 Grad-CAM++

Gradient-weighted Class Activation Mapping++ (Grad-CAM++) generates visual explanations for decisions from various CNN-based models [25]. It is an extension of the Grad-CAM method, which uses the gradient information flowing into the last convolutional layer of the CNN to understand each neuron for a decision of interest. Grad-CAM++ improves upon this by providing more fine-grained saliency maps and works with a broader range of CNN model architectures.

Grad-CAM++ uses the gradients of any target concept (e.g., logits for 'dog' or even a caption), flowing into the final convolutional layer to produce a coarse localization map

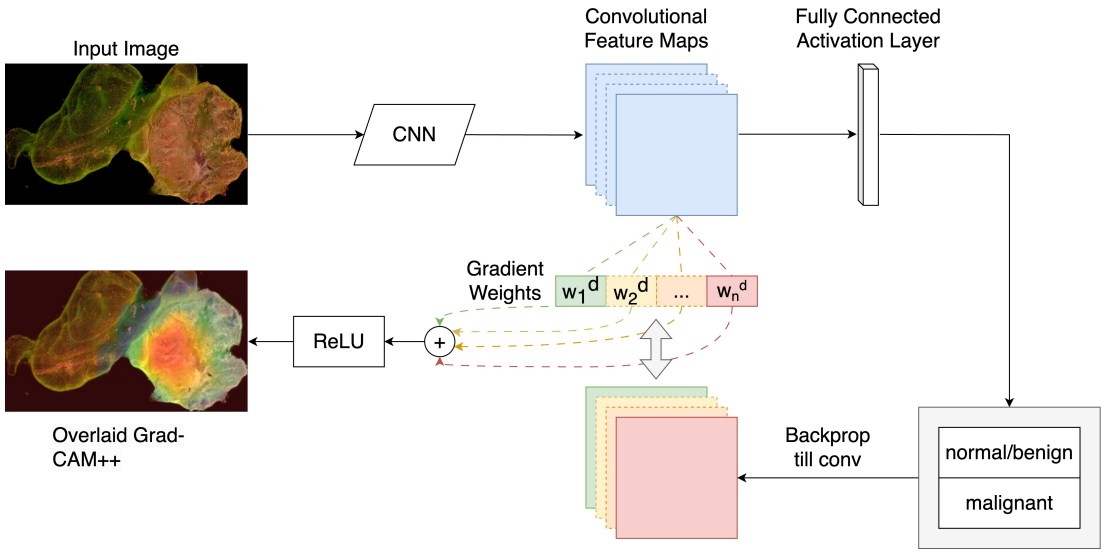

Figure 2.6: Top-level diagram of how the process for Grad-CAM++ works with an input image and a trained convolutional neural network.

highlighting the critical regions in the image for predicting the concept. These localization maps, also known as heatmaps, provide visual explanations of the model's decision by highlighting the critical regions in the image for the prediction. This helps in understanding what features the model focuses on to make its decisions.

In this study, Grad-CAM++ generates regional importance maps for the DUV Whole Surface Images, highlighting the most critical areas for the classification decision. This visually explains the model's decision, making it easier to understand and interpret its predictions.

By combining these techniques, this study aims to enhance the robustness of breast cancer classification in DUV Whole Surface Images and potentially improve real-time assessment of margin status during breast cancer surgery. The proposed method leverages the power of deep learning, transfer learning, and visual explanations to achieve high classification accuracy and provide interpretable results.

CHAPTER 3
METHODS

3.1 Data Collection

This study comprises 60 DUV images, including 24 normal/benign and 36 malignant samples. These images were collected from the Medical College of Wisconsin (MCW) tissue bank using a custom DUV Fluorescence Scanning Microscopy (DUV-FSM) system [5]. The data preprocessing steps for the imaging system that captured images of the DUV samples involve several stages to ensure accurate and enhanced visualization of the tissue surfaces.

To begin with, an inverted fluorescence microscope was transformed into a DUV-FSM system by integrating a 285nm LED and a fused silica ball lens. The LED was positioned on the right side of the microscope to provide oblique back illumination for the samples. A 325nm short-pass filter was placed in front of the LED to block emission spectrum tails in the visible range, preventing any potential overlap with fluorescence signals from the samples. The fused silica ball lens served as a condenser, converging the LED radiation to achieve a smaller illumination field with improved power density. These components were housed within a lens tube mounted on an adjustable arm for easy positioning. The power density of the LED light at the tissue surface was measured at $9.4 mW/cm$.

For imaging a lumpectomy specimen, the specimen was placed on one of its six margins in a quartz dish to minimize autofluorescence from the glass. The quartz dish had a diameter of 70mm and a thickness of 1.0mm. It was mounted on a motorized XY stage designed explicitly for fast mosaic imaging. This stage allowed for precise scanning of the specimen in a large area. The microscope's excitation/emission filter block was switched

to the open position, enabling the capture of fluorescent signals from multiple fluorophores using a color camera without switching emission filters during the imaging process.

A cooled USB 3.0 camera with a large image sensor, large pixel size, low dark noise, and high image transfer speed was employed for image acquisition. The camera had dimensions of 2748×2200pixels, with a pixel size of $4.54\mu m$ and an active area of $14.6 \times 12.8mm$. To balance good lateral resolution (2 to $3\mu m$) and a large effective imaging area of $3.48 \times 2.78mm$, a $4\times$apochromatic long working distance objective lens with a numerical aperture of 0.13 was selected. The field-of-view (FOV) of the objective lens was slightly larger than the imaging area of the camera to avoid distortions at the edge of the FOV. The microscope was enclosed in a dark enclosure to prevent exposure to DUV light and eliminate the background from room light.

Before imaging the specimens, they underwent staining using propidium iodide (PI) for nuclear staining and eosin Y (EY) for staining of cytoplasm and connective tissues. Both PI and EY were capable of effective excitation at $285nm$. Staining was carried out by dissolving PI and EY in pH 7.2 phosphate-buffered saline (PBS) to $100\mu g/ml$ and $1.0mg/ml$, respectively. Each specimen was sequentially stained in PI solution for 1 minute, followed by EY solution for 20 seconds, then rinsed in PBS for 10 seconds. After staining, the specimen was placed on the quartz plate of the specimen holder. Air bubbles between the tissue and plate were gently removed by flattening the tissue using a broad pallet knife. Excess liquid on the edges was carefully removed using a Kimwipe. Only images from one side of the samples were included in the data analysis, as both sides were thin and similar.

The specimen holder, containing the tissue specimen, was immobilized on the mo-

torized XY stage, with the focal plane set at the bottom surface of the specimen. Mosaic images were captured with overlapping regions of $0.75mm$ in the X direction and $0.60mm$ in the Y direction, striking a balance between speed and stitching accuracy. The number of scanning steps depended on the size of the specimen, the effective imaging area, and the chosen dimensions for overlapping. The camera temperature was set to $-18C$ to minimize electronic noise. Customized software developed in Microsoft Visual C#.NET controlled image acquisition and motor movements. The acquired images of each specimen, captured with a constant exposure time ranging from 50 to $100ms$, depending on the tissue type, were saved in TIFF format with dimensions of 2748×2200 pixels.

After imaging, the raw images were transformed into the hue-saturation-lightness (HSV) color space. The Fiji open-source image processing package was utilized for processing the tissue images. The saturation and lightness channels of the images underwent background and shading correction using a Fiji plugin called BaSiC. This correction compensated for any uneven and tilted illumination caused by the DUV LED. The color space transformation was necessary to maintain the original color information during the illumination correction. Subsequently, the images were transformed back into the red-green-blue (RGB) color space.

To stitch the mosaic images together seamlessly, a Fiji plugin was used to ensure the comprehensive combination of the captured images, resulting in a holistic view of the tissue surface. Finally, histogram equalization was applied to the red and green color channels of the stitched image to enhance visual contrast and improve the overall quality of the image.

In summary, the data collection steps for the DUV imaging system involved sample preparation, image acquisition with mosaic scanning, color space transformation, back-

Extracted Patches

	Number of Patches
Normal/Benign	25,024
Malignant	9,444
Total	34,468

DUV Whole Surface Images

	Number of Samples
Normal/Benign	24
Malignant	36
Total	60

Table 3.1: Dataset Information: Extracted Patches provide the number of patches that were extracted for normal/benign and malignant. DUV Whole Surface Images provides several samples for normal/benign and malignant.

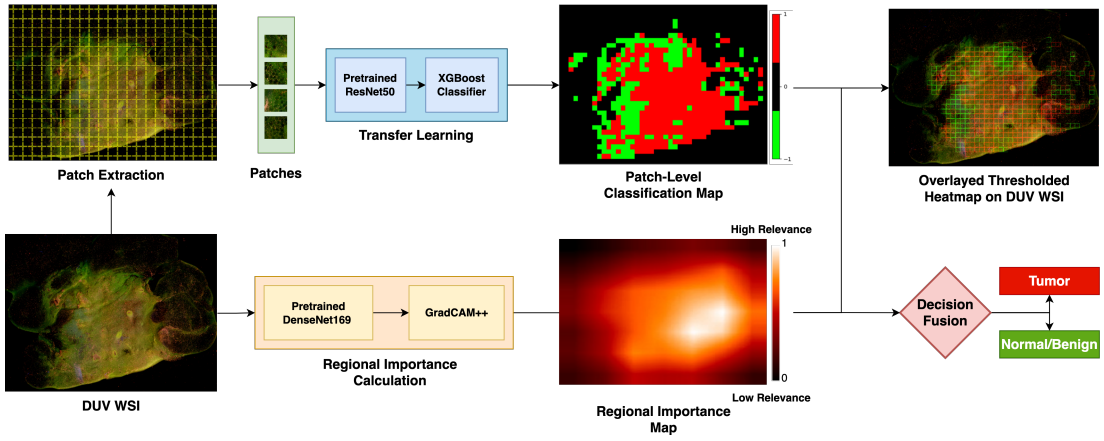

Figure 3.1: Overview of the proposed method: The Whole Surface Images are divided into patches to boost the dataset and localize cancer detection. With a pre-trained ResNet50 model, the convolutional features are extracted for each patch and used to train an XGBoost classifier for patch-level classification. Grad-CAM++ on a pre-trained DenseNet169 model calculates the regional importance map for the DUV Whole Surface Images. The patch-level classification results are merged with the regional importance map in a decision fusion for the Whole Surface Images-level prediction.

ground, shading correction, image stitching, and contrast enhancement. These steps ensured accurate visualization and improved the quality of the captured DUV images for further analysis.

3.2 Method

Fig. 3.1 describes the proposed breast cancer classification method for DUV image margin assessment. First, DUV Whole Surface Images are divided into patches to enhance the dataset and localize cancerous regions. A pre-trained ResNet50 model [18] extracts convolutional features for each patch, which are then used to train an XGBoost classifier

[26] for patch-level classification. The regional importance map for a DUV Whole Surface Image is calculated using Grad-CAM++ [25] on a pre-trained DenseNet169 model [19]. Finally, the patch-level classification results merge with the regional importance map in a decision fusion for the Whole Surface Image-level prediction.

3.2.1 Patch Extraction

The first step in our method involves dividing a Deep Ultraviolet (DUV) Whole Slide Image into multiple DUV patches. Each DUV Whole Surface Image, denoted as x_i, is divided based on a 2D grid system designed around each sample's field of view (FOV) Ω_i with non-overlapping patches Ω_i^j. This process is mathematically represented as:

$$\Omega_i = \cup_{j=1}^{N} \Omega_i^j, and \Omega_i^k \cap \Omega_i^l = \emptyset for \forall k, l, \tag{3.1}$$

where N represents the total number of patches in x_i, and k and l represent any patch indices. Since each DUV Whole Surface Image has different dimensions, the images are resized to the closest dimensions divisible into non-overlapping patches Ω_i^j with a constraint size of 400×400 pixels. The minimal alterations to the dimensions should not affect the quality of the morphological characteristics, such as cell density and infiltration, because the DUV Whole Surface Images are enormous. The patch images are extracted from each non-overlapping patch Ω_i^j of a DUV Whole Surface Image. To determine the validity of a patch, it is converted to grayscale, and its pixels are analyzed. If most of the pixels in the patch Ω_i^j are at least 80% foreground, then it is counted as a valid patch p_i^j. A pixel is a foreground if its grayscale value is greater than 5. It is important to remove the dark background to help with the localization and detection of breast cancer. This process

is shown in **Algorithm 1** *ExtractPatches*, and **Algorithm 2** *GetPercentage*.

Algorithm 1 *ExtractPatches* is used to extract foreground patches for each DUV Whole Surface Image.

$x_i \leftarrow$ DUV Whole Surface Image sample i
$\Omega_i \leftarrow$ 2D grid system on each sample's FOV
$\Omega_i^j \leftarrow$ Non-overlapping patch with 400×400 pixels
$p_i^j \leftarrow$ valid patch image
$M \leftarrow$ number of DUV Whole Surface Image samples
for $i < M$ **do**
  Load in sample x_i
  for Ω_i^j in Ω_i **do**
    if GetPercentage(Ω_i^j) ≥ 0.2 **then**
      Save Ω_i^j as valid patch p_i^j for Ω_i
    end if
  end for
end for

Algorithm 2 *GetPercentage* is used to determine the foreground percentage of a patch.

BackgroundThreshold $\leftarrow 5$
ImageSize $\leftarrow 400 \times 400$ pixels
PixelCounter $\leftarrow 0$
for $x < 400$ **do**
  for $y < 400$ **do**
    if Image[x][y] $\geq$ BackgroundThreshold **then**
      PixelCounter $\leftarrow$ PixelCounter$+1$
    end if
  end for
end for
return PixelCounter/ImageSize

3.2.2 Patch Classification with Transfer Learning

Convolutional neural networks (CNNs) have become a cornerstone in image classification, particularly in medical imaging, where they have shown great promise in diagnosing diseases. CNNs, such as ResNet, are designed to automatically and adaptively learn spatial hierarchies of features from the input images. However, deep architectures like ResNet can suffer from the 'vanishing gradient' problem. This issue arises when the back-

propagation of errors during the training process results in the gradients of the loss function becoming infinitesimally small. As the network depth increases, these small gradients can cause information loss, preventing the network from learning effectively.

To overcome this challenge, this study employs a transfer learning approach for patch classification. Transfer learning is a machine learning technique using a pre-trained model on a new problem. It is a popular approach in deep learning because it allows us to build accurate models in a time-saving way. This study uses transfer learning to determine whether each patch is malignant or normal/benign for all N DUV patches $p_i^j, j = 1, \ldots, N$.

The combination of transfer learning with the ResNet50 model and the XGBoost classifier allows for determining the tumor regions of interest (ROI) in the DUV Whole Surface Images. This determination is based on y_i^j with the relative patch locations. The proposed method can provide more accurate and localized predictions by focusing on these regions, enhancing the overall classification performance.

Deep Residual Learning with ResNet50

Residual learning is a crucial concept in ResNet architecture. It is based on learning the residual, or difference, between the input and output of a stack of layers rather than learning the original mapping directly. This is represented as $\mathscr{H}(\mathbf{x}) - \mathbf{x}$, where $\mathscr{H}(\mathbf{x})$ is the desired underlying mapping to be learned by the layers, and $\mathbf{x}$ is the input to the first of these layers.

The motivation behind residual learning is to address the degradation problem, where the network's performance degrades with the increase in depth. This problem arises due to the difficulty in approximating identity mappings by multiple non-linear layers. By reformulating the problem to learn the residual function, the network can more easily ap-

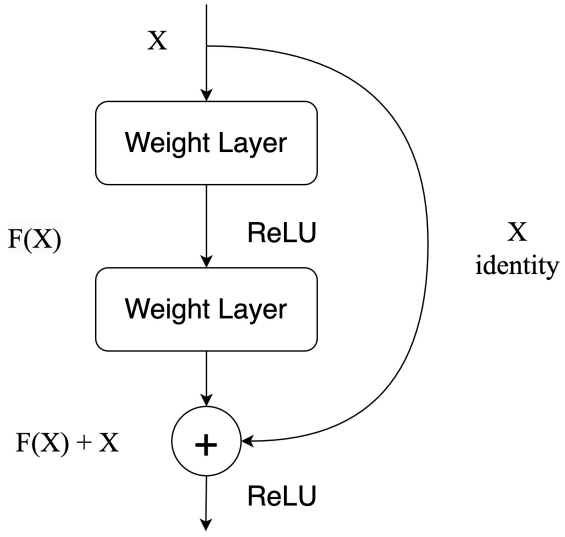

Figure 3.2: A block diagram of a residual block that features the shortcut connection with identity mapping

proximate the identity mapping by driving the weights of the multiple non-linear layers toward zero.

In practice, while identity mappings may not always be optimal, the residual learning reformulation helps to precondition the problem, making it easier for the solver to find perturbations concerning an identity mapping.

In the ResNet architecture, identity mapping is achieved through shortcut connections, which skip one or more layers. These connections are added to the residual function's output, forming the block's final output. This is represented as $\mathbf{y} = \mathscr{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}$, where $\mathbf{y}$ is the output vector of the layers, $\mathscr{F}(\mathbf{x}, \{W_i\})$ is the residual mapping to be learned, and $\mathbf{x}$ is the input vector.

The shortcut connections introduce neither extra parameters nor computation complexity, making them an attractive feature of the ResNet architecture. They also allow for a fair comparison between plain and residual networks that have the same number of

parameters, depth, width, and computational cost.

In cases where the dimensions of $\mathbf{x}$ and $\mathscr{F}$ are not equal, a linear projection W_s can be performed by the shortcut connections to match the dimensions. However, experiments have shown that identity mapping is sufficient for addressing the degradation problem and is economical, so W_s is only used when necessary to match dimensions.

The form of the residual function $\mathscr{F}$ is flexible and can represent multiple convolutional layers. The element-wise addition is performed on two feature maps, channel by channel, allowing the network to learn complex hierarchical features directly from the data.

This approach extracts the features from the final layer of a pre-trained ResNet50 network on the ImageNet dataset. ImageNet is an extensive database of annotated images designed for use in visual object recognition software research. The ResNet50 model, pre-trained on this dataset, has already learned to recognize various features from various images. We can leverage these learned features for classifying DUV patches using this pre-trained model.

Understanding XGBoost

The extracted features from the DUV patches are then fed into an XGBoost classifier. XGBoost, short for eXtreme Gradient Boosting, efficiently implements the gradient boosting framework. It is known for its speed and performance, outperforming other algorithms in machine learning competitions and applied work. It leverages creating a model as an ensemble of weak prediction models, typically decision trees. The model is trained in an additive manner, where new models are added to correct the errors made by existing models. The XGboost classifier predicts the binary output $y_i^j \in -1, +1$ as malignant or normal/benign for each patch p_i^j.

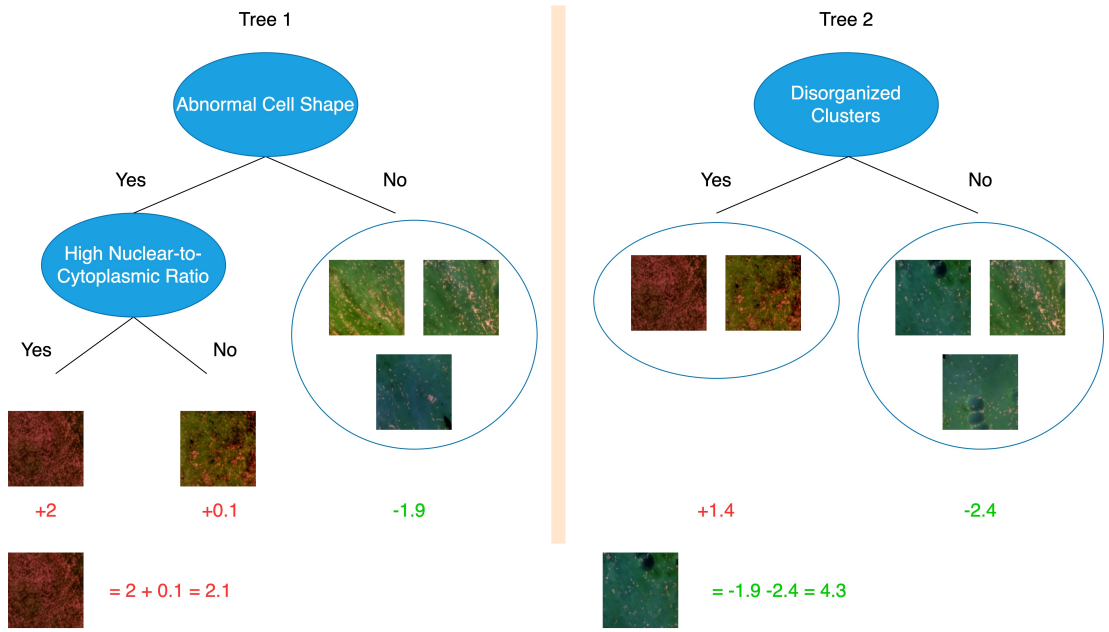

Figure 3.3: Example of how classification and regression trees (CART) work for gradient trees like XGBoost.

The tree ensemble model is a critical component of the XGBoost algorithm. It is an additive model that combines the predictions of multiple base learners (in this case, decision trees) to generate the final output. The model is defined as follows:

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \quad f_k \in \mathscr{F}$$

In this equation, $\phi(\mathbf{x}_i)$ represents the prediction for the i-th instance, and $f_k(\mathbf{x}_i)$ represents the k-th base learner (a decision tree) in the ensemble. The base learners are functions from the function space $\mathscr{F}$, the space of all possible decision trees. Each base learner f_k corresponds to an independent tree structure and leaf weights.

The ensemble model makes its prediction $\hat{y}_i$, for instance, $\mathbf{x}_i$ by summing up the predictions of all the base learners. This is an additive model, meaning that each base learner contributes to the final prediction independently of the others. The goal of the

learning process is to find the set of base learners that minimize the regularized objective function.

The learning objective in XGBoost is a regularized objective function that combines a differentiable convex loss function and a regularization term. The loss function measures the difference between the prediction and the target, while the regularization term penalizes the complexity of the model, helping to avoid overfitting. The regularization term is crucial as it encourages the model to use simple and predictive functions, making it more interpretable and less prone to overfitting.

Here, $l(\hat{y}_i, y_i)$ is the loss function that measures the difference between the prediction $\hat{y}_i$ and the target y_i, and $\Omega(f_k)$ is the regularization term that penalizes the complexity of the base learner f_k. The goal of the learning process is to find the set of base learners that minimize this regularized objective function.

The regularized learning objective is defined as:

$$\mathscr{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \text{ where } \Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2$$

Here, l is the loss function, Ω is the regularization term, γ is the complexity control parameter, T is the number of leaves in the tree, λ is the L2 regularization term on weights, and w is the vector of scores on leaves.

In XGBoost, the model is trained using a second-order gradient boosting algorithm. This involves iteratively adding new functions to the model that most improve the regularized objective function. The improvement is approximated using a second-order Taylor expansion, which provides a good approximation for a wide range of loss functions.

The objective function at each step is given by:

$$\mathscr{L}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t)$$

Here, g_i and h_i are the first and second-order gradients of the loss function, and f_t is the function to be added at the t-th iteration.

XGBoost also incorporates two additional techniques to prevent overfitting: shrinkage and column subsampling. Shrinkage scales the contribution of each tree by a factor after each boosting step, reducing the influence of each tree and leaving space for future trees to improve the model. Column subsampling is a technique used to prevent overfitting and speed up computations, where a subset of features is randomly sampled for each tree. This technique is similar to the one used in Random Forests and effectively prevents overfitting.

3.2.3 Calculation of Regional Importance with Grad-CAM++

Grad-CAM++ is a method that provides visual explanations for the decisions made by a model, making it a valuable tool in the realm of explainable artificial intelligence. When applied to deep neural networks, Grad-CAM++ can visualize gradients with pixel-wise weighted feature maps, explaining the decisions made at the output layer while considering the spatial information and high-level semantics from the preceding convolutional layers. Given the complexity of the details in the patches and the DUV images, applying Grad-CAM++ to a network that can retain as much information as possible is crucial.

DenseNet's unique connectivity pattern and feature concatenation make it suitable for applying Grad-CAM++. One of the unique aspects of DenseNet is that it eliminates redundant feature maps after the concatenation of the feature maps, which results in fewer

parameters. By concatenating these feature maps from different feature maps, it increases the variation in the inputs of the subsequent layers. The bottleneck and compression layers of the DenseNet architecture are effective against overfitting due to fewer parameters needed. The bottleneck layers help reduce the number of inputs from previous k-output feature maps, which improves computational efficiency. For the compression layers, they improve model compactness by reducing the feature maps that a dense block generates. The feature maps will overlap at the last output layer, highlighting the most relevant features which benefit Grad-CAM++. These characteristics make DenseNet a practical choice for generating visual explanations with Grad-CAM++ in complex image analysis tasks, such as those involving DUV images.

DenseNets: An Overview

Dense Convolutional Networks, or DenseNets, are a type of convolutional neural network (CNN) that are unique in their connectivity pattern. Unlike traditional CNNs, where each layer is only connected to the next layer, DenseNets connect each layer to every other layer in a feed-forward fashion. This dense connectivity pattern has several key advantages, including improved gradient flow, encouraging feature reuse, and substantially reducing the number of parameters.

In a DenseNet, the ℓ^{th} layer receives the feature-maps of all preceding layers, $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{\ell-1}$, as input:

$$\mathbf{x}_\ell = H_\ell([\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{\ell-1}])$$

This dense connectivity pattern dramatically improves the flow of information and

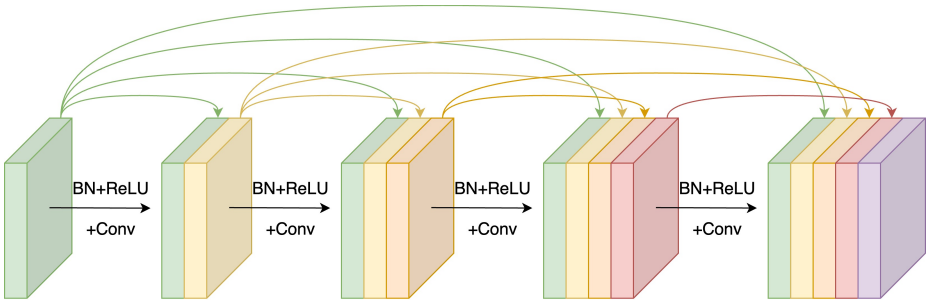

Figure 3.4: A diagram of how a Dense Block has its layer connected in the forward pass with all previous layers combined

gradients throughout the network, which makes the optimization of very deep networks tractable. Each layer has direct access to the gradients from the loss function and the original input signal, leading to implicit deep supervision, as seen in Fig. 3.4.

Each layer in a DenseNet consists of a composite function of three consecutive operations: batch normalization (BN), followed by a rectified linear unit (ReLU), and a 3×3 convolution (Conv). This sequence of operations is applied to the concatenated output of all preceding layers.

DenseNets are divided into multiple densely connected dense blocks. Layers between these blocks, referred to as transition layers, change the size of feature maps and are used to perform down-sampling. These transition layers consist of a batch normalization layer, a 1×1 convolutional layer, and a 2×2 average pooling layer.

The growth rate of a DenseNet, denoted as k, refers to the number of feature maps that each layer contributes to the global state of the network. A relatively small growth rate is sufficient to obtain state-of-the-art results, as each layer has access to all the preceding feature maps in its block and, therefore, to the network's "collective knowledge."

To improve computational efficiency, a 1×1 convolution can be introduced as a bottleneck layer before each 3×3 convolution to reduce the number of input feature maps.

The number of feature maps can be reduced at transition layers to improve model compactness further. This is achieved by introducing a compression factor $0 < \theta \leq 1$. When $\theta = 1$, the number of feature maps across transition layers remains unchanged. DenseNets with $\theta < 1$ are referred to as DenseNet-C, and DenseNets with both bottleneck and transition layers with $\theta < 1$ are referred to as DenseNet-BC.

In DenseNet implementations, the network is divided into three dense blocks, each having an equal number of layers. Before entering the first dense block, a convolution with 16 output channels is performed on the input images. For convolutional layers with kernel size 3×3, each side of the inputs is zero-padded by one pixel to keep the feature-map size fixed. Transition layers between two contiguous dense blocks consist of a 1×1 convolution followed by 2×2 average pooling. At the end of the last dense block, a global average pooling is performed, followed by the attachment of a softmax classifier. The feature-map sizes in the three dense blocks are $32 \times 32, 16 \times 16$, and 8×8, respectively. These feature maps allow an explainable artificial intelligence tool such as Grad-CAM++ to help visualize model predictions.

Grad-CAM++ Extraction of Feature Maps

Grad-CAM++ extracts weight maps from feature maps using a series of equations. The final classification score is defined as Y^c for class c as a linear combination of global average pooled convolutional feature maps A^k in the last layer over the FOV Ω_i:

$$Y^c = \sum_k w_k^c . \sum_{l \in \Omega_i} A_l^k \tag{3.2}$$

The gradient of the classification score Y^c for class c before the softmax layer con-

cerning the final convolutional layer feature map activation A^k is defined as $\frac{\partial Y^c}{\partial A^k}$. The weights w_k^c for a particular feature map A^k are defined by its global average pooled gradients:

$$w_k^c = \frac{1}{Z} \sum_{l \in \Omega_i} \frac{\partial Y^c}{\partial A_l^k}, \tag{3.3}$$

where Z represents the activation map's number of pixels.

With Grad-CAM implementations, visualizations are limited if there are multiple instances of a class in the input image x_i, as different spatial footprints of classes can cause different feature maps. The feature maps with small footprints will not be seen in the final saliency map.

To address this issue, a weighted average of pixel-wise gradients can be taken instead, where Eq. 3 is restructured as:

$$w_k^c = \sum_{l \in \Omega_i} \alpha_l^{kc} . ReLU(\frac{\partial Y^c}{\partial A_l^k}), \tag{3.4}$$

where $ReLU$ is the rectified linear unit activation function and α_l^{kc} correspond to the pixel-wise gradients for class c and convolutional feature map A^k.

The gradient weights α_l^{kc} can be derived for a particular class c and activation map A^k. Combining Eq. 2 and Eq. 3:

$$Y^c = \sum_k \{ \sum_{l \in \Omega_i} \alpha_l^{kc} . ReLU(\frac{\partial Y^c}{\partial A_l^k}) \} [\sum_{l \in \Omega_i} A_l^k] \tag{3.5}$$

Without a threshold need, the $ReLU$ can be removed from the derivation since there

is no loss of generality. Hence:

$$\frac{\partial^2 Y^c}{\partial (A_l^k)^2} = 2.\alpha_l^{kc}.\frac{\partial^2 Y^c}{\partial (A_l^k)^2} + \sum_{l\in\Omega_i} A_l^k\{\alpha_l^{kc}.\frac{\partial^3 Y^c}{\partial (A_l^k)^3}\} \tag{3.6}$$

To isolate α_l^{kc}, Eq. 6 is rearranged:

$$\alpha_l^{kc} = \frac{\frac{\partial^2 Y^c}{(\partial A_l^k)^2}}{2\frac{\partial^2 Y^c}{(\partial A_l^k)^2} + \sum_{l\in\Omega_i} A_l^k \frac{\partial^3 Y^c}{(\partial A_l^k)^3}} \tag{3.7}$$

The Grad-CAM++ weights are calculated by substituting Eq. 7 in Eq. 4:

$$w_k^c = \sum_{l\in\Omega_i} [\frac{\frac{\partial^2 Y^c}{(\partial A_l^k)^2}}{2\frac{\partial^2 Y^c}{(\partial A_l^k)^2} + \sum_{l\in\Omega_i} A_l^k \frac{\partial^3 Y^c}{(\partial A_l^k)^3}}].relu(\frac{\partial Y^c}{\partial A_l^k}) \tag{3.8}$$

The regional importance map R^i for a DUV Whole Surface Image x_i is computed using a linear combination of forward activation maps.

$$R^i = ReLU(\sum_k w_k^c.A_l^k), \tag{3.9}$$

This highlights the most significant features in the final classification with a positive correlation with pixel intensity and classification score by applying the *ReLU* function to a linear combination of activation maps. In this study, the Grad-CAM++ implementation is applied to a pre-trained DenseNet169 model with ImageNet weights and extracts the feature map as the regional importance map at the Norm5 layer. The architecture of the DenseNet169 model can be visualized in Fig. 2.

3.2.4 Whole Surface Image Decision Fusion

A decision fusion approach achieves the classification of Whole Surface Images. This process takes into account both the classification labels at the patch level, denoted as $y_i^j \in +1,-1$ for all patches $j = 1,\ldots,M_i$, and the regional importance map, R^i. Initially, the regional importance, r_i^j, for each patch p_i^j is computed by averaging the values of R^i over the field of view (FOV) Ω_i^j of the patch.

$$r_i^j = \frac{1}{|\Omega_i^j|} \sum_{l \in \Omega_i^j} R_l^i, \tag{3.10}$$

Here, $|\Omega_i^j|$ represents the total pixel count for each patch, which is 400×400 pixels. A weight, w_i^j, is then assigned to each patch p_i^j, based on the thresholded regional importance value r_i^j.

$$w_i^j = \begin{cases} 0 & \text{if } r_i^j < 0.25 \\ r_i^j & \text{otherwise} \end{cases} \tag{3.11}$$

This weighting strategy disregards patches of low importance, whether malignant or normal/benign, in the final decision for a DUV Whole Surface Image. Subsequently, each patch's value u_i^j is computed by multiplying the patch-level classification label $y_i^j \in +1,-1$ and its corresponding weight w_i^j.

$$u_i^j = w_i^j \cdot y_i^j, \tag{3.12}$$

The total count of malignant patches, H_i, is then calculated,

$$H_i = \sum_{j=1} \begin{cases} 1 & \text{if } u_i^j > 0 \\ \\ 0 & \text{otherwise} \end{cases}, \tag{3.13}$$

Finally, the Whole Surface Image-level classification label $y_i \in -1, +1$ is determined by comparing H_i to a specific percentage q of the total foreground patches M_i.

$$y_i = \begin{cases} +1 & \text{if } H_i > q \cdot M_i, 0 \leq q \leq 1 \\ \\ -1 & \text{otherwise} \end{cases} \tag{3.14}$$

This equation represents the positive (malignant) and negative (benign) values by $+1$ and -1, respectively. This final step ensures that the Whole Surface Image level classification considers the overall distribution of malignant and benign patches within the Whole Surface Image, providing a more accurate and robust classification result.

CHAPTER 4
EXPERIMENT AND RESULTS

4.1 Experimental Setup

The experimental setup for this study involved automated DUV Whole Surface Image classification in discerning between malignant and normal/benign tissue samples. The proposed method was evaluated using a 5-fold cross-validation process. In this process, the DUV Whole Surface Image data was divided while maintaining a balance in class labels. Care was taken to ensure that specific patient samples were either in the training or testing data splits, but not both. The pre-trained ResNet50 model with ImageNet weights, used in the Transfer Learning part of the study, needed to be fine-tuned. Similarly, the pre-trained DenseNet169 model with ImageNet weights, used for the Regional Importance Calculation, underwent no hyperparameter tuning. The XGBoost classifier's hyperparameters were left at their default settings, with no adjustments made. The proposed method was benchmarked against a standard ResNet50 model and a Patch Classification with a Majority Voting approach. The ResNet50 model was trained for 100 epochs on the limited DUV Whole Surface Image data, using a batch size of 4, a learning rate of 0.006, and a dropout rate of 40%. Hyperparameter tuning was performed for the ResNet50 model. The Patch Classification with a Majority Voting approach was derived from the architecture of the proposed method's patch classification and transfer learning portion. This approach was enhanced with a patch majority voting scheme for binary classification between malignant and normal/benign for the Whole Surface Images. The hardware used for this research included an AMD Threadripper 2960x and an NVIDIA 2080ti graphics card. This setup provided the computational power necessary for the resource-intensive tasks of training

and testing the machine learning models. The primary machine learning library used in this study was PyTorch version 2.00.00. PyTorch is a widely-used open-source machine learning library that offers a range of tools and functionalities for developing and training machine learning models. Its flexibility and efficiency made it an ideal choice for this research.

4.2 Visual Inspection

The visual inspection process is a crucial part of the proposed method, providing a qualitative evaluation of the breast cancer location and detection results on various DUV images. This process involves examining the Grad-CAM++ influenced results, highlighted as red (malignant) and green (benign) patches on the DUV images, indicating areas of high importance for the classification task.

Fig. 4.1 presents these results, with the last column showing the DUV images overlaid with the Grad-CAM++ influenced results. H&E images annotated by pathologists are also displayed alongside their corresponding DUV images. This side-by-side comparison directly evaluates the proposed method's performance in identifying and localizing malignant and benign areas in the DUV images.

Upon visual inspection, it is evident that the DUV images exhibit a higher color contrast between malignant (pink/yellow) and normal/benign (light/dark green) tissues compared to the H&E images. This enhanced contrast is a crucial feature of DUV imaging that aids in accurately locating malignant and benign areas in the DUV Whole Surface Images. The Grad-CAM++ regional importance maps further enhance this localization by highlighting the areas of the highest importance for the classification task.

The overlaid regions on the DUV Whole Surface Images, represented by red squares

for malignant tissue and green for normal/benign adipose tissue, visually represent the proposed method's classification results. These overlaid regions demonstrate the effectiveness of the decision fusion method combined with the regional importance maps in localizing the regions of interest (ROI) for the Whole Surface Image level decision. The high degree of prediction accuracy observed in these images underscores the efficacy of the proposed method.

Furthermore, the visual inspection process reveals that the proposed deep learning approach, combined with the patching strategy, can capture key pathological traits such as high cell density and infiltration. This capability is particularly noteworthy given the limited training data used in this study. Overall, the visual inspection process provides a qualitative validation of the proposed method's ability to accurately identify and localize malignant and benign areas in DUV Whole Surface Images, highlighting its potential for real-world clinical applications.

4.3 Classification Performance

The performance of the proposed method for binary classification of normal/benign and malignant DUV Whole Surface Images was evaluated quantitatively using several metrics, including accuracy, sensitivity, specificity, and AUC score. The results of this evaluation, conducted using 5-fold cross-validation across 60 DUV Whole Surface Images, are presented in Fig. 4.1.

The accuracy, sensitivity, and specificity were calculated using the following equations where TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives:

	(1)	(2)	(3)
Accuracy	81.7%	93.3%	95.0%
Sensitivity	91.7%	94.4%	100%
Specificity	66.7%	91.2%	87.5%

Table 4.1: 5-fold Cross-Validation Classification Performance on 60 DUV Whole Surface Images: The proposed method (3) (Patch Classification with Regional Importance) significantly increases the classification accuracy, sensitivity, and specificity compared with the standard ResNet50 approach (1), and Patch Classification with Majority Voting (2).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \tag{4.1}$$

$$Sensitivity = \frac{TP}{TP+FN}, \tag{4.2}$$

$$Specificity = \frac{TN}{TN+FP}, \tag{4.3}$$

The proposed method (Patch Classification with Regional Importance) significantly outperforms the standard ResNet50 approach and Patch Classification with Majority Voting, as shown in the table. Compared to the standard ResNet50, the proposed method improves classification performance with a 13.3% increase in accuracy. The proposed method also outperforms the ResNet50 in sensitivity, specificity, and AUC scores by 8.3%, 20.8%, and 17.0%, respectively.

The ROC Curve of the proposed method, illustrated in Fig. 4.3, shows that the AUC is 17% higher than with the ResNet50 model. Compared with Patch Classification with Majority Voting, the proposed method outperforms by 1.7% and 5.6% for accuracy and sensitivity, respectively. Although the proposed method has a specificity that is 3.7% lower than another approach, it is still a good option for breast cancer classification.

The proposed method achieved a perfect sensitivity rate of 100% compared to the

other approach. Sensitivity is crucial in detecting malignant cases, a primary concern in breast cancer classification. The higher sensitivity rate of the proposed method makes it a valuable option despite having slightly lower specificity. These metrics demonstrate the advantage of the proposed method for intra-operative margin assessment, as it should reduce the likelihood of breast cancer margins being undetected during BCS.

4.4 Misclassified Samples

Three normal/benign DUV Whole Surface Images that were incorrectly classified as malignant are presented in Fig. 4.4. These misclassifications provide valuable insights into the challenges and potential areas of improvement for the proposed method. The top sample in Figure 6 is a mixture of fat and fibrotic breast tissue. The fibrotic tissue's dense, irregular, and bumpy nature likely contributed to its misclassification as malignant. This suggests that the model may need further training to better distinguish between fibrotic and malignant tissue. The middle sample contains a pool of bleeding, which likely led to its misclassification. This indicates that the model may be sensitive to the presence of blood, possibly interpreting it as a sign of malignancy. Future iterations of the model could benefit from additional training data that includes examples of bleeding in benign samples. The bottom sample was classified as normal/benign breast tissue with high fibro glandular density, indicating dense breast tissue. However, it was misclassified as malignant by the model. This suggests that the model may struggle to correctly classify samples with high tissue density, potentially mistaking it for malignancy. It is important to note that other confounding factors may contribute to these misclassifications. Therefore, additional analysis and validation are recommended to confirm these classifications and refine the model.

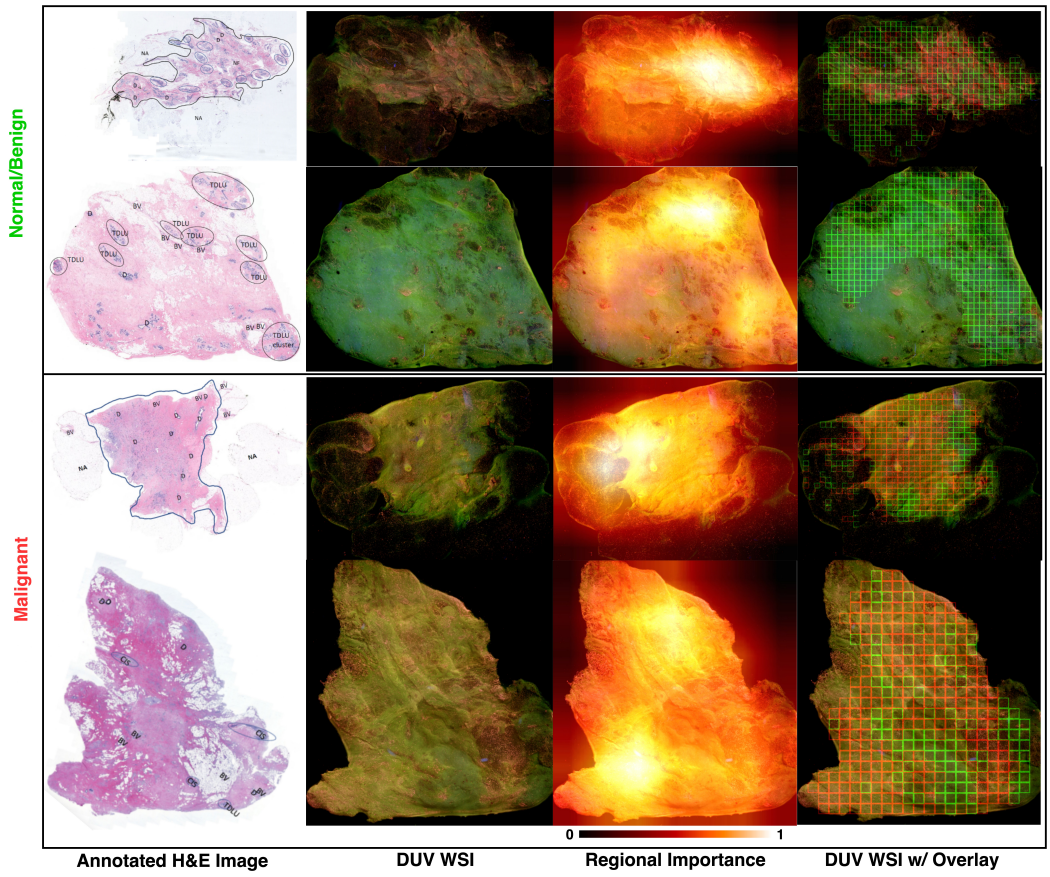

Figure 4.1: DUV Whole Surface Images alongside corresponding H&E images and regional importance maps for malignant and normal/benign samples: The areas of regional importance, depicted in the 3rd column, are characterized by high entropy and rich semantic content. These areas are emphasized to aid in the classification process. It is important to note that these images are fed into a pre-trained DenseNet169 model with ImageNet weights, and the heatmaps, which determine the relevance of tissue areas on a scale from 0 to 1, are extracted using Grad-CAM++. In the DUV Whole Surface Images (shown in the last column), red and green bounding boxes represent malignant and normal/benign patches, which are thresholded based on regional importance. For visualization, the large DUV Whole Surface Images are scaled down.

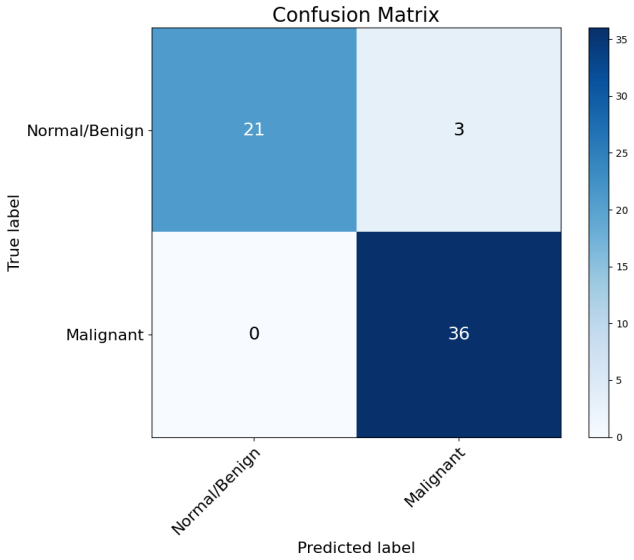

Figure 4.2: Confusion Matrix of the proposed method (Patch Classification with Regional Importance).

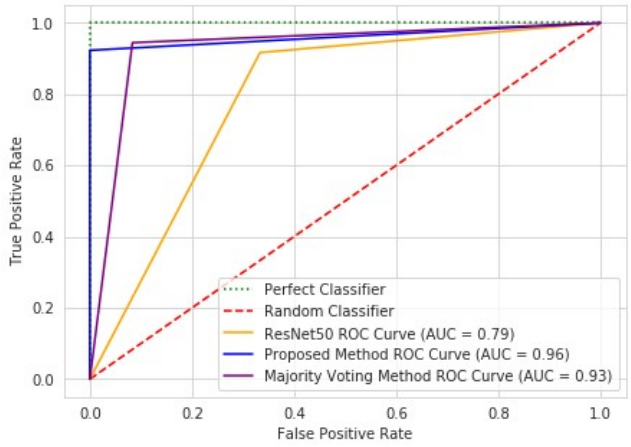

Figure 4.3: ROC graph of the proposed method (Patch Classification with Regional Importance), ResNet50 approach, and Patch Classification with Majority Voting.

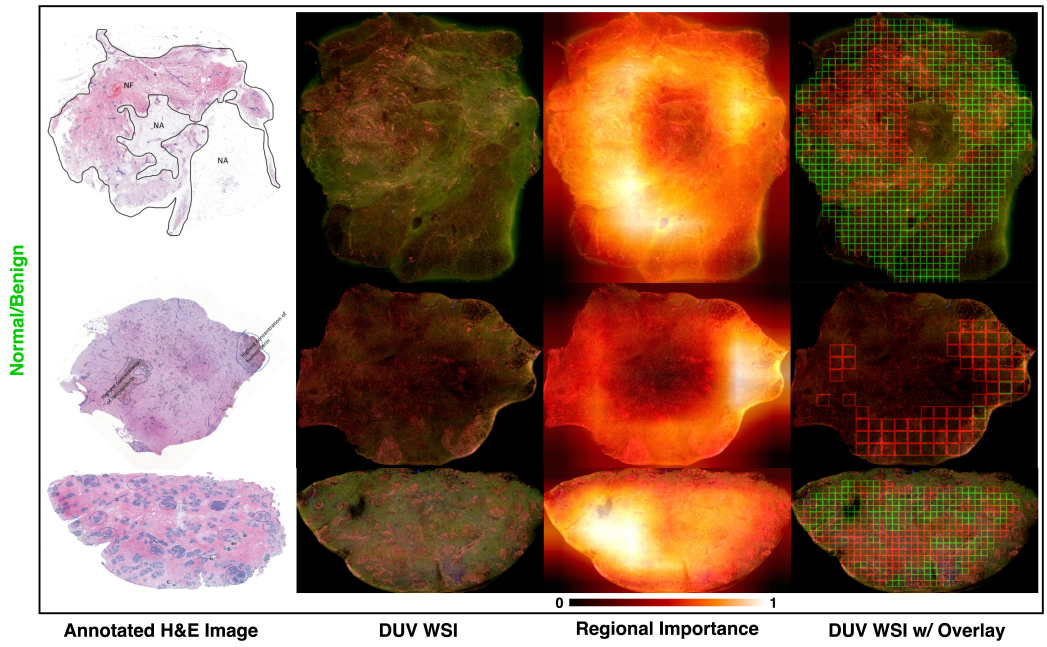

Figure 4.4: Three misclassified DUV Whole Surface Images normal/benign samples from the proposed approach. H&E image and regional importance map are shown for each sample, respectively. Large DUV Whole Surface Images are scaled down for visualization.

CHAPTER 5
DISCUSSION

5.1 Methodology and Results

The proposed method for breast cancer classification in DUV image margin assessment demonstrated impressive performance, achieving an accuracy of 95% in determining DUV Whole Surface Images. Notably, it displayed a perfect sensitivity score of 100% in detecting malignant cases. This high sensitivity is crucial in breast cancer surgery, where missing malignant cases can have severe consequences. Another significant strength is the method's ability to accurately localize malignant and normal/benign tissue areas. This localization is achieved through the use of regional importance maps generated by the Grad-CAM++ algorithm. These maps highlight the most critical regions in the image responsible for the classification decision, focusing on areas with the highest level of detail and semantic information within the DUV Whole Surface Image. This approach allows the model to concentrate on the most relevant image regions that could contain malignant or benign tissue, enhancing the accuracy of the classification. The proposed method offers several advantages compared to standard deep-learning classification methods on DUV breast surgical samples. It combines patch-level classification using a transfer learning approach with regional importance maps, providing a more robust and accurate representation of the input data. This ensemble learning-based approach effectively utilizes the limited dataset available while minimizing overfitting, a critical aspect when dealing with a small dataset like the one used in this study. The proposed method also simplifies the task for surgeons by focusing on binary classification, assigning a malignant or normal/benign label to each Whole Surface Image. This streamlined decision-making

process is tailored for margin assessment during breast cancer surgery, making it easier for surgeons to determine whether the margins are clear or require further resection. In contrast, standard deep-learning methods, such as the ResNet50-based approach, classify images into multiple tissue types. While this provides valuable information regarding the tissue types present, it may not directly apply to margin assessment in the same way as the proposed method, as it requires additional interpretation to decide the margin status. Overall, the proposed method demonstrates the effectiveness of deep learning in accurately identifying cancerous regions in DUV Whole Surface Images. Utilizing transfer learning, a robust gradient-boosting algorithm, and regional importance maps paves the way for further research and development in this area, potentially contributing to improved surgical outcomes and reduced breast cancer re-excision rates after BCS.

5.2 Limitations and Challenges

Despite the promising results, this study has its limitations. One of the main challenges was the relatively small size of the dataset used, which consisted of only 60 images available. This small sample size may limit the generalizability of the findings, as the model's performance on more extensive or diverse datasets still needs to be investigated. Training a robust and accurate model on such a small dataset can be challenging, requiring careful selection of training parameters and model architecture to prevent overfitting.

Another limitation lies in the complexity of the tissue samples. Many samples contained mixed tissue types, meaning that the label assigned to each patch may only represent a portion of the area within the patch. This complexity could complicate the classification process. Furthermore, determining the optimal patch size remains an open question. Better classification results could be achieved by identifying and using an optimal patch size that

maximizes the model's discriminatory power.

The dataset used in this study mostly contained samples that were either predominantly benign/normal or malignant. Six samples exhibiting a complex mixture of benign/normal and malignant tissues were disregarded for the sake of model training. While this approach allowed the model to be trained on clean and pure samples, it may limit the generalizability of the proposed method to samples with a greater complexity of benign/normal and malignant tissues.

Consequently, further investigation of the model's performance on more complex and diverse samples is essential. This will help validate the proposed method's efficacy and robustness in real-world scenarios, ensuring it can accurately classify a wide range of tissue samples. Despite these limitations, the study contributes to breast cancer classification, demonstrating the potential of deep learning techniques in improving surgical outcomes.

CHAPTER 6
CONCLUSIONS

6.1 Summary of Findings and Contributions

This study presented an innovative method for accurately assessing margins during breast cancer surgery, leveraging the power of deep learning techniques. The proposed methodology combined patch-level classification using transfer learning with regional importance maps generated through Grad-CAM++. This approach focused on the most significant regions within DUV Whole Surface Images, assigning a malignant or normal/benign label to each Whole Surface Image with increased confidence. The method was evaluated on a dataset of 60 authentic DUV Whole Surface Images, achieving an impressive classification accuracy of 95.0%. This result underscores the potential of the proposed method for real-time assessment of margin status during breast cancer surgery and potentially other types of cancer surgeries. The study's contributions lie in its novel approach to breast cancer classification in DUV Whole Surface Images, enhancing the robustness of intra-operative margin assessment.

6.2 Implications for Clinical Practice

The proposed method has significant potential to impact clinical decision-making. Providing real-time guidance to surgeons during operations can ensure more informed decisions about resection margins, leading to more accurate surgical interventions and potentially reducing the need for re-excision. Furthermore, the method can aid in developing personalized treatment plans and tailoring surgical strategies to individual patients' needs. The potential for integrating the proposed method into surgical navigation systems is particularly promising. This integration could allow for real-time visualization of can-

cerous regions and their boundaries, improving surgical precision and minimizing damage to healthy tissues.

6.3 Future Work

Despite the promising results, there is room for further refinement and validation of the proposed method. Future work could expand the dataset to enhance the model's generalizability, optimize patch size to improve classification performance, and investigate alternative evaluation metrics to compare the proposed method more comprehensively with other techniques. Additionally, assessing the clinical impact of the proposed method will be crucial to understand its potential to guide surgical decision-making and reduce the need for re-excision.

6.4 Final Remarks

This study demonstrates the effectiveness of deep learning techniques in breast cancer margin assessment using DUV Whole Surface Images. The proposed method, with its potential clinical applications and possibilities for further refinement, offers a promising direction for future research and development in medical imaging. By integrating this method into surgical navigation systems, we can improve the precision of surgical interventions and minimize damage to healthy tissues. The proposed method's versatility and applicability to different diagnosis techniques could contribute to better patient outcomes, underscoring the value of deep learning in medical imaging and healthcare.

BIBLIOGRAPHY

[1] Adrienne G Waks and Eric P Winer. Breast cancer treatment: a review. *Jama*, 321(3):288–300, 2019.

[2] Taylor D Ellington, Jacqueline W Miller, and Henley. Trends in breast cancer incidence, by race, ethnicity, and age among women aged ≥20 years - united states, 1999-2018. *MMWR Morb Mortal Wkly Rep*, 71(2):43–47, January 2022.

[3] Kristy L. Kummerow, Liping Du, and David F. Penson. Nationwide trends in mastectomy for early-stage breast cancer. *JAMA Surgery*, 150(1):9–16, Jan 2015.

[4] Olga Kantor, Catherine Pesce, and Katherine Kopkash. Impact of the society of surgical oncology-american society for radiation oncology margin guidelines on breast-conserving surgery and mastectomy trends. *Journal of the American College of Surgeons*, 229(1):104–114, Jul 2019.

[5] Tongtong Lu, Julie M Jorns, Mollie Patton, Renee Fisher, Amanda Emmrich, Todd Doehring, Taly Gilat Schmidt, DongHye Ye, Tina Yen, and Bing Yu. Rapid assessment of breast tumor margins using deep ultraviolet fluorescence scanning microscopy. *Journal of Biomedical Optics*, 25(12):126501, 2020.

[6] Tongtong Lu, Julie M. Jorns, and Dong Hye Ye. Automated assessment of breast margins in deep ultraviolet fluorescence images using texture analysis. *Biomedical Optics Express*, 13(9):5015–5034, Sep 2022.

[7] Farnaz H Foomani, DM Anisuzzaman, Jeffrey Niezgoda, Jonathan Niezgoda, William Guns, Sandeep Gopalakrishnan, and Zeyun Yu. Synthesizing time-series wound prognosis factors from electronic medical records using generative adversarial networks. *Journal of biomedical informatics*, 125:103972, 2022.

[8] Yabo Fu, Yang Lei, Tonghe Wang, Walter J Curran, Tian Liu, and Xiaofeng Yang. Deep learning in medical image registration: a review. *Physics in Medicine & Biology*, 65(20):20TR01, 2020.

[9] Yu-Dong Zhang, Suresh Chandra Satapathy, David S Guttery, Juan Manuel Górriz, and Shui-Hua Wang. Improved breast cancer classification through combining graph convolutional network and convolutional neural network. *Information Processing & Management*, 58(2):102439, 2021.

[10] Rishav Singh, Tanveer Ahmed, Abhinav Kumar, Amit Kumar Singh, Anil Kumar Pandey, and Sanjay Kumar Singh. Imbalanced breast cancer classification using transfer learning. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(1):83–93, 2020.

[11] Paul Gamble, Ronnachai Jaroensri, and Hongwu Wang. Determining breast cancer biomarker status and associated morphological features using deep learning. *Communications Medicine*, 1(1):14, Jul 2021.

[12] Mustafa I. Jaber, Bing Song, and Clive Taylor. A deep learning image-based intrinsic molecular subtype classifier of breast tumors reveals tumor heterogeneity that may affect survival. *Breast Cancer Research*, 22(1):12, Jan 2020.

[13] Yuqian Li, Junmin Wu, and Qisong Wu. Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning. *IEEE Access*, 7:21400–21408, 2019.

[14] Saman Farahmand, Aileen I. Fernandez, and Fahad Shabbir Ahmed. Deep learning trained on hematoxylin and eosin tumor region of interest predicts her2 status and trastuzumab treatment response in her2+ breast cancer. *Modern Pathology*, 35(1):44–51, Jan 2022.

[15] Heather D. Couture, Lindsay A. Williams, and Joseph Geradts. Image analysis with deep learning to predict breast cancer grade, er status, histologic subtype, and intrinsic subtype. *npj Breast Cancer*, 4(1):30, Sep 2018.

[16] Jiandong Ye, Yihao Luo, Chuang Zhu, Fang Liu, and Yue Zhang. Breast cancer image classification on wsi with spatial correlations. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1219–1223. IEEE, 2019.

[17] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *CoRR*, abs/1511.08458, 2015.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[19] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016.

[20] Farzad Fereidouni, Zachary T Harmany, Miao Tian, Austin Todd, John A Kintner, John D McPherson, Alexander D Borowsky, John Bishop, Mirna Lechpammer, Stavros G Demos, et al. Microscopy with ultraviolet surface excitation for rapid slide-free histology. *Nature biomedical engineering*, 1(12):957–966, 2017.

[21] Tadayuki Yoshitake, Michael G Giacomelli, Liza M Quintana, Hilde Vardeh, Lucas C Cahill, Beverly E Faulkner-Jones, James L Connolly, Daihung Do, and James G Fujimoto. Rapid histopathological imaging of skin and breast cancer surgical specimens using immersion microscopy with ultraviolet surface excitation. *Scientific reports*, 8(1):4476, 2018.

[22] Weisi Xie, Ye Chen, Yu Wang, Linpeng Wei, Chengbo Yin, Adam K Glaser, Mark E Fauver, Eric J Seibel, Suzanne M Dintzis, Joshua C Vaughan, et al. Microscopy with ultraviolet surface excitation for wide-area pathology of breast surgical margins. *Journal of biomedical optics*, 24(2):026501–026501, 2019.

[23] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.

[24] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.

[25] Aditya Chattopadhyay, Anirban Sarkar, Prantik Howlader, and Vineeth N. Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. *CoRR*, abs/1710.11063, 2017.

[26] Arash Shokouhmand, Nicole D Aranoff, Elissa Driggin, Philip Green, and Negar Tavassolian. Efficient detection of aortic stenosis using morphological characteristics of cardiomechanical signals and heart rate variability parameters. *Scientific reports*, 11(1):1–14, 2021.