NUST COLLEGE OF

ELECTRICAL AND MECHANICAL ENGINEERING

Derma Vision - A Cloud-Based Application For Visualizing And Analyzing Whole Slide Skin Images

A PROJECT REPORT

DE-41 (DC&SE)

Submitted by

PC Tayyab Rashid

NS Tazeen Hina

NS Hamid Rasul

BACHELORS

IN

COMPUTER ENGINEERING

YEAR

2023

PROJECT SUPERVISOR

Dr. Usman Akram

Dr. SAJID GUL KHAWAJA

AP. Jehanzeb

COLLEGE OF

ELECTRICAL AND MECHANICAL ENGINEERING

PESHAWAR ROAD, RAWALPINDI

Acknowledgements

Abstract

Histopathology, an essential field in medical research, involves the examination of cells and tissues to understand various diseases and conditions. To aid in this study, tissue samples are traditionally stained to enhance color and contrast, enabling detailed analysis. However, staining procedures are costly, time-consuming, and prone to introducing inconsistencies. To address these challenges, we present an innovative cloud-based application that leverages deep learning techniques to analyze skin tissues and provide valuable insights to dermopathologists. Our solution revolves around the concept of virtual staining, achieved through the utilization of generative networks. By employing advanced algorithms, we replicate the staining process digitally, eliminating the need for physical staining and the associated drawbacks. This approach not only significantly reduces costs but also accelerates the analysis process, allowing for quicker diagnosis and treatment decisions. Moreover, our application incorporates cutting-edge techniques such as semantic segmentation with uncertainty maps to accurately classify skin layers and cells into 12 distinct dermatological classes. This sophisticated methodology enables precise identification and differentiation of various tissue components, aiding in the diagnosis of complex skin conditions and diseases. The development of this project has been a collaborative effort with Nidi Skin, a renowned institution in the field of dermatology based in the United States. By combining their expertise with our technical advancements, we have created a robust and efficient tool that has the potential to revolutionize histopathological analysis in dermatology. In conclusion, our cloud-based application, powered by deep learning algorithms, provides a cost-effective, time-saving, and consistent alternative to traditional staining techniques. With virtual staining and advanced segmentation capabilities, it offers dermopathologists detailed insights into skin tissues, empowering them to make accurate diagnoses and informed treatment decisions.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Skin cancer is a prevalent disease that affects millions of people worldwide. According to the American Academy of Dermatology, skin cancer is the most common cancer in the United States, with one in five Americans developing it in their lifetime. Dermatologists use their specialized experience to diagnose skin cancer based on the results of pathological tests from skin biopsy and sonography imaging of the skin tissue. Biopsy on whole slide skin images is a traditional method used for skin cancer detection. It is a painful and invasive method that requires laboratory testing, which is not very efficient and time-consuming to detect skin lesions.

Early detection, appropriate treatment, and prevention of recurrence are major challenges for researchers today. Determining precise skin disease on brightfield images using manual technique requires considerable time, complex screening and could be erroneous, whereas automated methods face many other challenges like presence of hair, inconspicuous lesion margins, low contrast on dermoscopic images, and variability in skin lesion color, texture, and shape. Advanced computation and optimized code can be used to extract some meaningful information from brightfield images that may not be readily perceived by humans. The tissue

undergoes investigation after the process of staining, where different anomalies are investigated. This investigation is again influenced by the type of processing it has received, leaving a lot of ambiguities even after huge laborious work. It has been highlighted that due to these challenges, pathologists disagree on up to 60 percent of cases. Research in related computations and segmentation algorithms to help in quick and uniformed outcome for easy diagnosis of different anomalies present in the tissue is need of the hour.

This thesis report aims to investigate the use of staining and segmentation techniques on WSIs of skin biopsies for the automated diagnosis of skin cancer. The study will explore the effectiveness of different staining techniques, especially Hematoxylin and Eosin (HE) staining, in enhancing image contrast. The report will also evaluate the performance of various segmentation algorithms, and deep learning-based approaches, in accurately identifying cancerous regions in WSIs.

1.1 Scope

Data has emerged as a valuable resource in various domains, including healthcare. The volume of data being generated has surpassed the capacity of current systems to handle it effectively. In the field of healthcare, there is a vast amount of data available, as Worldwide digital healthcare was estimated to be equal to 500 petabytes and was expected to reach 25 exabytes in 2020. Hughes has also predicted that the global growth of healthcare data will be between 1.2 and 2.4 exabytes a year. This wealth of medical information can be leveraged for various purposes, such as patient follow-up and advice. However, to fully harness the potential of this information, advanced data analysis techniques such as machine learning, deep learning and statistics are necessary. Machine learning, a branch of artificial intelligence, focuses on enabling computers to learn from data without explicit programming. Deep learning, a subfield of machine learning, em-

ploys complex algorithms inspired by human neurons, enabling the processing of unstructured data, including images, signals, and text.Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) were the pioneering architectures used to process images and text, respectively. This research aims to leverage the power of deep learning techniques to improve the facilities for diagnosing skin cancer through automated, accurate, and reliable segmentation and staining of skin cancer lesions. The proposed methods have the potential to improve the accuracy and efficiency of skin cancer diagnosis aiding healthcare professionals such as histopathologists and dermatologists in making informed decisions and ultimately improving patient outcomes.

1.2 Motivation

Skin cancer, including melanoma and non-melanoma skin cancer (NMSC), is a significant public health problem globally, with increasing incidence rates each year. Melanoma is the deadliest form of skin cancer, with a high mortality rate, and early detection is the most promising means of decreasing morbidity and mortality. Non-melanoma skin cancer (NMSC) has 12 classes, including basal cell carcinoma (BCC) and squamous cell carcinoma (SCC), represents the most common type of malignancy in the white population. The high treatment cost of skin cancer causes an economic burden of disease in the country and negatively impacts public health issues. BCC and SCC represent the most frequently observed malignancy among Caucasians, with BCC accounting for approximately 75-80% of these malignancies in individuals with fair skin. In the United States, approximately 5.4 million cases of non-melanoma skin cancers were treated in 2012, with skin cancer accounting for more than 40% of all malignancies. The incidence of non-melanoma skin cancers has been rising over the past three decades. There are different subtypes of melanoma, including cutaneous melanoma, uveal

melanoma, and mucosal melanoma. In a study conducted in North Sumatera, squamous cell carcinoma was found to be the most common form of skin cancer. The authors of another study reported that BCC is a slow-progressing, non-melanocytic cancer arising from basal cells, while CSCC arises from keratinocytes.

Skin cancer is prevalent in Pakistan, particularly in the southwestern region. A retrospective study conducted on skin cancer cases in the Centre for Nuclear Medicine and Radiotherapy (CENAR) in Quetta found that skin cancer was the second most prevalent category of cancer in the area, with a prevalence of 12.5% of total cancer patients. The study also found that skin cancer was more prevalent in males than females, with squamous cell carcinoma(SCC) being the most prevalent category of skin cancer. Another study conducted in Saudi Arabia found that skin cancer was uncommon but not rare, with basal cell carcinoma and squamous cell carcinoma being the most common types of nonmelanoma skin cancer (NMSC). The study also found that the head and neck was the most common location for skin cancer, and skin cancers in individuals younger than 50 years of age require more careful evaluation of possible risk factors.

Early diagnosis of melanoma lesions is crucial for successful treatment, as it can spread to other parts of the body if not caught and treated early. Whole slide images (WSIs) are prepared for skin cancer diagnosis by scanning glass slides containing tissue samples at high resolution to create a digital image of the entire slide. This involves staining tissue samples with dyes to enhance the visibility of specific cellular structures and identify abnormalities. This process is time-consuming and labor-intensive, as it involves manually staining the slides and scanning them. Skin cancer specialists are required for the current diagnostic process, which can be expensive and not easily available in developing countries. However, deep learning models are being developed to replace this tedious process. These models use artificial intelligence to analyze the digital images and

identify cancerous cells with high accuracy. The use of WSI and deep learning models has the potential to improve the efficiency and accuracy of skin cancer diagnosis. This has led to the development of automated diagnosis systems using artificial intelligence (AI)-based methods, such as machine learning (ML) and deep learning (DL). DL has been integrated into skin cancer diagnosis in recent years, with dermoscopic images being at the center of this revolution. Deep learning models have been explored for skin Whole Slide Skin Images staining, segmentation and classification, including for the detection of melanoma skin cancer. Many ML and DL methods show high performance as classifiers of skin lesions, which bodes well for the inclusion of these techniques in clinical practice.

1.3 Objectives

The purpose of this research study is to develop a cloud-based application for staining and segmentation of whole slide skin images using deep learning algorithms. The objective is to provide an efficient and cost-effective tool for identifying cancerous regions in skin slide images, aiding in early detection and treatment planning. This includes:

Digitally Stain and Segment Skin Whole Slide Images

Design and develop a deep learning-based framework to enhance the visibility of cellular structures and accurately delineate different tissue components within skin whole slide images.

Identify Cancerous Regions

Utilize the digitally stained and segmented images to automatically identify and highlight potential cancerous regions, assisting pathologists in early detection and treatment planning.

Manage Data in a Cloud-based Application

Create a comprehensive cloud-based application for uploading, storing, and organizing skin slide images, annotated regions, and associated metadata. This facilitates collaboration and provides a centralized platform for efficient data management in the digital pathology workflow.

1.4   Problem Statement

The early diagnosis of melanoma and non melanoma is vital for effective treatment, as the disease can rapidly spread to other parts of the body if not detected and treated promptly. The current diagnostic process for skin cancer relies on whole slide images (WSIs), which are created by scanning glass slides containing tissue samples at high resolution to generate a digital representation of the entire slide. However, this process presents several challenges.

Firstly, the preparation of WSIs involves manually staining the tissue samples with dyes to enhance the visibility of cellular structures and identify abnormalities. This staining process is time-consuming, labor-intensive, and requires skilled technicians. The dependency on manual staining introduces the potential for human error and inconsistency in the quality of staining, which can impact the accuracy of diagnosis.

Secondly, the scanning of stained slides to create WSIs is also a time-consuming task. This process requires specialized equipment and expertise, making it costly and often inaccessible in resource-limited settings, particularly in developing countries. The lack of accessibility to scanning facilities and skilled skin cancer specialists hinders early detection and diagnosis, potentially compromising patient outcomes.

Moreover, the current diagnostic process heavily relies on the expertise of skin

cancer specialists, who are in high demand and may not be easily accessible in certain regions. This dependency on specialized professionals further limits the availability and affordability of accurate and timely skin cancer diagnosis, particularly in underserved areas.

Therefore, there is a pressing need for a solution that addresses these challenges and provides a more efficient and accessible approach to staining, scanning, and diagnosing skin cancer using whole slide images. A cloud-based application that incorporates deep learning algorithms for automated staining and segmentation, coupled with efficient data management and analysis tools, can potentially revolutionize the field of digital pathology and enable early and accurate diagnosis of skin cancer in a cost-effective manner.

1.5 Contributions

To achieve the above-mentioned objectives of this research study, the following contributions are made.

1. A novel modified Generative Adversarial Network (GAN) architecture specifically designed for digital staining of whole slide skin images by digitally replicating the staining process, enabling enhanced visibility of cellular structures.

2. Developed a reliable deep learning model for accurate segmentation and classification of skin tissue regions within whole slide images.

3. To address the challenges of time-consuming processing and limited accessibility, a cloud-based web application is developed as part of this thesis. The application leverages the power of cloud computing to accelerate the processing of whole slide skin images, enabling faster staining, segmentation, and analysis.

1.6   Structure of Thesis

The structure of thesis is as follows: Chapter 2 presents a brief overview about the different skin diseases and also gives an idea about the individual skin layers. Chapter 3 in this thesis highlights the existing work in this domain and recognizes the gaps in the said references. It also lists down the contributions that our work has made to the said problem. Chapter 4 lists down the different datasets we used, along with explaining any pre processing that was done.

Chapter 2

Skin Anatomy and Diagnostic Methods:

Skin disorders, particularly skin cancers like Melanoma, and non-melanoma variations such as Basal Cell Carcinoma (BCC), Intraepidermal Carcinoma (IEC), Squamous Cell Carcinoma (SCC), and inflammatory conditions, pose significant health challenges globally. These ailments are often linked to various factors such as environmental influences, occupational risks, and personal behaviors related to skin maintenance. Regrettably, skin conditions, especially skin cancers, have historically received comparatively less attention and research funding compared to diseases like cardiovascular disorders, cancer, strokes, diabetes, and Alzheimer's disease. (Add reference) Therefore, it is crucial to acquire a comprehensive understanding of the occurrence, impact, and mortality rates associated with skin disorders, both on a global scale and within specific regions. Such knowledge will facilitate informed strategies for prevention, early detection, treatment, and focused research endeavors in the realm of dermatology. This section offers an overview of the structure and composition of the skin, shedding light on its intricate anatomy. Moreover, it explores diverse categories of skin disorders, with

particular emphasis on skin cancers like Melanoma, and non-melanoma variations such as BCC, IEC, SCC, and inflammatory conditions. The discussion encompasses the causes, effects, and implications of these ailments on overall well-being. Furthermore, it briefly touches upon various diagnostic methods and imaging techniques utilized in dermatology, as well as approaches for documenting and communicating findings within the field.

2.1 Skin Anatomy

The skin, the largest organ of the human body, is composed of various layers and structures that work together to provide protection, sensation, and regulation of body temperature. In this section, we will discuss different skin layers and the diseases associated with them.
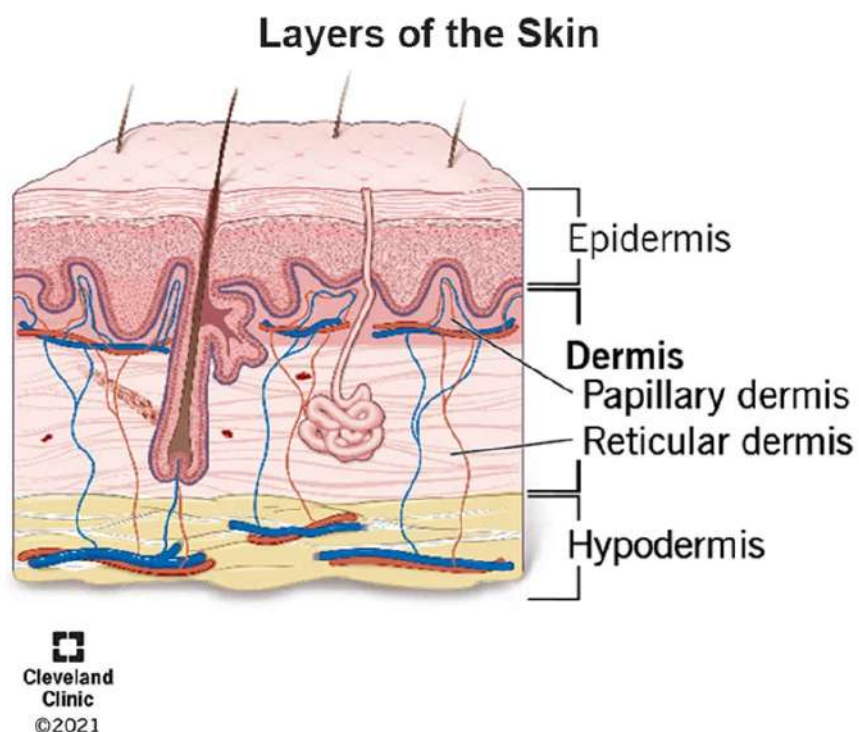
Figure 2.1: Diagram showing individual skin layers.

2.1.1 Epidermis

The epidermis is the outermost layer of the skin, providing a protective barrier against external factors. It comprises several sublayers, including the stratum corneum, stratum granulosum, stratum spinosum, and basal layer. Its primary function is to shield the underlying tissues and organs from potential harm.
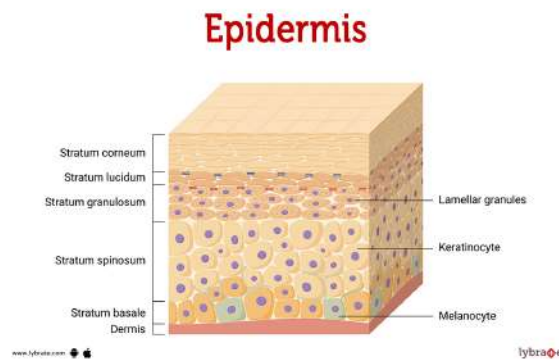
Figure 2.2: Epidermis, the outermost layer.

2.1.2 Glands

Within the skin, there are glands responsible for maintaining skin health. These include sebaceous glands, which produce sebum for skin and hair lubrication, and sweat glands, aiding in body temperature regulation through sweat secretion.

2.1.3 Reticular Dermis

Situated beneath the epidermis, the reticular dermis is composed of dense connective tissue. It provides structural support, strength, and houses blood vessels, nerves, and lymphatic vessels.

2.1.4 Hair Follicles

Hair follicles are distributed throughout the skin and are responsible for hair production. Each follicle consists of a hair shaft and a root extending into the
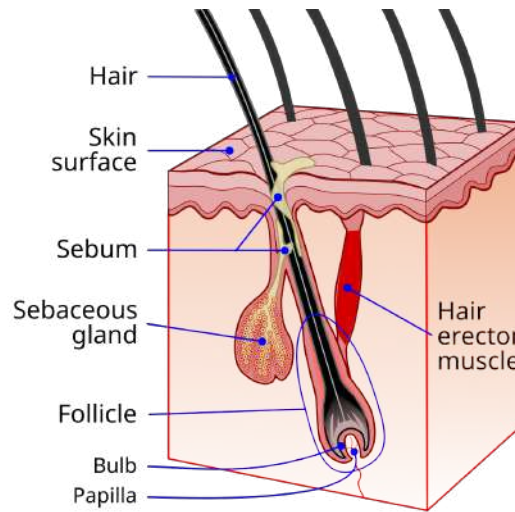
Figure 2.3: Glands.

dermis, connecting to sebaceous glands.

2.1.5 Papillary Dermis

The papillary dermis, located in the uppermost layer beneath the epidermis, contains papillae that form ridges on the skin's surface. These ridges contribute to fingerprint patterns and enhance tactile sensitivity.

2.1.6 Hypodermis

The hypodermis, or subcutaneous tissue, resides beneath the dermis. Composed of adipocytes (fat cells), it provides insulation, cushioning, and energy storage. The hypodermis also contains blood vessels and nerves.

2.2 Staining Techniques in Skin Biopsies and Whole Slide Images

Biomedical imaging has become a common practice in the field of dermatology, offering non-invasive methods for diagnosing skin cancer and aiding in histopatho-

logical evaluations. These imaging techniques assist doctors, dermatologists, and pathologists in accurately identifying skin diseases at early stages, enabling timely interventions and improving patient outcomes. The section explains skin biopsies and the process of staining and conversion to whole slide images.

2.2.1 Skin Biopsy

Skin biopsy is a medical procedure in which a small sample of skin tissue is collected for diagnostic evaluation. It is often performed when there is a suspicious skin lesion, rash, or abnormality that requires further examination. The need for a skin biopsy arises from the importance of obtaining a precise diagnosis in dermatological conditions. Through a skin biopsy, dermatologists and pathologists can assess the microscopic features of the skin tissue, such as cellular morphology, architecture, and the presence of any abnormalities. This information is crucial for determining the nature of skin diseases, distinguishing between benign and malignant lesions, and guiding appropriate treatment decisions. By providing a direct examination of the affected area at a cellular level, skin biopsies play a vital role in accurate diagnosis, prognosis, and management of various skin conditions, including skin cancers, inflammatory disorders, autoimmune diseases, and infections. The samples obtained from biopsies are unstained and the use of various stains in skin biopsies enables a comprehensive evaluation of tissue samples and aids in the accurate diagnosis and classification of skin diseases.

2.2.2 Staining

Various staining techniques are employed in skin biopsies to enhance the visualization and analysis of tissue samples. Staining plays a crucial role in highlighting specific cellular components and structures, aiding dermatologists and pathologists in identifying and characterizing different types of skin lesions and abnormalities. Some common stains used in skin biopsies include Hematoxylin

and Eosin (H&E), which provides general information about tissue architecture and cellular morphology. Additionally, immunohistochemical stains are utilized to detect specific proteins or markers, helping in the classification and subtyping of skin cancers. Examples of immunohistochemical stains used in skin biopsies include S100, Melan-A, and Mart-1 for melanoma, and AE1/AE3 and p40 for squamous cell carcinoma. Special stains such as Periodic Acid-Schiff (PAS) and Gram stain may be employed to identify specific microorganisms or evaluate specific features in skin samples. Though H&E is the most popular stain around the world and the most common.

H&E Staining

Hematoxylin and Eosin (H&E) is the most commonly used stain in histopathology. Hematoxylin stains nuclei blue-purple, providing information about cellular architecture and organization. Eosin stains cytoplasm and extracellular components pink, allowing for the evaluation of cellular morphology and tissue structure. H&E staining provides essential insights into tissue composition and helps in the initial assessment of various skin diseases.
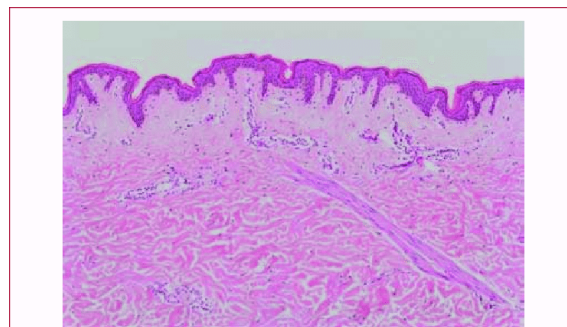
Figure 2.4: Diagram showing H&E Staining.

Immunohistochemical Stains

Immunohistochemical stains involve the use of specific antibodies to detect proteins or markers in skin tissue samples. They aid in the classification and subtyp-

ing of skin cancers. Examples of immunohistochemical stains include:

- **S100**: A marker used to identify melanocytes, aiding in the diagnosis of melanoma.

- **Melan-A and Mart-1**: Stains that help in confirming the presence of melanoma cells.

- **AE1/AE3**: Antibodies used to detect cytokeratins, which are found in squamous cell carcinoma.

- **p40**: A marker used to differentiate squamous cell carcinoma from other types of carcinoma.
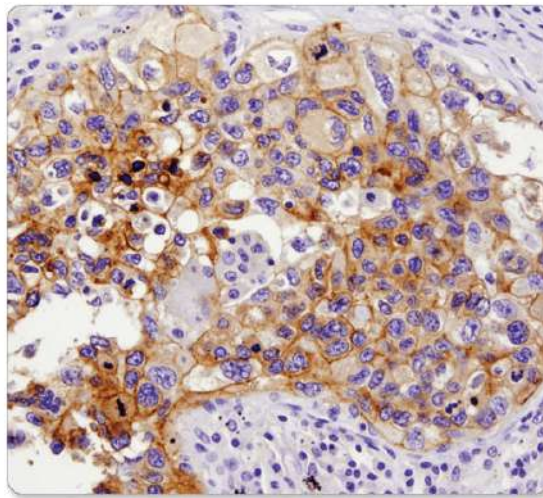
Figure 2.5: Diagram showing Immunohistochemical Stain.

Special Stains

Special stains are employed for specific purposes in skin biopsies. Some examples include:

- **Periodic Acid-Schiff (PAS)**: This stain is used to identify glycogen, basement membranes, and certain fungal organisms in skin samples.

- **Gram Stain**: Gram staining is utilized to identify and classify bacteria based

on their cell wall composition. It can help identify bacterial infections or evaluate specific features in skin samples.
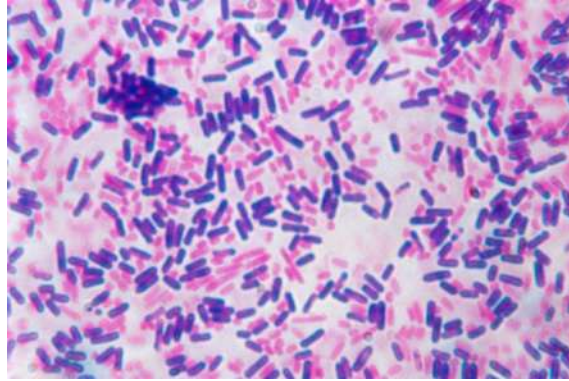
Figure 2.6: Diagram showing Gram Stain.

2.3 Skin Diseases

Skin diseases such as those mentioned below necessitate prompt medical attention and appropriate management, such as surgical interventions, chemotherapy, radiation therapy, immunotherapy, or targeted therapies. Early detection through regular skin screenings is crucial for improving outcomes and reducing the risk of complications associated with these conditions.

2.3.1 Inflammation

Inflammation is a vital response of the skin to injury or infection. It involves a complex cascade of immune cells and chemical mediators, protecting against pathogens and facilitating the healing process. The severity and duration of inflammatory responses vary depending on the underlying cause.

2.3.2 Basal Cell Carcinoma (BCC)

BCC is the most common form of skin cancer, originating from basal cells in the epidermis. It typically manifests as a small, pearly bump or a waxy, translucent
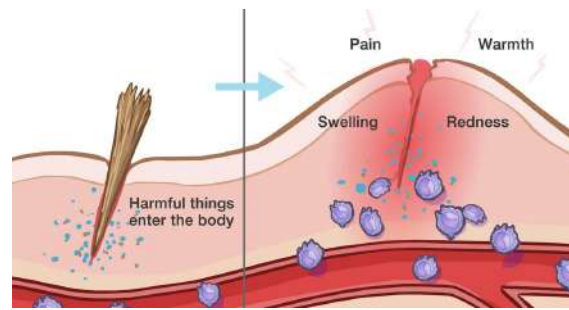
Figure 2.7: Diagram showing Inflammation.

lesion. Although slow-growing and rarely metastasizing, BCC can cause local tissue destruction if left untreated.
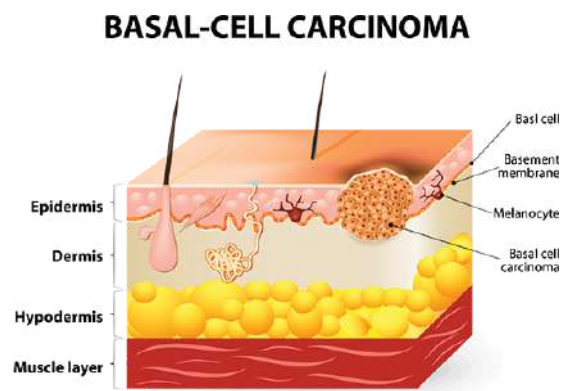
Figure 2.8: Diagram showing Basal Cell Carcinoma (BCC).

2.3.3 Intraepidermal Carcinoma

Also referred to as Bowen's disease or squamous cell carcinoma in situ, this condition affects the upper layers of the epidermis. It appears as red, scaly patches that may be itchy or tender. If untreated, it has the potential to progress to invasive squamous cell carcinoma.

2.3.4 Squamous Cell Carcinoma (SCC)

SCC arises from squamous cells in the epidermis and can develop in various body areas, including the skin. It typically presents as a firm, red nodule or a flat,

scaly lesion. Compared to basal cell carcinoma, SCC has a higher propensity for metastasis.
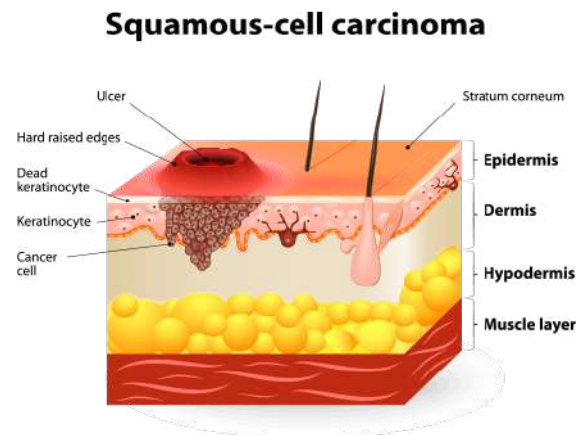
Figure 2.9: Diagram showing Squamous Cell Carcinoma (SCC).

2.3.5 Melanoma

Melanoma originates in the melanocytes, the pigment-producing cells of the skin. It often appears as an irregularly shaped mole or a dark lesion with uneven borders. Melanoma has a high potential for rapid metastasis and is considered the most dangerous type of skin cancer.
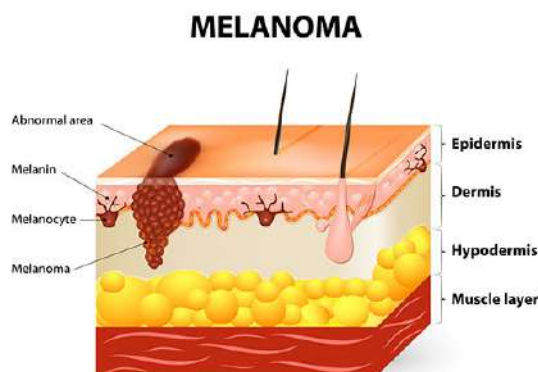
Figure 2.10: Diagram showing Melanoma Skin cancer.

Chapter 3

Literature Review

3.0.1 Existing Work

Early detection of skin cancer can increase five year survival rate of patients from 18 - 98 percent. Automated computerized solutions of skin disease consists segmentation, feature extraction, and its classification. Deep fully convolutional networks have achieved significant success in the task of semantic segmentation [4]. Segmentation has its own importance as feature extraction and classification rely on this part [5]. Adi Wibowo et al. [6] presented a light weight encoder-decoder for segmentation by treating variability with the augmentation. The method shows efficiency in terms of computation. Accuracy in image segmentation is very critical to diagnosis, hence putting more pressure on the performance of the model, and it is always challenging as images have heterogeneous texture, color and shape with fuzzy boundaries. UNet is a deep learning architecture proposed for semantic segmentation tasks, which aims to accurately classify each pixel in an image to a certain category. It was introduced by Ronneberger et al. in 2015 [2]. The UNet architecture consists of an encoder and a decoder part. The encoder part extracts high-level features from the input image, while the decoder part upsamples these features to the original image size, and then concatenates them

with the corresponding features from the encoder to obtain the final segmentation map. The architecture has shown impressive performance in various biomedical image segmentation tasks and has been widely adopted and extended for other image segmentation applications.

EfficientNet is another deep learning architecture that was proposed in 2019 by Tan and Le in their paper [1]. The EfficientNet architecture achieves state-of-the-art performance on various image classification tasks with much fewer parameters and FLOPS compared to previous state-of-the-art models, by adopting a novel compound scaling method that efficiently scales up the depth, width, and resolution of the network. The authors also introduced a new scaling parameter called "compound coefficient" that uniformly scales all dimensions of the network based on a single scaling parameter. The EfficientNet architecture has been widely adopted in various computer vision tasks and has set a new standard for model efficiency. Utilizing these, Gouse Mohiddin et al. [7] has offered a theory of preprocessing of images for color consistency, hair removal, noise filtering and edge enhancement. Pre-processed image is then fed to convolutional network to get better results. Fatemah Bagheri et al. [5] has also proposed a two stage method to get the segmentation and masking separately to address different factors of dermoscopic images. This combination strategy has achieved the Jaccard Coefficient index of 80 percent for overall lesion segmentation. To address the boundry accuracy in leison segmentation, Lituan Wang et al. [8] has proposed Deep edge convolutional neural networks based on an encoder-decoder structure to focus more on the skin lesion boundary information. An edge information guided module is designed to introduce more information about the boundary. A new loss function including full loss, center loss and edge loss is proposed to pay more attention to boundary optimization.

Different researchers have worked for the purpose of skin segmentation with

different ideas. Some of them have used complete image whereas others have argued that to get more precision with utmost use of smart methodology, it is better to find the respective region of interests on the image and apply Deep Learning models on that specific region rather than using it for complete image. Hao Zheng et al. [4] has provided representative captions as an alternative for better image captioning. The proposed method relies on an unsupervised network that extracts features by directly targeting critical instances in the image followed by a fully connected trained supervised network to segment the image. The segmentation method selects the representative human annotations with reduced inter- and intra-cluster redundancy. Nooshin Moradi et al. [9] has put forward a multi class image segmentation instead of binary segmentation based on combining data from different feature spaces to build more informative structure.Two dictionaries are jointly learned using the K-SVD algorithm and then final segmentation is accomplished by a graph-cut method to distinguish background and foreground based on topological information of lesions and the learned dictionaries.

Medical data are scarce and collecting them is a difficult and time consuming process, while their annotation has to be performed by multiple specialists to ensure its validity [10]. Different deep learning models have shown varying degrees of achievement to date which has shown acceptability of the clinicians. Blind evaluation of these results by board-certified pathologists has also demonstrated similarities with gold standard [11]. The grouping of radiology through the application of information technology has led to its digitization. The images are digitally created, stored, rapidly transmitted over long distances, and consulted by medical professionals. New advancements in information technology has brought the possibility of 3D/ 4D in MRI [12] and further adding to the process of rapid diagnosis. Today, images are clearer, more detailed than ever and annotated [10], which allows healthcare teams to ultimately take a better approach towards patient care.

The literature indicates that different algorithms which have been used for the image segmentation purposes and have provided good results. It has also been observed that combination of different algorithms either for pre-processing or post-processing, shows more promising result some extra computation. This factor has remained motivational factor and many segmentation processes are now being implemented with grouping of different algorithms targeting different features in the image for better results. Y Zhang et al. [13] has used an improved inception module in the encoder to efficiently extract and synthesize information from different receptive fields, followed by a new mesh synthesis strategy to gradually refine shallow features and further smooth the semantic gap for brain tumor segmentation. Similarly for breast cancer detection, KB Soulami et al. [14] has used U-Net followed by loss function to improve the accuracy of the model. For segmentation of retinal vessel, H Wu et al. [15] has first proposed a scale-aware feature synthesis module, which aims to dynamically adjust receptive fields to efficiently extract multi-scale features. An adaptive feature matching module is then designed to guide the efficient combination of adjacent hierarchical features to capture more semantic information.

3.0.2 Gaps

The existing work suffers from several limitations, including inadequate accuracy, difficulties in extracting relevant features, and ineffective performance of different backbones on U-Net. Additionally, lower performance on some cancerous classes needed to be addressed especially with respect to segmentation. Another issue was the interpretation of overlap in patches of Whole Slide Images (WSIs) and the limited information of neighboring patches, which led to the problem of ragged patches.

Digital staining on the other hand undergoes certain limitations including varia-

tions in staining, tissue thickness and laboratory preprocessing which can affect the accuracy of automated segmentation methods. Another challenge is the analysis of staining condition in digital histopathological whole slide images which is essential for developing which is aiding in diagnostic systems but is challenging due to high data volume. Moreover, there might be issues with incomplete segmentation of the epidermal area in digitized H&E stained WSIs. Finally, there may be nuances in anatomy, inter and intra-variations in staining and cellular infiltration and skin appendages that can affect the accuracy of automated image analysis algorithms.

3.0.3 Contributions

Our work aims to address above mentioned limitations and improve upon them. Our Efficient-Net model builds upon the works of Simon et al. [16] by training the model to disburse the image with respect to different layers available in that specific image and assign each layer to a class. Overall, by experimenting we are able to:

- Provide a more efficient and powerful encoder architecture using Efficient-Net B3.

- Address limitations by training the model to disburse the image into layers and assign each layer to a class, leading to smoother predictions.

- Achieve better results with improved accuracy and reliability as shown by uncertainty maps.

- Analyze the performance of different backbones on a UNet for semantic segmentation tasks.

- For staining a DCL GAN Model for generating images

Chapter 4

MATERIALS AND METHODS

In this investigation, we present a comprehensive framework for examining histopathological skin whole slide images (WSIs) by employing Generative Networks to simulate staining and Convolutional Neural Networks for segmentation. Our primary objective is to identify various skin manifestations and generate detailed pathology reports using advanced deep learning models. This chapter provides an overview of our proposed system and an in-depth discussion on the datasets employed in this study.

A crucial element of our research involves utilizing both unstained and stained WSIs. We collect unstained WSIs and leverage Generative Networks to mimic the staining process, resulting in corresponding stained WSIs. Additionally, for each stained WSI, we generate a mask labeled by medical professionals, which serves as a reliable reference for segmentation.

The selection of representative datasets plays a vital role in our investigation. We ensure that the datasets encompass a wide range of skin pathologies, covering various manifestations and conditions. Careful consideration is given to the suitability of the datasets for our specific requirements. We delve into the distinctive attributes, origins, and peculiarities associated with each dataset, alongside a de-

tailed explanation of the proposed model, in subsequent sections. These datasets form a robust foundation for training and evaluating our deep learning models for the analysis of histopathological skin WSIs.

4.1 Datasets

Public datasets, as well as private datasets from our research collaborators are used in this study. The following sections provides a comprehensive overview of all benchmark datasets used in this study, along with the necessary pre-processing steps:

4.1.1 Queensland Histopathology Dataset

The primary dataset utilized for segmentation was the Histopathology Non-Melanoma Skin Cancer Segmentation dataset, provided by Queensland University [16]. The dataset consisted of 290 high-resolution images of skin cancer specimens, which were down sampled at various factors, including 1x, 2x, 5x, and 10x. The dataset includes three cancer classes: Basal Cell Carcinoma (*BCC*), Squamous Cell Carcinoma (*SCC*), and Intra-Epidermal Carcinoma (*IEC*), with a total of 290 images. The specimens were obtained from patients aged between 34 and 96 years, with a median age of 70 years. The gender distribution in the dataset was 2/3 female and 1/3 male. The ground-truth segmentation was created by color-coding the images into 12 classification categories, as shown in figure 1, including Glands (*GLD*), Inflammation (*INF*), Hair Follicles (*FOL*), Hypodermis (*HYP*), Reticular Dermis (*RET*), Papillary Dermis (*PAP*), Epidermis (*EPI*), Keratin (*KER*), Background (*BKG*), *BCC*, *SCC*, and *IEC*. To determine the optimal down-sampling factor, the dataset was subjected to a series of experiments, which showed that the 5x and 10x down-sampling factors had minimal impact on the segmentation performance. Therefore, the 10x down-sampling factor was used

in this study to ensure that the segmentation results were representative of the underlying data and that the performance of the method was not affected by external factors. Thus the Histopathology Non-Melanoma Skin Cancer Segmentation dataset, combined with the selection of an appropriate down-sampling factor, provided a reliable and rigorous foundation for the investigation of the proposed segmentation method in this research.

BKG	GLD	INF	FOL	HYP	RET	PAP	EPI	KER	BCC	SCC	IEC
0,0,0	108,0,115	145,1,122	216,47,148	254,246,242	181,9,130	236,85,157	73,0,106	248,123,168	127,255,255	127,255,142	255,127,127
Background	Glands	Inflammation	Hair Follicle	Hypodermis	Reticular Dermis	Papillary Dermis	Epidermis	Keratin	Basel Cell Carcinoma	Squamous Cell Carcinoma	Intra Epidermal Carcinoma
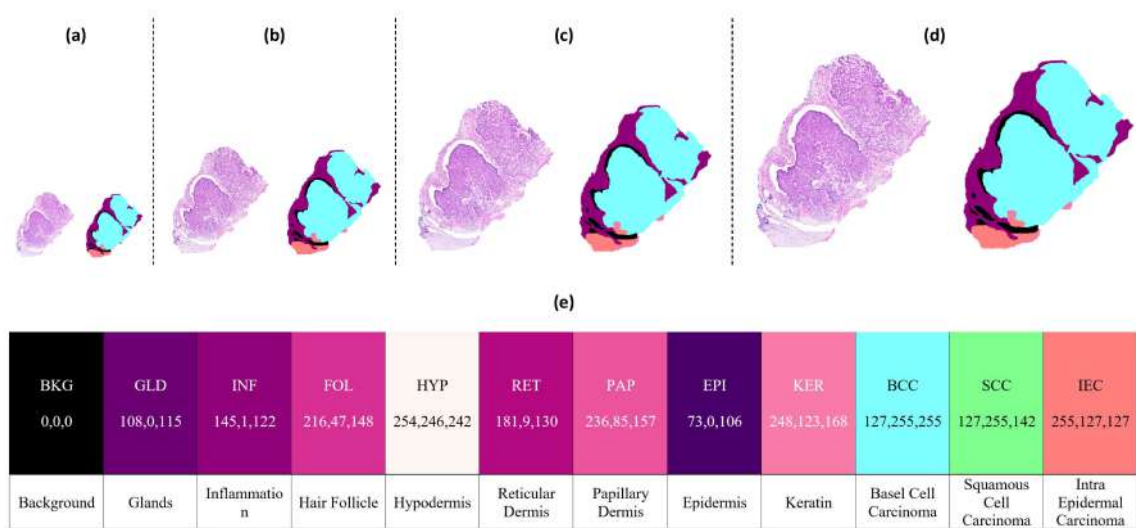
Figure 4.1: Image with different magnifications and the color coding to segment different parts of the skin. This helps visualize how down sampling plays a role in our study. a) Image sample with its mask, at 10x downsampled. b) Image sample with mask 5x downsampled. c) Image sample with mask 2x downsampled. d) Image sample with mask 1x downsampled. e) The color palette visualized for the mask, with the individual mask RGB values.

Figure 7.1 highlights the distribution of whole slide images and the corresponding layer masks. Each layer is represented by a unique RGB color code. It also shows how the different scales of magnification compare with each other.

4.1.2   NIDI Skin Dataset

The Skin Histopathological Dataset, provided by Non Invansive Diagnostic in Skin (NIDI Skin) consists of 93 samples of histopathological images, specifically Whole Slide Images (WSIs) of human skin tissue.

Staining Dataset

The dataset includes both H&E (Hematoxylin and Eosin) stained and unstained images. The raw dataset initially provided contained 27 pairs of TIFF (.tif) files, with each file containing multiple samples. Prior to analysis and research purposes, the dataset underwent preprocessing, which involved extracting each individual sample from the pairs of files. This resulted in a final dataset comprising 93 distinct samples. The stained images represent H&E staining, a common technique used in dermatology and pathology to visualize cellular structures and tissue morphology. The unstained images provide a baseline reference for comparison and analysis. The images in the dataset are stored in the TIFF format, ensuring high-quality and lossless preservation of the histopathological details necessary for accurate analysis. This Skin Histopathological Dataset can be valuable for various applications in dermatological research, pathology analysis, and developing computer vision algorithms for skin image analysis.

Segmentation Dataset

Additionally, out of the 93 samples in the Skin Histopathological Dataset, 40 samples have been annotated with labeled masks provided by doctors. The labels on these masks correspond to specific classes established by the Queensland University. The inclusion of these labeled masks enhances the dataset's value, as it allows for supervised machine learning and computer vision tasks, such as segmentation and classification, to be performed on the dataset. The labeled masks provide ground truth information about the regions of interest and different classes within the histopathological images, facilitating more precise analysis and accurate algorithm training.
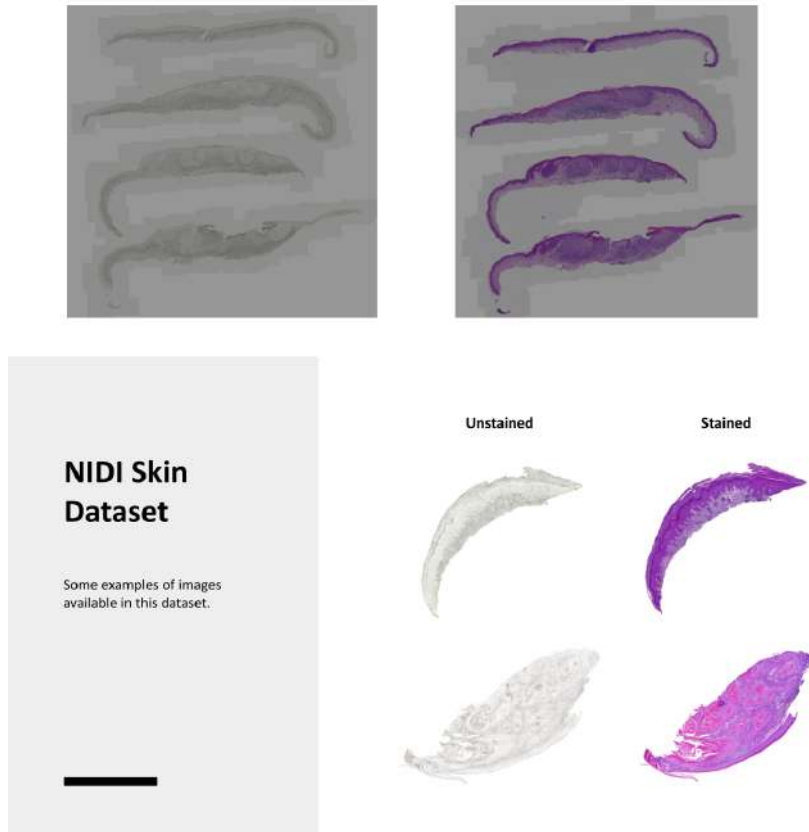
Figure 4.2: Dataset provided by the NIDI skin. The dataset provided had both unstained and stained samples, which after processing were fed to a generative network.

4.1.3 Data Processing

For ensuring a uniform standard as to how both the datasets were linked, heavy pre processing of the dataset was done. We had to ensure two aspects to stay constant: position and color features. The positional consistency was maintained through applying the Scale Invariant Feature Transform (SIFT). While the stain consistency was ensure through the approach defined by Macenko.et.al [17].
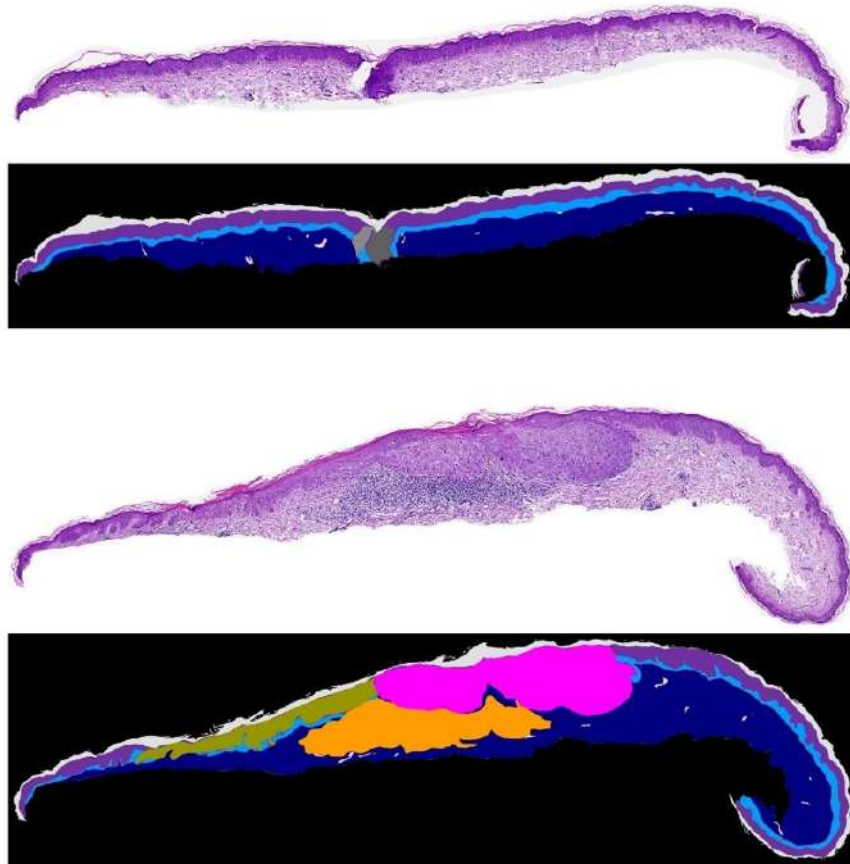
Figure 4.3: Segmentation dataset provided by the NIDI skin. For easier fine tuning of the model, the image was downsampled 5x to ensure each pixel corresponded to the same width as done by Simon et al add ref

SIFT

SIFT helps in aligning images by detecting and describing local features in the images that are invariant to scale, rotation, and other transformations. The process involves the following steps:

1. Compute SIFT descriptors for the input images.

2. Compare the descriptors to determine the geometric transformation needed to align the images.

3. Apply the transformation to align the images

Index	Name	Abbreviation	RGB Values	Color Fill
0	Background	BKG	0, 0, 0	
1	Keratin	KER	224, 224, 224	
2	Epidermis	EPI	112, 48, 160	
3	Papillary Dermis	PAP	0, 156, 255	
4	Reticular Dermis	RET	0, 0, 128	
5	Hypodermis	HYP	0, 255, 0	
6	Glands	GLD	127, 96, 255	
7	Hair Follicle	FOL	96, 96, 96	
8	Inflammation	INF	255, 156, 0	
9	Basal Cell Carcinoma	BCC	127, 255, 255	
10	Squamous Cell Carcinoma	SCC	255, 0, 255	
11	Intra Epidermal Carcinoma	IEC	150, 150, 0	

Table 4.1: The color key used by the NIDI dataset to differentiate between 2 or more layers

Feature-based alignment, such as SIFT, searches for an image transformation that matches keypoint locations between the images. This approach is more robust than direct alignment, which relies on pixel-based similarity in the overlap region SIFT's ability to detect features at different scales and angles makes it a reliable method for image alignment. The application of SIFT was done on ImageJ, as it provided the best results and visual changes could be made to the keypoints if any changes were observed in alignment.
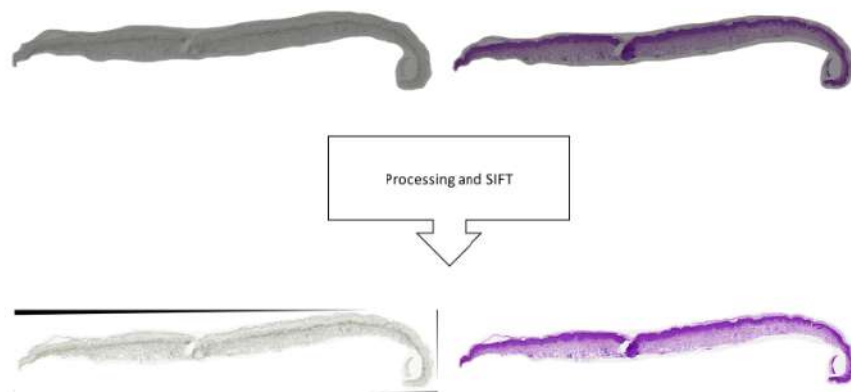
Figure 4.4: Scale Invaraint Feature transform being applied to the images.This would then be aligned with its unstained counterpart.
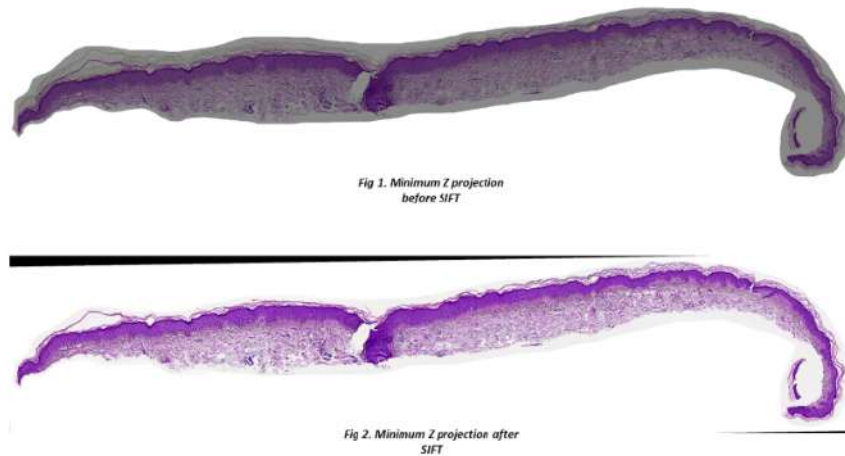
Figure 4.5: Scale Invaraint Feature transform being applied to the images.Fig 1 shows Z projection showing overlap before SIFT and figure 2 shows the same projection after SIFT.

Stain Normalization

The Macenko stain normalization algorithm [17] is a method used to normalize the color distribution in histopathology images. It estimates the stain vectors of the Whole Slide Images (WSIs) by using a singular value decomposition (SVD) approach applied to the non-background pixels of the input image. Macenko's method assumes that each image contains a pixel that represents a single stain vector and applies SVD to find the singular values. This technique helps to overcome inconsistencies in the staining process, allowing for more accurate and reliable quantitative analysis of histopathology slides. For this apporach the target image used was that provided by queensland.

Patches

Whole slide images are often processed in overlapping patches of size 256x256 due to their large dimensions. This approach is primarily chosen to enhance computational efficiency when dealing with these high-resolution images, typically obtained from histopathology slides. Dividing the images into smaller patches allows for easier handling and analysis of the data, as working with smaller units
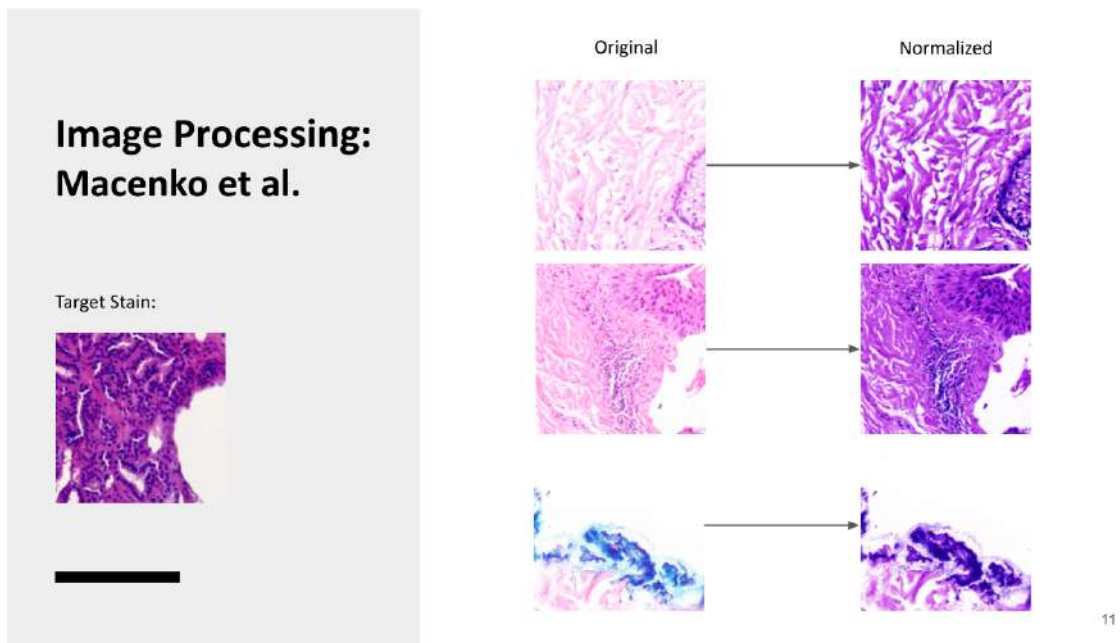
Figure 4.6: Normalized image result along with its target stain based on the approach defined by Macenko.et.al

is more manageable. Moreover, the overlapping strategy ensures that no vital information is lost at the patch boundaries, enabling a comprehensive examination of the entire slide. Another advantage of processing whole slide images in overlapping patches is the increase in available training data. Since each patch represents a distinct region of the slide, this technique expands the dataset, providing a diverse range of image samples for training machine learning models. Consequently, this expanded training data improves the model's capacity to generalize and make accurate predictions on new whole slide images.

4.2 Proposed Model

Our proposed model involves using a generative model for the staining part in whole slide images.First an unstained image would be passed to a GAN model and this would then be passed to a segmentation model of UNET with an efficient net b3 backbone.
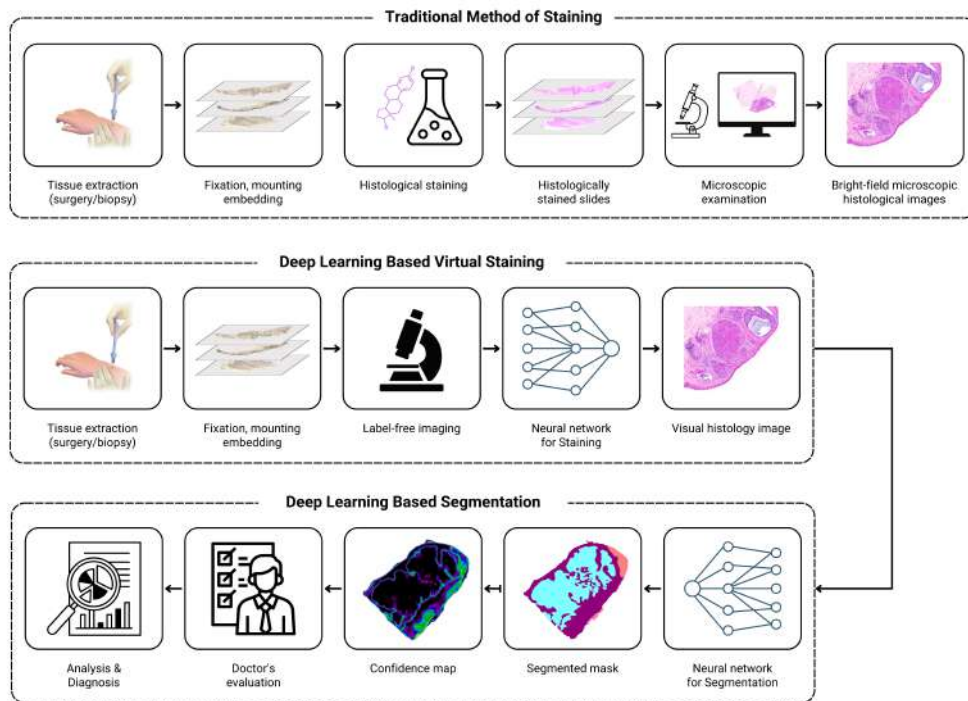
Figure 4.7: The block diagram for the proposed framework chest composed of Staining, segmentation and application modules.

The proposed model is therefore divided into two modules as shown in figure 4.7. The first module performs the staining parts by taking patches of 256x256 and this is then fed into the segmentation model already trained.

Chapter 5

PROPOSED FRAMEWORK FOR STAINING USING DUAL CONTRASTIVE LEARNING GAN'S

5.1 Overview

The Dual Contrastive Learning for Unsupervised Image-to-Image Translation (DCLGAN) model is a sophisticated approach that enhances the process of image conversion in unsupervised image-to-image translation tasks. DCLGAN expands upon the principles of CycleGAN and Contrastive Learning for Unpaired image-to-image Translation (CUT) to address their shortcomings. CycleGAN, which relies on cycle consistency to maintain image quality, encounters difficulties with significant geometric transformations between input and target domains. Conversely, CUT utilizes contrastive learning to develop feature embeddings that bring related features closer.

DCLGAN refines the CUT method by integrating the concept of two-sided mappings from CycleGAN. In the scenario of converting unstained skin histopathol-

ogy samples into stained ones, DCLGAN applies contrastive learning to maximize the mutual information shared between input and generated image patches. By mastering cross-domain similarity and embedding mappings, DCLGAN achieves exceptional outcomes in image-to-image translation. This effectively tackles the challenge of converting unstained histopathology samples into their stained equivalents.

DCLGAN is denoted as $\mathcal{A} \subset \mathbb{R}^{H \times W \times C}$ and $\mathcal{B} \subset \mathbb{R}^{H \times W \times 3}$. Here, H, W, and C represent the height, width, and number of channels, respectively. The model leverages datasets $\mathcal{A}$ and $\mathcal{B}$, which consist of sets of images a and b belonging to domains $\mathcal{A}$ and $\mathcal{B}$, respectively. These datasets are denoted as $\mathcal{A} = a \in \mathcal{A}$ and $\mathcal{B} = b \in \mathcal{B}$.

5.2 Architecture

The DCLGAN architecture consists of two generators, $G_{\mathcal{A} \to \mathcal{B}}$ and $G_{\mathcal{B} \to \mathcal{A}}$, and two discriminators, $D_{\mathcal{A}}$ and $D_{\mathcal{B}}$.

5.2.1 Generators

The generator $G_{\mathcal{A} \to \mathcal{B}}$ takes an image $a \in \mathcal{A}$ from domain $\mathcal{A}$ as input and generates a corresponding image $b' \in \mathcal{B}$ in domain $\mathcal{B}$. Similarly, the generator $G_{\mathcal{B} \to \mathcal{A}}$ takes an image $b \in \mathcal{B}$ from domain $\mathcal{B}$ as input and generates a corresponding image $a' \in \mathcal{A}$ in domain $\mathcal{A}$.

The generator $G_{\mathcal{A} \to \mathcal{B}}$ can be represented by the function $G_{\mathcal{A} \to \mathcal{B}} : \mathcal{A} \to \mathcal{B}$, which learns the mapping from domain $\mathcal{A}$ to domain $\mathcal{B}$. Similarly, the generator $G_{\mathcal{B} \to \mathcal{A}}$ can be represented by the function $G_{\mathcal{B} \to \mathcal{A}} : \mathcal{B} \to \mathcal{A}$, which learns the mapping from domain $\mathcal{B}$ to domain $\mathcal{A}$.

5.2.2 Discriminators

The discriminator $D_{\mathcal{A}}$ is responsible for distinguishing between real images from domain $\mathcal{A}$ and images generated by $G_{\mathcal{B} \to \mathcal{A}}$. It takes an image $a \in \mathcal{A}$ as input and outputs a scalar value indicating the probability that the image is real.

Similarly, the discriminator $D_{\mathcal{B}}$ is responsible for distinguishing between real images from domain $\mathcal{B}$ and images generated by $G_{\mathcal{A} \to \mathcal{B}}$. It takes an image $b \in \mathcal{B}$ as input and outputs a scalar value indicating the probability that the image is real.

The discriminator $D_{\mathcal{A}}$ can be represented by the function $D_{\mathcal{A}} : \mathcal{A} \to [0, 1]$, which learns to classify images from domain $\mathcal{A}$ as real or fake. Similarly, the discriminator $D_{\mathcal{B}}$ can be represented by the function $D_{\mathcal{B}} : \mathcal{B} \to [0, 1]$, which learns to classify images from domain $\mathcal{B}$ as real or fake.

5.2.3 Loss Functions

The generators and discriminators are trained using adversarial loss and Patch Noise Contrastive Estimation loss.

The adversarial loss for $G_{\mathcal{A} \to \mathcal{B}}$ is defined as:

$$\mathcal{L}_{\text{adv}}(G_{\mathcal{A} \to \mathcal{B}}, D_{\mathcal{B}}, \mathcal{A}, \mathcal{B}) = \mathbb{E}_{b \sim \mathcal{B}}[\log(D_{\mathcal{B}}(b))] + \mathbb{E}_{a \sim \mathcal{A}}[\log(1 - D_{\mathcal{B}}(G_{\mathcal{A} \to \mathcal{B}}(a)))] \quad (5.1)$$

Similarly, the adversarial loss for $G_{\mathcal{B} \to \mathcal{A}}$ is defined as:

$$\mathcal{L}_{\text{adv}}(G_{\mathcal{B} \to \mathcal{A}}, D_{\mathcal{A}}, \mathcal{B}, \mathcal{A}) = \mathbb{E}_{a \sim \mathcal{A}}[\log(D_{\mathcal{A}}(a))] + \mathbb{E}_{b \sim \mathcal{B}}[\log(1 - D_{\mathcal{A}}(G_{\mathcal{B} \to \mathcal{A}}(b)))] \quad (5.2)$$

The Patch Noise Contrastive Estimation loss is defined as:

$$\mathcal{L}\text{PatchNCE}(G_1, H_A, H_B, A) = \mathbb{E}a \sim A \sum_{l=1}^{L} \sum_{s=1}^{S_l} \ell(z_{\hat{l}}^s, z_l^s, z_l^{(S \setminus s)}) \quad (5.3)$$
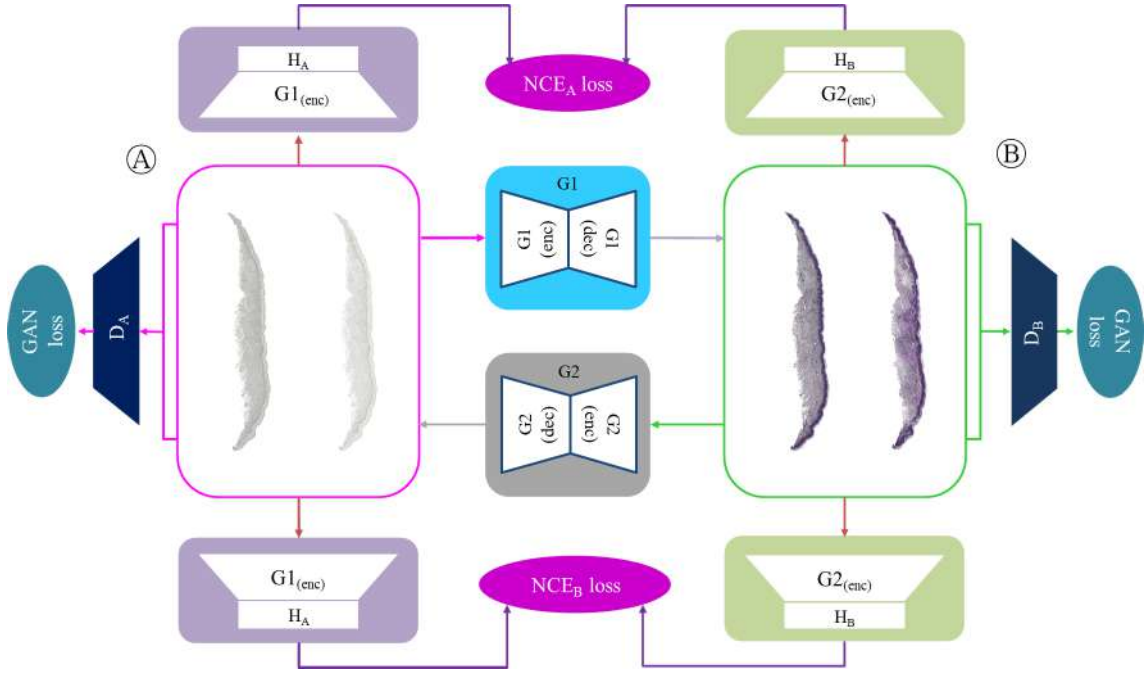
Figure 5.1: DCL GAN Architecture along with Patch contrastive Learning and loss explained.

The total generator loss is given by:

$$\mathcal{L}_{\text{gen}}(G_{\mathcal{A}\to\mathcal{B}}, G_{\mathcal{B}\to\mathcal{A}}, D_{\mathcal{A}}, D_{\mathcal{B}}, \mathcal{A}, \mathcal{B}) = \mathcal{L}_{\text{adv}}(G_{\mathcal{A}\to\mathcal{B}}, D_{\mathcal{B}}, \mathcal{A}, \mathcal{B}) + \mathcal{L}_{\text{adv}}$$
$$(G_{\mathcal{B}\to\mathcal{A}}, D_{\mathcal{A}}, \mathcal{B}, \mathcal{A}) + \lambda\mathcal{L}_{\text{NCE}}(G_{\mathcal{A}\to\mathcal{B}}, G_{\mathcal{B}\to\mathcal{A}}, \mathcal{A}, \mathcal{B})$$

(5.4)

where λ is a hyperparameter controlling the weight of the NCE Loss.

The discriminators are trained using the adversarial loss defined as:

$$\mathcal{L}_{\text{adv}}(D_{\mathcal{A}}, D_{\mathcal{B}}, G_{\mathcal{A}\to\mathcal{B}}, G_{\mathcal{B}\to\mathcal{A}}, \mathcal{A}, \mathcal{B}) = \mathbb{E}_{a\sim\mathcal{A}}[\log(D_{\mathcal{A}}(a))]$$
$$+ \mathbb{E}_{b\sim\mathcal{B}}[\log(1 - D_{\mathcal{A}}(G_{\mathcal{B}\to\mathcal{A}}(b)))] + \mathbb{E}_{b\sim\mathcal{B}}[\log(D_{\mathcal{B}}(b))] + \mathbb{E}_{a\sim\mathcal{A}}[\log(1 - D_{\mathcal{B}}(G_{\mathcal{A}\to\mathcal{B}}(a)))]$$

(5.5)

The final objective function for training DCLGAN is to minimize the generator loss $\mathcal{L}_{\text{gen}}$ and the discriminator loss $\mathcal{L}_{\text{adv}}$ simultaneously.

5.3 Results

Overall the results were evaluated using both quantitative and qualitative measures. The results were constantly reviewed by the team of doctors at NIDI Skin and were approved or disapproved based on the findings. The quantitative analysis was done with different metrics as explained below.

5.3.1 Evaluation Metrics

In order to evaluate the performance of Generative Adversarial Networks (GANs), it is important to utilize appropriate evaluation metrics. Two commonly employed metrics for GAN result evaluation are Kernel Inception Distance (KID) [18] and Fréchet Inception Distance (FID) [19].

Kernel Inception Distance is a metric that quantifies the dissimilarity between the generated samples and the real data distribution based on their feature representations obtained from an Inception model. KID assesses the quality of generated samples by measuring the similarity of their feature embeddings to those of real samples. A lower KID value implies higher quality generated samples.

To compute KID, let F represent the Inception model used for feature extraction. Given a set of generated samples $\mathcal{G}$ and a set of real samples $\mathcal{R}$, the KID is calculated as follows:

$$
KID(\mathcal{G}, \mathcal{R}) = \frac{1}{n_g^2} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} k(f(\mathbf{g}_i), f(\mathbf{g}_j)) - \frac{2}{n_g n_r} \sum_{i=1}^{n_g} \sum_{j=1}^{n_r} k(f(\mathbf{g}_i), f(\mathbf{r}_j))
$$
$$
+ \frac{1}{n_r^2} \sum_{i=1}^{n_r} \sum_{j=1}^{n_r} k(f(\mathbf{r}_i), f(\mathbf{r}_j))
$$

(5.6)

Here, $\mathbf{g}_i$ and $\mathbf{r}_j$ represent the i-th generated sample and j-th real sample, respectively. $f(\cdot)$ denotes the feature embedding obtained from the Inception model F,

and $k(\cdot, \cdot)$ represents a kernel function (e.g., Gaussian kernel). n_g and n_r denote the number of generated and real samples, respectively.

Fréchet Inception Distance is another popular metric for evaluating GAN results. It measures the similarity between the generated samples and the real data distribution by comparing their multivariate Gaussian distributions in the feature space of an Inception model.

To calculate FID, let M represent the Inception model used for feature extraction. Given a set of generated samples $\mathcal{G}$ and a set of real samples $\mathcal{R}$, the FID is computed as follows:

$$FID(\mathcal{G}, \mathcal{R}) = \|\mu_{\mathbf{g}} - \mu_{\mathbf{r}}\|_2^2 + \mathrm{Tr}(\Sigma_{\mathbf{g}} + \Sigma_{\mathbf{r}} - 2(\Sigma_{\mathbf{g}}\Sigma_{\mathbf{r}})^{1/2}) \tag{5.7}$$

Here, $\mu_{\mathbf{g}}$ and $\mu_{\mathbf{r}}$ represent the mean vectors, and $\Sigma_{\mathbf{g}}$ and $\Sigma_{\mathbf{r}}$ represent the covariance matrices of the feature embeddings for the generated and real samples, respectively.

Both KID and FID scores provide quantitative measures of the quality of generated samples. A lower KID or FID score indicates higher quality and greater similarity to the real data distribution. It is important to note that both metrics rely on pretrained Inception models, and the choice of the Inception model can influence the results. Additionally, comparing the scores obtained by different GAN models trained on the same dataset helps evaluate their relative performance.

Metric	Unstained & H&E Stained	Unstained & Virtually Stained
KID	0.25	0.31
FID	322.3	345

Table 5.1: Evaluation Metrics: KID and FID

5.3.2 Visual Results

Upon visual inspection, the transformed stained images exhibit convincing color variations, highlighting the presence of different cellular structures and tissue components. The staining patterns and intensity are realistically replicated, allowing for better visual interpretation of the histopathology samples. Notably, the GAN model effectively incorporates the staining-specific geometrical transformations, such as the differential absorption of stains by different tissue components, resulting in accurate representations of stained images.
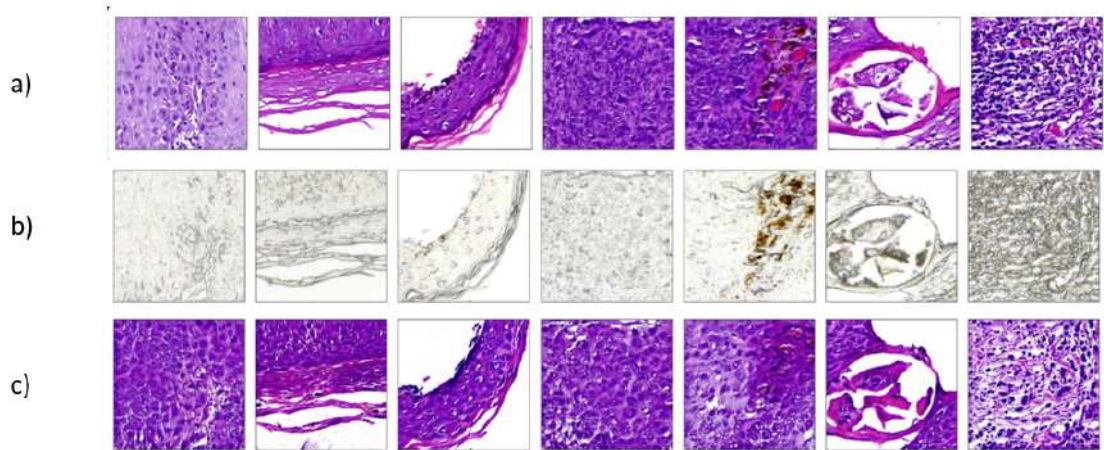
Figure 5.2: Results shown from the DCL GAN model. In the diagram below a) Shows the images stained with normal H&E Stains in a laboratory. b) Shows Unstained WSI patches that were fed to the network. c) Shows corresponding generated images from the DCL GAN model.

Chapter 6

PROPOSED FRAMEWORK FOR SEGMENTATION

6.1 Models

6.1.1 Overview

In recent years, the field of dermatopathology has benefited greatly from the integration of deep learning techniques for the segmentation of whole slide skin images. Among the notable architectures used for this purpose are U-Net and EfficientNet-B3. In this study, we have leveraged the capabilities of both U-Net and EfficientNet-B3 to achieve accurate and efficient segmentation of whole slide skin images.

The U-Net architecture, proposed by Ronneberger et al. in 2015, has gained significant recognition in the biomedical imaging domain, including dermatopathology. Its unique encoder-decoder structure and skip connections enable the network to effectively capture both high-level and low-level features, facilitating precise localization and boundary delineation. By utilizing U-Net, we aim to leverage its

proven ability to achieve accurate segmentation results in the context of whole slide skin images.
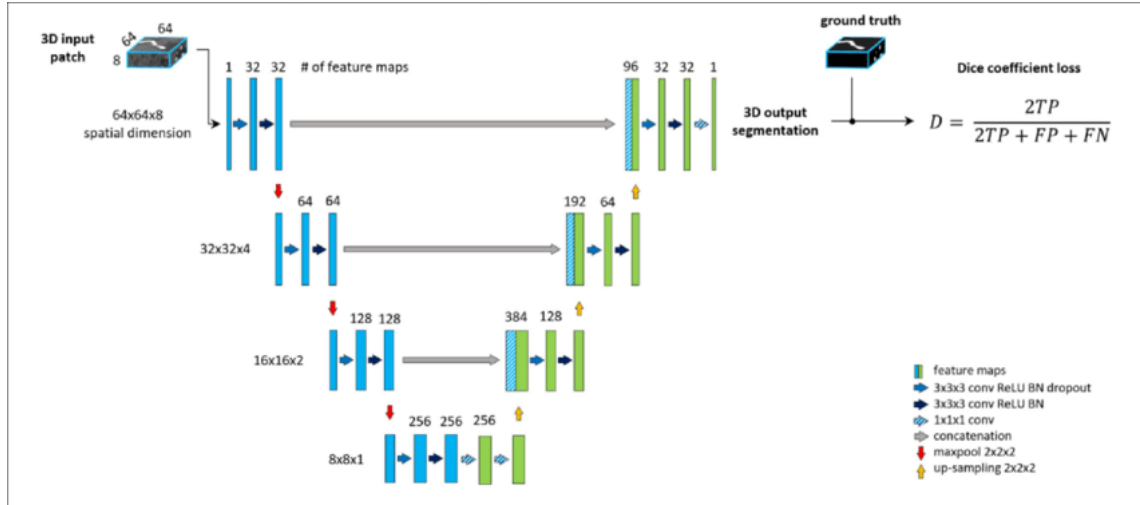
Figure 6.1: Basic UNet Architecture.

In addition to U-Net, we have also incorporated EfficientNet-B3, which belongs to the EfficientNet family of architectures proposed by Tan and Le in 2019. EfficientNet-B3 is specifically designed to strike a balance between model complexity and computational efficiency. It achieves this through a compound scaling method that optimizes the network's depth, width, and resolution simultaneously. By efficiently utilizing computational resources, EfficientNet-B3 has demonstrated state-of-the-art performance across various computer vision tasks.

By integrating U-Net and EfficientNet-B3, we aim to capitalize on their respective strengths to enhance the segmentation of whole slide skin images. U-Net's encoder-decoder structure and skip connections enable accurate localization of tissue components, while EfficientNet-B3's efficient architecture ensures computational efficiency and high performance.

U-NET

U-Net architecture includes context detection followed by correct localization. Usually it is difficult to have a large amount of training data in the field of biomedical, so a good deal about this algorithm is that it can be trained end to end with fewer images and outcomes with relatively the star performance. UNet has a proven track record of multi-layer image segmentation by giving a specific class to each pixel in the image. The input data is downsampled (encoded) to extract smaller features from the image and also to represent it for better segmentation, followed by upsampling (decoding) with concatenation of the skip connections from the downsampled layer to restore the original image. Pairing this helps the model get more details from the image while it was being encoded. To put it in simple way, the downsampling process extracts the "what" from the image, and upsampling connects it to the "where" in that image [2].

In the first half of the architecture that consists of pre trained classification encoder network, we gave our input images of size 256x256 with 3 channels. It goes through couple of 3x3 convolutional layers that comprise of 64 filters, then maxpooling operation of 2x2 is performed to down samples the spatial dimensions of image to 128x128 with 128 filters. The steps were repeated to gradually increase the depth and reduce the image size from 128x128 with 128 filters to 16x16 with 1024 filters progressively.

Decoder network has 1024 filters with 16x16 dimension of lower resolution onto the pixel space. The up sampling is done by 2x2 transpose convolution by concatenation from corresponding blocks of encoder network, and to keep the spatial resolution intact ReLu function is performed. Here our 16x16 image with 1024 filter is up sampled to 32x32 with 512 filter along with the concatenation from its corresponding 512 features map from encoder. The depth is gradually decreased and image size is increased until we reach our final layer. In the end 1x1 con-

volution is used to map our desired 3 channeled 256x256 image to 12 classes for Segmentation.

EfficientNet-B3

ConvNet provides better result if scaled up in width, depth or image resolution, but after achieving a certain level, the result show very minute improvement. Scaling some or all components has also been tested, but requires a lot of manual fine-tuning [20] and consistently produces substandard accuracy and efficacy. Extending network depth is the most common means used by many ConvNets where the intuition is that the deeper ConvNet can capture richer and more complex functionality and generalize well to new tasks [21]. At the same time, deeper networks are more difficult to form due to the vanishing gradient problem. Network width scaling is often used for small models as larger networks tend to be able to capture features in more detail and are easier to train [22]. However, ultra-wide but shallow networks tend to have difficulty capturing higher-level peculiarities. With higher resolution input images, ConvNets is capable of capturing better segments [23], but accuracy decreases for very high resolutions. Efficient Net is a variety of ConvNet but with balanced scaling with respect to width, depth and Image resolution. It is observed that the different scale sizes are not independent. Intuitively, for higher resolution images we should increase the depth, so that larger receptive fields can help capture the identical features including more pixels in the large image. Accordingly, we should also increase the width when the resolution is higher, to capture finer samples with more pixels in the high resolution image. These perception suggest that we need to coordinate and balance different scale sizes rather than conventional uni-directional scaling [1].

Our Efficient Net model consisted of 12M parameters and consisted of 9 layers each in the encoder and decoder. The base layer had a total of 1536 features which then grew into the class wise mask segmentation. Model is trained to disburse

44

the image with respect to different layers available in that specific image and assign each layer to a class. The result of comparison made to test the accurate classification with Simon et al. [24] and the results of comparison of segmentation with different models are explained in table 2 and 3 respectively
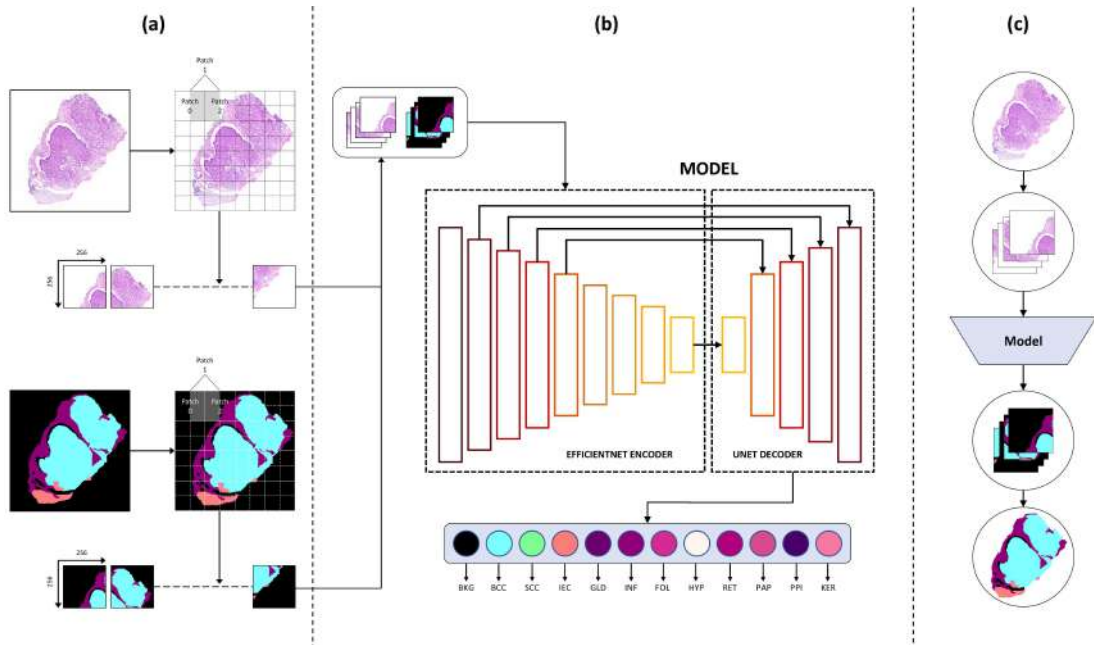
Figure 6.2: An overview of the entire working of the segmentation pipeline. **(a)** Shows the data input pipeline for training the model architecture, the input WSI is split into overlapping patches of 256x256 and the same is done to the corresponding masks. This encompasses the input pipeline of the WSI and the corresponding masks to the architecture.**(b)** The overlapping patches are taken and, in the form of batches, is fed into the model architecture. The model architecture is an Efficient Net-B3 [1] encoder with a simple U-Net decoder [2]. The decoder transforms the 256x256x3 input to a 256x256x12, one hot encoded, output where each layer's probability is shown. **(c)** This stage shows the testing pipeline, whereby a new image is tested on the network. This uses the approach shown by [3] where the input image is patched, rotated and augmented in 4 different directions and then fed to the inference model. This ensures smooth predictions and the network to achieve maximum confidence by taking into account all neighbours.

6.1.2 Models Training

The fine-tuning of the U-Net backbone models in the Queensland dataset was performed using TensorFlow and trained on a GeForce RTX 2070 8GB GPU. It will be noted that during training, the model parameters were adjusted using the

Adam optimization algorithm with a batch size of 4 and maximum 50 epochs, based on the gradients of loss function. The parameters resulting in the minimum validation loss during training were used for testing the model. Early stopping and learning rate reduction callbacks were used to improve model performance. The Efficient net b3 model showed a training loss of 0.3 and a training accuracy of 94, while a validation loss of 0.38 and a validation accuracy of 92. The loss used in the training was the categorical cross-entropy loss, this was done as in the loss the sum over all classes ensures that the loss is computed for all classes and the negative sign in front of the sum ensures that the loss is minimized during training. The loss is defined as:

$$\mathcal{L}CE = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \tag{6.1}$$

During the testing phase, the input images underwent a smoothing process to improve the accuracy of the predictions. The smoothing process included a combination of rotations, mirroring, and windowing techniques. The Dihedral Group D_4 was applied to augment the images by generating eight different versions of each image at different angles, which were then combined to produce a more robust prediction. Additionally, a simple second-order spline window function was used for interpolation to blend the predictions together, with a default overlap of 50% between merged windows. This approach helped to ensure that the predictions were merged smoothly and effectively, leading to better overall performance of the model.

6.1.3   Models Training

The fine-tuning of the U-Net backbone models in the Queensland dataset was performed using TensorFlow and trained on a GeForce RTX 2070 8GB GPU. It will be noted that during training, the model parameters were adjusted using the

Adam optimization algorithm with a batch size of 4 and maximum 50 epochs, based on the gradients of loss function. The parameters resulting in the minimum validation loss during training were used for testing the model. Early stopping and learning rate reduction callbacks were used to improve model performance. The Efficient net b3 model showed a training loss of 0.3 and a training accuracy of 94, while a validation loss of 0.38 and a validation accuracy of 92. The loss used in the training was the categorical cross-entropy loss, this was done as in the loss the sum over all classes ensures that the loss is computed for all classes and the negative sign in front of the sum ensures that the loss is minimized during training. The loss is defined as:

$$\mathcal{L}CE = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \tag{6.2}$$

During the testing phase, the input images underwent a smoothing process to improve the accuracy of the predictions. The smoothing process included a combination of rotations, mirroring, and windowing techniques. The Dihedral Group D_4 was applied to augment the images by generating eight different versions of each image at different angles, which were then combined to produce a more robust prediction. Additionally, a simple second-order spline window function was used for interpolation to blend the predictions together, with a default overlap of 50% between merged windows. This approach helped to ensure that the predictions were merged smoothly and effectively, leading to better overall performance of the model.

6.2   Experiments & Results

The proposed model has been tested using skin biopsy images which in turn were segmented by histopathologists.

6.2.1 Evaluation Metrics

After determining the set of ground-truth (GT) images, the final segmented whole slide images were compared to them. The testing results in terms of class-wise recall are shown in Table 1. To define these metrics, the individual pixels in both sets of GT and model outputs were categorized into four classes: True-Positive (TP), True-Negative (TN), False-Positive (FP), and False-Negative (FN) as follows:

- TP: pixels segmented as the said class by the model that match the manual annotations.

- TN: pixels that are not segmented as the said class in the model, which match the manual annotations.

- FP: pixels segmented as the said class by the model, which are not present as the said class in the manual annotations.

- FN: pixels that are not segmented as the said class by the model, while they are segmented as the said class in the manual annotations.

The class-wise accuracy is defined as the recall:

$$Recall = \frac{TP}{TP + FN} \qquad (6.3)$$

Another useful metric is the Dice F-1 Score which shows intersection of a said class in the terms of the total area:

$$F1 - Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \qquad (6.4)$$

Overall the per-pixel accuracy score is calculated, which is given by:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (6.5)$$
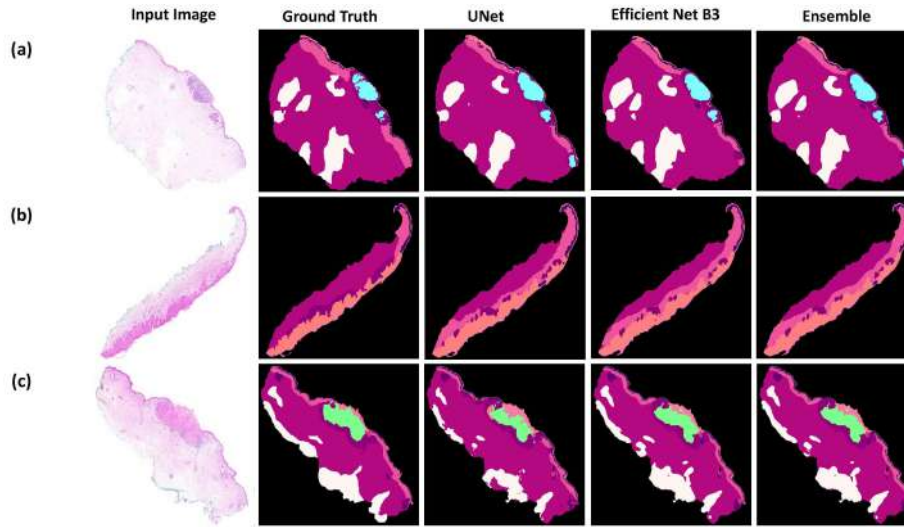
Figure 6.3: Visualizing the Accuracy: Generated Masks of Skin Cancer Types from the Queensland Dataset compared to Ground Truth: a) Generated Mask with Predominant BCC Cancer Visually Represented. b) Generated Mask with Predominant IEC Cancer Displayed. c) Generated Mask Illustrating Predominant SCC Cancer.

6.2.2   Ablation study

To evaluate the proposed segmentation model, we have conducted detailed ablation studies. As shown in figure 6.3, our studies explored how different backbones on a U-Net performed. The predictions from these models were later also aggregated to form an ensemble, which included the original vanilla U-Net architecture as well. All of our experiments are on the Queensland dataset and we reported both qualitative and quantitative results. The hyperparameters used in the experiments are shown in Table 1.

Parameter	Value
Batch Size	4
Buffer Size	40
Epochs	50
Learning Rate	0.0001
Dimensions	256 x 256

Table 6.1: Hyperparameters used in the ablation study. The learning rate was dynamic and changed if a plateau in training was observed, at the same time the epochs also had an early stopping callback.

Quantitative Results

Table 2 shows the results of the studies in the form of a performance evaluation of the proposed models, both when visualized independently and in the form of an ensemble. The results show how the proposed model outperforms the existing approach by a significant margin in cancerous classes, where an average increase of 6% is observed. Based on the results, it is also important to mention that an ensemble of models did not improve the results significantly, and an individual model with an efficient net b3 backbone would achieve the same results as those of an ensembled model approach. The results clearly highlight that proposed modified ENB3 backbone network show improved results specifically for cancerous regions i.e. BCC, SCC and IEC. Our model performed most poorly on the *FOL* class with a recall of 0.67. This was primarily due to its unbalance in the dataset while also depending on the depth of the biopsy in the case of shave biopsies for example, some hair follicles may be included in the specimen obtained. However, since shave biopsies only remove a superficial layer of skin, the hair follicles may not be fully intact and may not provide a complete view of the hair follicle structure.

Layer	UNet	ENB3 backbone	Ensemble	Simon.et.al [24]
BKG	**0.99**	**0.99**	**0.99**	0.95
BCC	0.91	0.90	**0.91**	0.86
SCC	0.70	**0.86**	0.83	0.85
IEC	0.82	0.82	**0.83**	0.70
EPI	0.70	0.78	0.78	**0.83**
GLD	0.81	**0.89**	**0.89**	0.87
INF	0.64	**0.70**	**0.70**	0.57
RET	**0.91**	**0.91**	**0.91**	0.70
FOL	0.65	0.66	**0.67**	0.61
PAP	0.68	0.73	0.72	**0.80**
HYP	0.85	0.89	0.89	**0.96**
KER	0.79	**0.85**	0.83	0.84

Table 6.2: Class wise recall . The bold shows best results where as second best are shown with underline format.

Qualitative Analysis

The visual results of the experiment support the findings in the quantitative analysis. We also see how an efficient net b3 backbone greatly improves the results of the experiment. The model predicts the *BCC* class much better than others, this is again evident from the skewness in the dataset. The model however does sometimes confuse between *IEC* and *SCC* as both the classes had very little differences and many times coexisted in the same images. Segmentation of the cancers is also evident from the results shown in Figure 6.4 where the model shows how it performed on the cancer as classified by the pathologist.
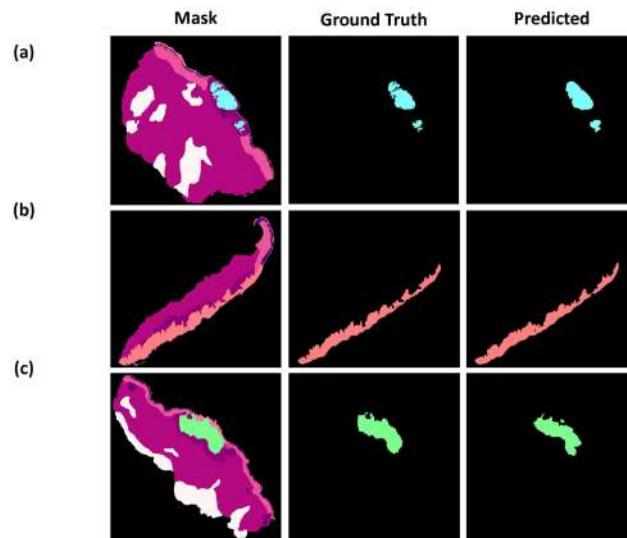
Figure 6.4: Isolating BCC, SCC, and IEC Masks in Different Models. a) Compares the Basal cell carcinoma labeling provided by a doctor, versus that of the trained model. b) Compares the Intra-epidermal cell carcinoma labeling provided by a doctor, versus that of the trained model. The IEC was difficult to detect as there was a lot of overlapping in it and the Epidermal class. c) Compares the Squamous cell carcinoma labeling provided by a doctor, versus that of the trained model.

For qualitative analysis in deep learning, as visualized in Figure 6.5, uncertainty maps are commonly used to visualize the output of a model and identify regions where the model needs improvement or additional data. It provides valuable information about how confident the models is about its predictions. For generating the maps, we have used the predictions from the softmax function that transforms

51

the output of the last convolutional layer into a probability distribution over the different 12 classes. The resulting probability map can be visualized as a heat map where the intensity of each pixel represents the confidence of the models prediction for that pixel. The formula for the softmax function is:

$$\text{softmax}(x) = \frac{e^x}{sum(e^x)} \tag{6.6}$$

where the x is the input to the softmax function, which is the output of the last convolutional layer of the model. The softmax function normalizes the output of the last convolutional layer into a probability distribution over the classes. After receiving the probability distribution we subtract it from 1 and with each pixel's maximum probability it is selected and then transformed based on the nipy-spectral color map to show uncertainty in each class.

$$P(x) = 1 - max(\text{softmax}(x)) \tag{6.7}$$

The heat maps shown again show that the model segments out the cancerous area with high precision and confidence, while however the layers where borders occur do show less confidence due to merging of the overlapping patches.

Comparison with literature

The proposed approach for segmentation of WSI skin images using deep learning was compared to the approach presented in Simon et al. [24] as shown in Table 6.2 and Table 6.3. The results showed that the proposed approach performed better in many cases, with an overall increase in accuracy and an improvement in average class accuracy. These results demonstrate the effectiveness of the proposed approach and its superiority over the existing approach in the literature. There is 5% and 10% increase in comparison to [24] for mean class accuracy and
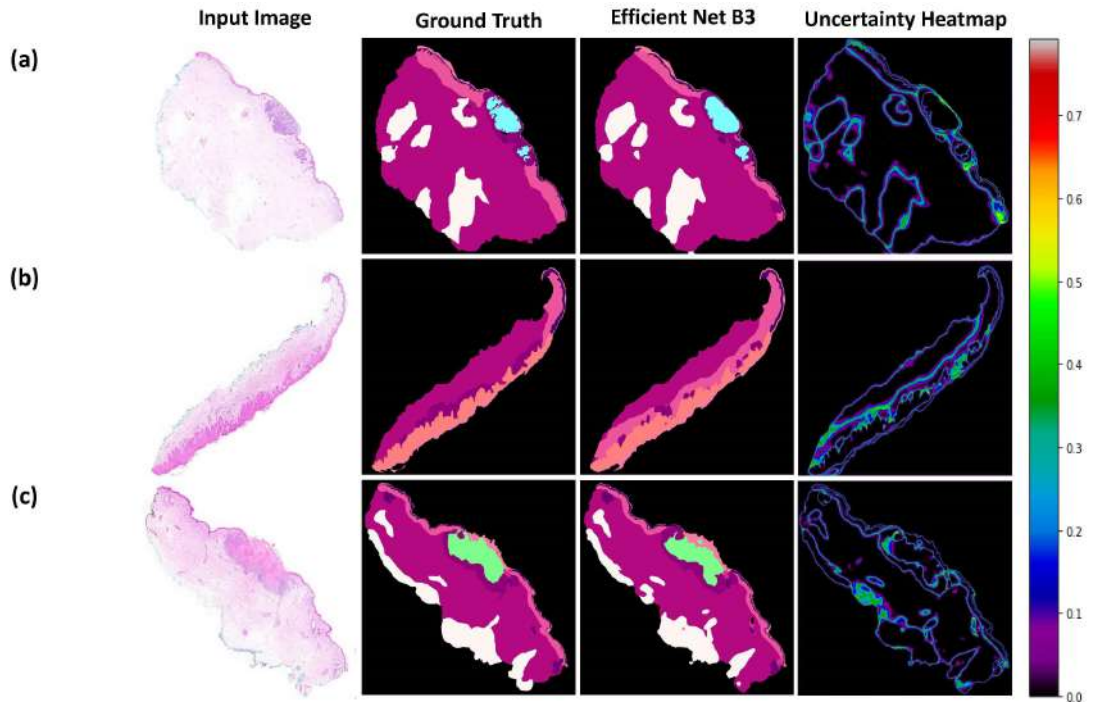
Figure 6.5: The model predictions with their respective uncertainty heatmaps.The Uncertainty heatmaps provided a highly interpret-able way of diagnosis, whereby a model's confidence indicates the probability for a physician to accurately judge the models work and where they need to pitch in for improved diagnoses. a) Generated Masks and uncertainty heatmaps with Predominant BCC Cancer Visually Represented. b) Generated Mask and uncertainty heatmaps with Predominant IEC Cancer Displayed. c) Generated Mask and uncertainty heatmaps Illustrating Predominant SCC Cancer.

overall accuracy respectively. The improvement in accuracy can be attributed to the use of an efficient net b3 backbone, which manages to extract and learn features more than the existent approach, and addressing class imbalance through data augmentation. These findings contribute to the ongoing research in the field and provide valuable insights for future studies.

Sr#	Model	F-1 Score	Accuracy	Mean Class Accuracy
1	Simon et al. [24]	-	0.85	0.79
2	U-Net	0.92	0.91	0.78
3	Efficient Net B3	0.93	0.95	0.83
4	Ensemble	0.94	0.95	0.84

Table 6.3: Comparison of models using different performance parameters

Chapter 7

DermaVision- A visual application for performing staining and segmentation on Whole Slide Images

7.1 Framework

The application is built using different tools for both front end and backend. For front end React.Js is used. And for deploying on backend Flask in Python is used.
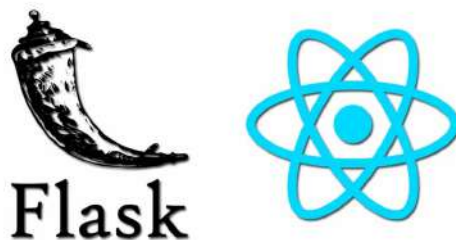
Figure 7.1: Web Frameworks used

7.2 Application

7.2.1 Starting page

At the landing page the user is asked to upload an unstained image. They are then provided with 2 options: if its an unstained image first stain it, if its an unstained image go for staining. If its already stained, go for analysis using segmentation.
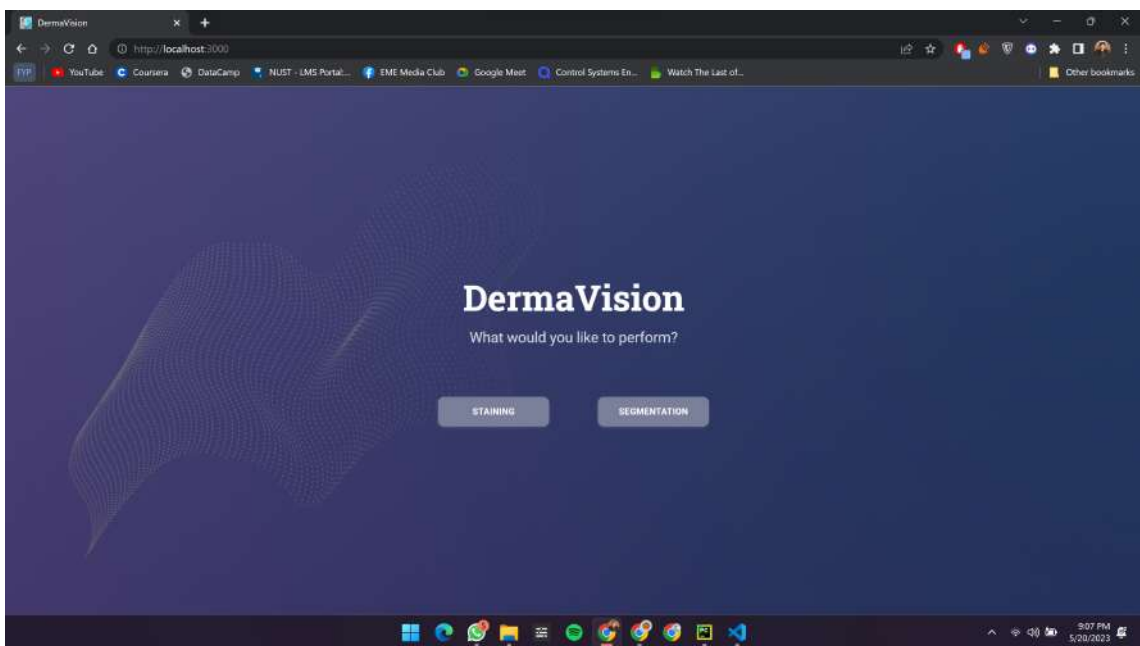
Figure 7.2: Web Frameworks used

7.2.2 Staining

If after opting for the staining and uploading the image, the image goes back to our server and gets processed through the GAN model converting unstained to stained images. First however, as explained above, patches are made from the input image for processing.
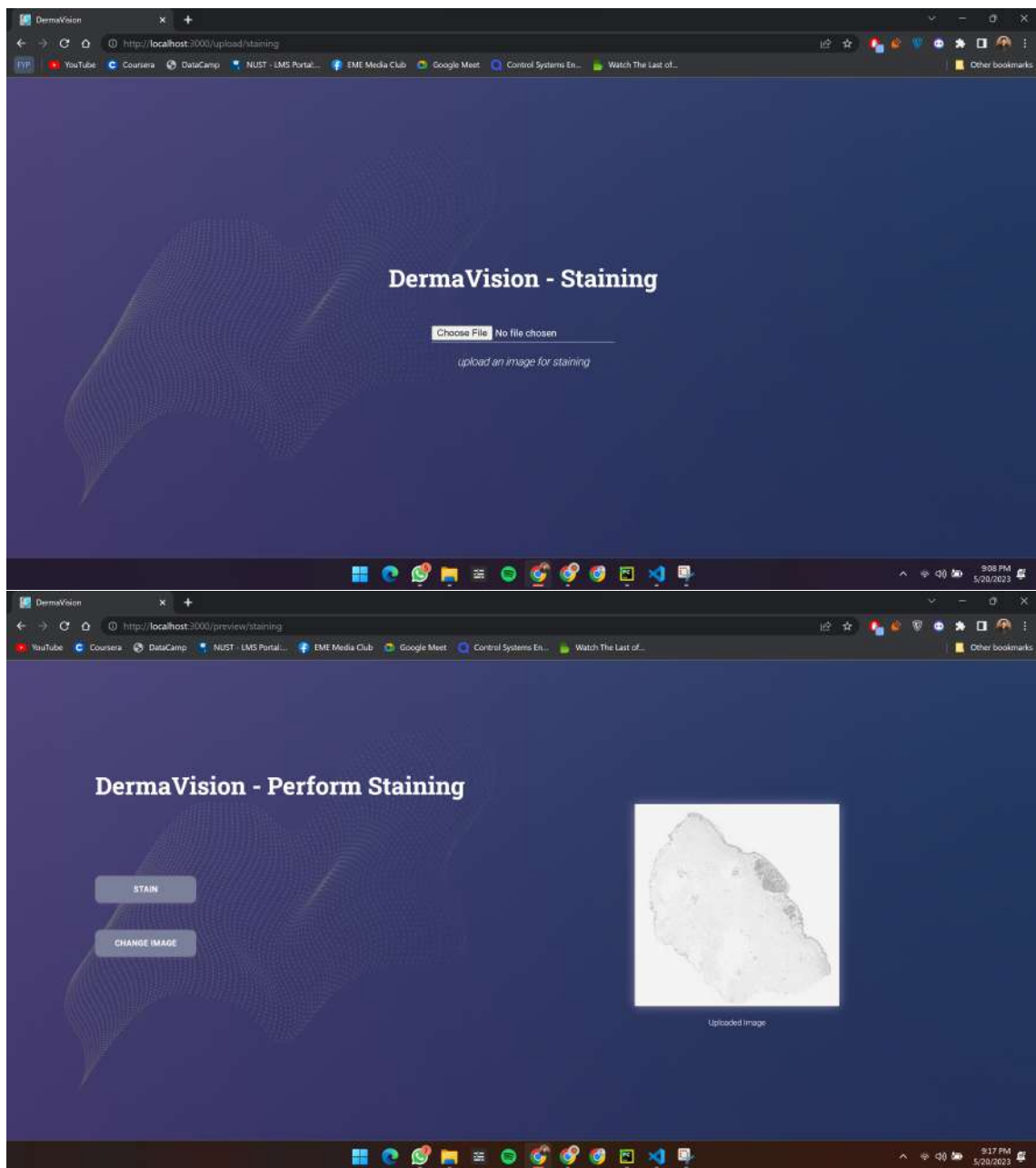
Figure 7.3: (a) Staining intro. (b) Staining File upload

7.2.3 Segmentation

If the user chooses to go for segmentation or decides to choose this option post staining, the input image patches go to the segmentation model. These are then later stitched back together in the smooth prediction approach [3]. From this the

heatmaps are also generated which are then shown to the user.

Figure 7.4: Segmentation process in the application.

7.2.4   Results

The final view of the application showcasing the unstained image along with its stained counterparts. The masks are also shown with its color key. The final image shows a heatmap that is showing uncertainity in predictions.

Figure 7.5: Final Result of the Application

Chapter 8

Conclusion and Future Work

8.1 Conclusion

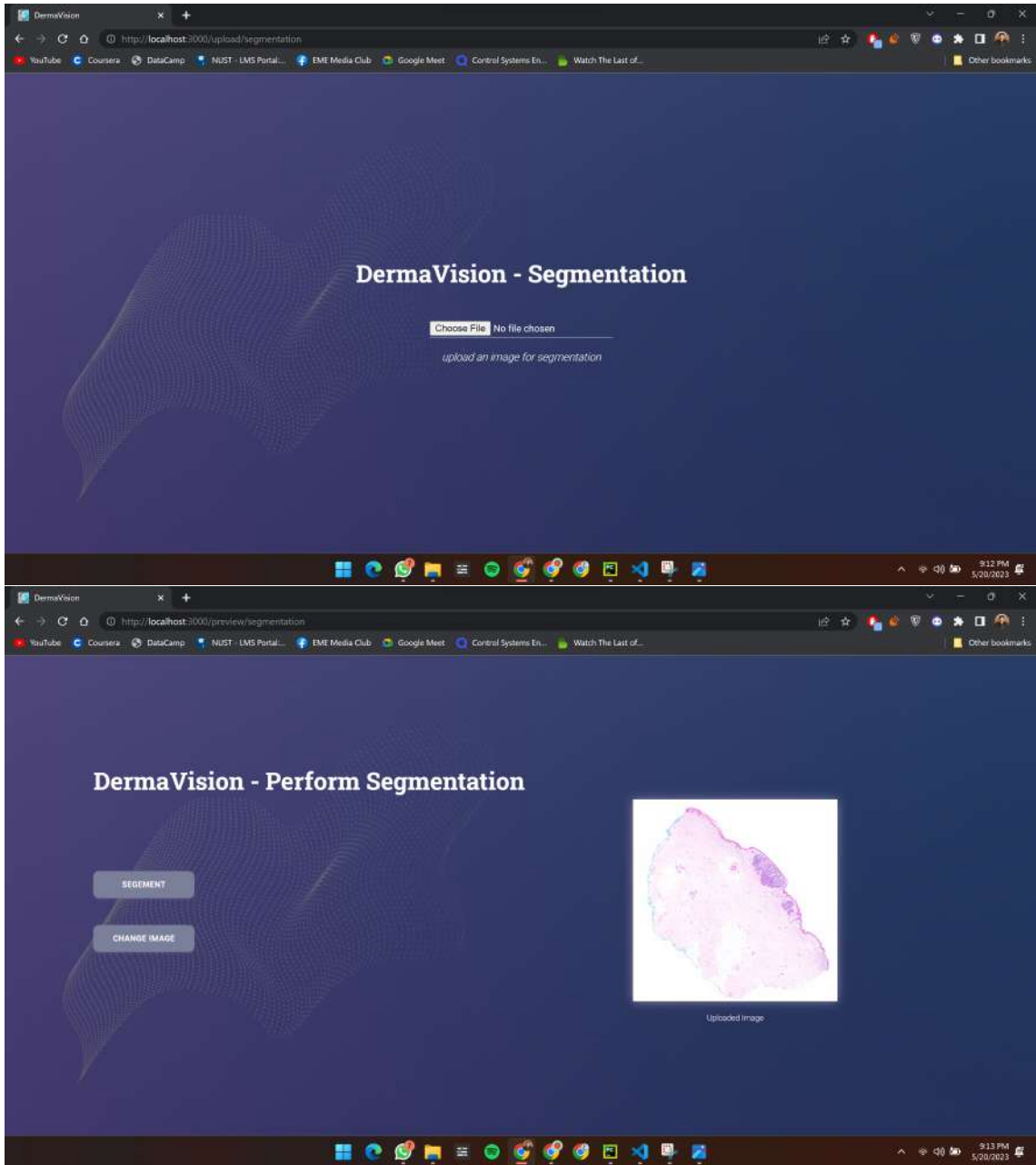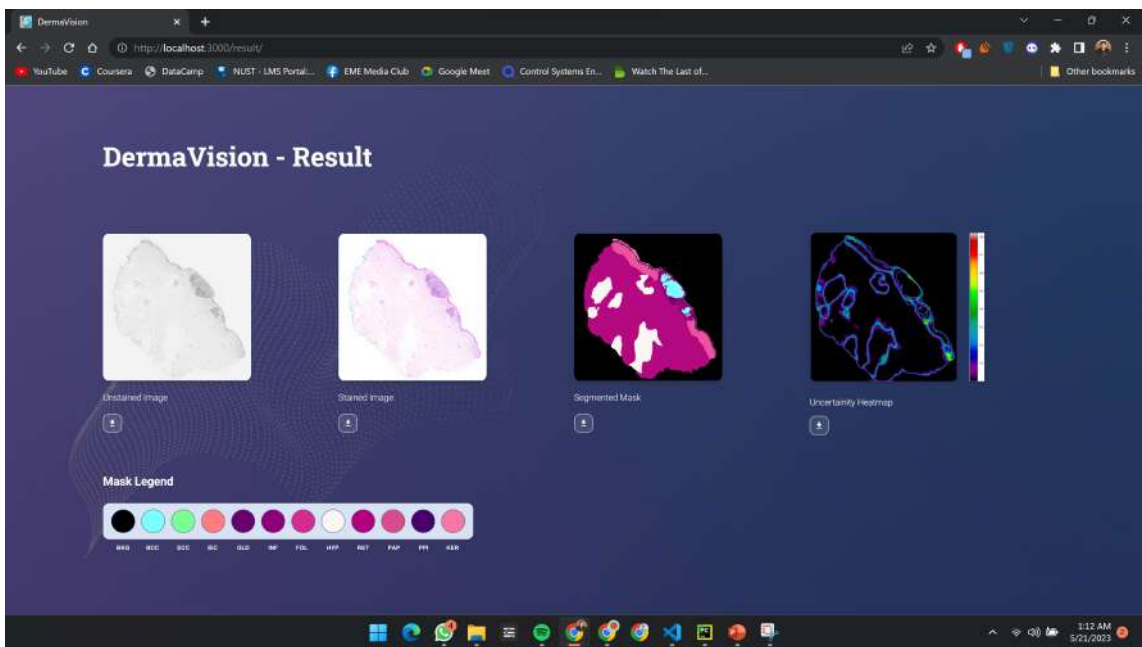In conclusion, our thesis report has explored the development of a cloud-based application for the segmentation and digital staining of whole slide images using deep learning models, specifically EfficientNet B3 for segmentation and DCL GANs for staining in the context of skin cancer. Through our research and analysis, we have demonstrated the potential of these models in improving the accuracy and efficiency of skin cancer diagnosis.

The use of deep learning models, such as EfficientNet B3, for automated segmentation of skin cancer regions in whole slide images can aid dermatologists and healthcare professionals in identifying and analyzing suspicious areas more effectively. By automating the segmentation process, the application can assist in speeding up the diagnostic workflow and reduce human error.

Furthermore, digital staining using DCL GANs provides an innovative approach to enhance the visualization of skin cancer patterns in histopathological images. By generating stained images that mimic traditional staining techniques, the application can facilitate a better understanding of the tissue structures and aid in

accurate diagnosis. This digital staining technique also offers the advantage of preserving the original slides while allowing for multiple virtual stains to be applied, enabling comprehensive analysis without the need for additional physical staining.

The cloud-based nature of the application provides several advantages, including scalability, accessibility, and cost-effectiveness. This enables healthcare institutions of varying sizes and resources to benefit from advanced image analysis techniques, improving access to accurate diagnosis and reducing the burden on healthcare professionals.

However, there are challenges that need to be addressed to ensure the successful implementation of cloud-based applications for skin cancer diagnosis. Issues related to data privacy, security, and ethical considerations must be carefully managed to protect patient information and maintain compliance with regulatory standards. Additionally, the integration of the application into existing clinical workflows and the establishment of robust validation protocols are crucial to ensure its practicality and reliability in real-world settings.

Our research has demonstrated the potential of a cloud-based application for the segmentation and digital staining of whole slide images in the context of skin cancer diagnosis. The combination of EfficientNet B3 for segmentation and DCL GANs for staining holds promise in improving the accuracy, efficiency, and accessibility of skin cancer analysis. Further collaboration between researchers, dermatologists, and healthcare professionals is essential to refine and validate these cloud-based approaches, paving the way for improved diagnostic capabilities and patient care in the field of skin cancer.

8.2 Future Work

Future work for this thesis holds several opportunities to further enhance the cloud-based application for segmentation and digital staining of whole slide images using deep learning models, particularly in the context of skin cancer. Some potential areas for future research and development include:

Refinement of segmentation model: While EfficientNet B3 has shown promising results in skin cancer segmentation, further improvements can be made to enhance its accuracy and robustness. Exploring alternative deep learning architectures and incorporating advanced techniques like attention mechanisms or multi-scale analysis, can potentially improve the segmentation performance.

Stain transfer techniques: The application can benefit from integrating advanced stain transfer techniques to improve the digital staining process. Stain transfer involves adapting stains from one histopathological image to another, enabling the visualization of multiple staining types on a single slide. Exploring state-of-the-art stain transfer algorithms, such as CycleGAN or StarGAN, can enhance the versatility and flexibility of the digital staining capabilities of the application.

Clinical validation and integration: Conducting extensive clinical validation studies involving dermatologists and pathologists is essential to assess the real-world performance and reliability of the cloud-based application. Collaborating with medical professionals to collect feedback, evaluate the accuracy of the model against ground truth annotations, and conducting comparative studies with existing diagnostic methods will validate its effectiveness and establish its clinical utility.

Deployment and scalability: Implementing the cloud-based application in real-world clinical settings requires careful consideration of deployment strategies and scalability. Optimizing the application to handle large-scale datasets, ensuring

efficient utilization of cloud resources, and addressing potential latency issues are crucial steps towards enabling seamless integration into clinical workflows.

By addressing these areas of future work, the cloud-based application for segmentation and digital staining of whole slide images using deep learning models can continue to evolve, improving diagnostic accuracy, efficiency, and accessibility in the field of skin cancer diagnosis and treatment.

Bibliography

[1] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[3] Vooban, "Smoothly-blend-image-patches." `https://github.com/Vooban/Smoothly-Blend-Image-Patches`, 2017.

[4] H. Zheng, L. Yang, J. Chen, J. Han, Y. Zhang, P. Liang, Z. Zhao, C. Wang, and D. Z. Chen, "Biomedical image segmentation via representative annotation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 5901–5908, 2019.

[5] F. Bagheri, M. J. Tarokh, and M. Ziaratban, "Skin lesion segmentation from dermoscopic images by using mask r-cnn, retina-deeplab, and graph-based methods," *Biomedical Signal Processing and Control*, vol. 67, p. 102533, 2021.

[6] A. Wibowo, S. R. Purnama, P. W. Wirawan, and H. Rasyidi, "Lightweight encoder-decoder model for automatic skin lesion segmentation," *Informatics in Medicine Unlocked*, vol. 25, p. 100640, 2021.

[7] G. M. Kosgiker and A. Deshpande, "A novel segcap algorithm based enhanced segmentation of dermoscopic images of interest," *Materials Today: Proceedings*, vol. 51, pp. 779–787, 2022.

[8] R. Gu, L. Wang, and L. Zhang, "De-net: A deep edge network with boundary information for automatic skin lesion segmentation," *Neurocomputing*, vol. 468, pp. 71–84, 2022.

[9] N. Moradi and N. Mahdavi-Amiri, "Multi-class segmentation of skin lesions via joint dictionary learning," *Biomedical Signal Processing and Control*, vol. 68, p. 102787, 2021.

[10] M. Anthimopoulos, S. Christodoulidis, L. Ebner, T. Geiser, A. Christe, and S. Mougiakakou, "Semantic segmentation of pathological lung tissue with dilated fully convolutional networks," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 714–722, 2018.

[11] D. Li, H. Hui, Y. Zhang, W. Tong, F. Tian, X. Yang, J. Liu, Y. Chen, and J. Tian, "Deep learning for virtual histological staining of bright-field microscopic images of unlabeled carotid artery tissue," *Molecular imaging and biology*, vol. 22, no. 5, pp. 1301–1309, 2020.

[12] Z. Stankovic, B. D. Allen, J. Garcia, K. B. Jarvis, and M. Markl, "4d flow imaging with mri," *Cardiovascular diagnosis and therapy*, vol. 4, no. 2, p. 173, 2014.

[13] Y. Zhang, Y. Lu, W. Chen, Y. Chang, H. Gu, and B. Yu, "Msmanet: A multi-scale mesh aggregation network for brain tumor segmentation," *Applied Soft Computing*, vol. 110, p. 107733, 2021.

[14] K. B. Soulami, N. Kaabouch, M. N. Saidi, and A. Tamtaoui, "Breast cancer: One-stage automated detection, segmentation, and classification of digital

mammograms using unet model based-semantic segmentation," *Biomedical Signal Processing and Control*, vol. 66, p. 102481, 2021.

[15] H. Wu, W. Wang, J. Zhong, B. Lei, Z. Wen, and J. Qin, "Scs-net: A scale and context sensitive network for retinal vessel segmentation," *Medical Image Analysis*, vol. 70, p. 102025, 2021.

[16] S. Thomas and N. Hamilton, "Histopathology non-melanoma skin cancer segmentation dataset," 2021.

[17] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," pp. 1107–1110, 2009.

[18] R. Zhang, P. Isola, A. A. Efros, and E. Shechtman, "Evaluation metrics for conditional image generation," *International Journal of Computer Vision*, vol. 129, no. 6, pp. 1733–1752, 2021.

[19] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, pp. 6626–6637, 2017.

[20] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, "Regularized evolution for image classifier architecture search," in *Proceedings of the aaai conference on artificial intelligence*, vol. 33, pp. 4780–4789, 2019.

[21] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep networks with stochastic depth," in *European conference on computer vision*, pp. 646–661, Springer, 2016.

[22] S. Zagoruyko and N. Komodakis, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.

[23] Y. Huang, Y. Cheng, A. Bapna, O. Firat, D. Chen, M. Chen, H. Lee, J. Ngiam, Q. V. Le, Y. Wu, *et al.*, "Gpipe: Efficient training of giant neural networks using pipeline parallelism," *Advances in neural information processing systems*, vol. 32, 2019.

[24] S. M. Thomas, J. G. Lefevre, G. Baxter, and N. A. Hamilton, "Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer," *Medical Image Analysis*, vol. 68, p. 101915, 2021.