AI Assisted Mole Detection for Online Dermatology Triage in Telemedicine Settings

by

Debarpan Das

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Medical Sciences - Radiology and Diagnostic Imaging
University of Alberta

Abstract

Skin moles are one of the most commonly occurring dermatological conditions prevalent nowadays. Early identification and diagnosis of moles are absolutely crucial since they often turn out to be precursors to serious conditions such as melanoma, a dangerous type of skin cancer. Therefore, to ensure an efficient treatment of cases based on their severity, they need to be assessed systematically. In this thesis, we present an artificial intelligence (AI)-enabled triage tool to identify moles from images uploaded by patients to a teledermatology platform. The proposed approach employs NesT, one of the latest state-of-the-art transformer-based network for classification. Our system acts as a filter by sending a warning flag if a mole is detected. This can be used to help dermatologists set up consultation appointments in a physical setting by giving priority if the patient has a mole on their image.

A comparative study of the prediction performance of the different neural network models has been provided for different performance metrics of interest. The results presented in this thesis have been obtained from two sets of data, consisting of more than 26,000 clinical pictures with combined different dermatological conditions. Multiple experiments using different models yielded a macro-average recall value as high as 0.955, along with overall accuracy and macro-average precision values of 0.962 and 0.958, respectively.

Preface

This thesis was submitted as partial fulfillment of the degree Master of Science (MSc) in Radiology and Diagnostic Imaging at the University of Alberta. The thesis is an original work of Debarpan Das and was completed between September 2020 and August 2023. The material presented in this thesis is based on the following paper:

Debarpan Das, Elcin Ergin, Bruno Morel, Michelle Noga, Derek Emery, Kumaradevan Punithakumar. AI-assisted mole detection for online dermatology triage in telemedicine settings. Informatics in Medicine Unlocked, vol. 41, p. 101311, 2023.

Acknowledgements

This project could not have been possible without the consistent support from industrial collaborators, OROHealth Inc., and the funding support from Mitacs. I would like to thank Dr. Elcin Ergin and Bruno Morel for giving me the opportunity to work on this amazing project and Shanel Gauthier for her insightful contributions to this work. I got the opportunity to learn a lot from their data science team and I am very grateful to be a part of it.

I also would like to express my deep gratitude to my supervisors, Dr. Derek Emery and Dr. Kumaradevan Punithakumar, and to the Director of the Servier Virtual Cardiac Center, Dr. Michelle Noga. They helped to pave the course of this project in the right direction and it would not have been successful without their constant help and support.

Finally, I would like to thank my family and colleagues who constantly support and encourage me to achieve my goals and follow my dreams.

Table of Contents

List of Tables

List of Figures

Abbreviations

ABCD Asymmetry, Border, Color, and Diameter.

AI Artificial Intelligence.

API Application Programming Interface.

BDD Body Dysmorphic Disorder.

BiT Big Transfer.

CAD Computer-aided Diagnosis.

CNN Convolutional Neural Network.

ConvLSTM Convolutional LSTM.

FFN Feed Forward Network.

FL Focal Loss.

GCP Google Cloud Platform.

GLCM Gray Level Co-occurrence Matrix.

GPU Graphical Processing Unit.

Inception-V4 Inception Version 4.

ISIC International Skin Imaging Collaboration.

LN Layer Normalization.

LSTM Long Short-Term Memory network.

MHSA, MSA Multi-Head Self-Attention.

NesT Nested Hierarchical Transformer.

NLP Natural Language Processing.

PHI Public Health Information.

ResNet Residual Network.

RMSprop Root Mean Squared Propagation.

sgd Stochastic Gradient Descent.

SOTA State-of-the-art.

ViT Vision Transformer.

Chapter 1

Introduction

1.1 Overview

The skin is the largest organ of the human body, covering our entire frame. It performs the vital functions of protecting our internal organs from external physical, chemical, and biological agents, as well as helping in thermoregulation. However, because of its large surface area, bacteria, viruses, allergies, or inflammatory reactions cause diseases to spread. Aging, environmental, and genetic factors, as well as trauma, can lead to various skin-related conditions with over 3000 different entities identified in the literature [1][2]. Studies indicate that in 2013, about 85 million Americans had consulted a physician for at least one skin condition [3]. In the same year, skin conditions contributed to direct health care costs of $75 billion and lost opportunity costs of $11 billion. With a reference age of 75 years, it was found that on average, the number of years of potential life lost was 10.8 and a total of 22,953 deaths were recorded. Therefore, early detection and diagnosis of skin conditions are extremely crucial in the present day.

The focus of this thesis is primarily on the detection of moles. When pigmented cells or melanocytes start to grow in clusters, they usually lead to the development of a mole on the skin. Most adults have 10 to 40 moles on average. Although commonly occurring moles are not generally cancerous, individuals with small or several large moles have an increased risk of developing melanoma, one of the deadliest forms of

cancer [4]. It has a high mortality rate since it can easily spread to other parts of the body if it is not treated in its early stages. The American Cancer Society estimates that in 2023, 97,610 new melanomas will be diagnosed and 7,990 people are expected to die. From 2015 to 2019, the melanoma rate was found to have increased by 1% per year for women but had stabilized in men. Moles that persist into old age are found to have elevated chances of malignant degeneration [5]. Therefore, it is highly crucial for newly appeared moles to be detected, properly diagnosed, and treated as early as possible to ensure a successful treatment.

1.2 Challenges and goal

Moles, in general, can range from harmless common moles to atypical moles, and, in the worst-case scenario, melanoma. Due to such a wide range of conditions, it becomes difficult and concerning for patients without adequate knowledge to understand the severity of their conditions without consulting a dermatologist. However, there is a long wait time for the time period between when a referral is initiated by a primary care provider (PCP) and when a dermatologist sees the patient in Canada [6]. It was found that dermatology was the second most common type of specialty referral constituting about 9.3% of the total number of referrals just after general surgery [6]. However, it has a median wait time of 92 days (75th percentile = 174) [6]. This prolongs physical consultation with a dermatologist, which must be minimized in critical cases. Figure 1.1 shows a bar graph of wait times for different specializations in Canada.

The advent of telemedicine has given us the opportunity to solve this problem in a much more efficient way. With the help of teledermatology, patients can now directly get access to fast and reliable treatment of their conditions. Furthermore, the COVID pandemic played a vital role in promoting and popularizing the use of teledermatology, as many types of skin conditions can potentially be treated without much physical interaction with doctors. Patients can upload images taken by their mobile phones of

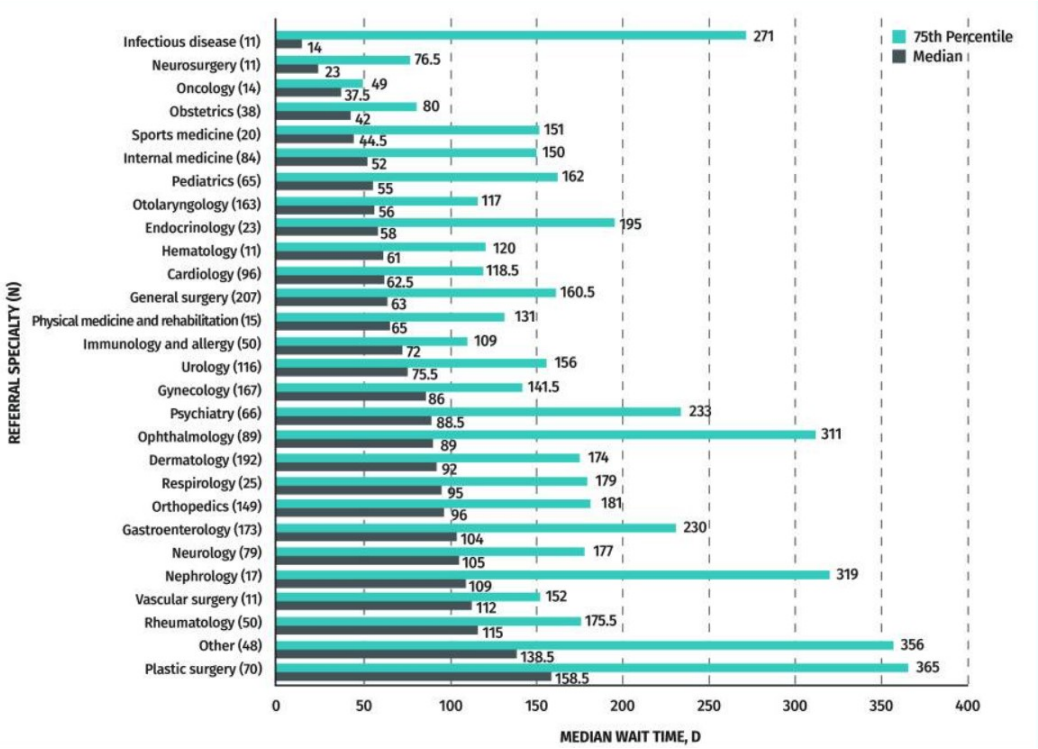


Figure 1.1: The wait time between a family doctor's referral to a specialist and the visit with the specialist, also called median wait time 1, for different specializations. Source: [6]

the affected skin region to an online platform and receive the corresponding diagnosis and treatment. However, unlike normal conditions, it becomes totally different when it comes to moles. Due to the potential risks associated with melanoma, they are preferably treated in person. It is absolutely necessary to minimize false negatives in such cases.

Hence, to mitigate these problems, it becomes necessary to make the system more efficient so that patients can first get a preliminary assessment of their skin condition, and if it turns out they have a mole, they should have an efficient way to get appointments with a dermatologist. This is possible by introducing a system in an online teledermatology platform that detects cases with moles. When a patient uploads pictures to the platform, the filter immediately sends a warning signal stating the potential presence of a mole and recommends that they consult a dermatologist. This would greatly help the triaging process and make the diagnosis and treatment procedure much more efficient.

1.3 Thesis contributions

In this thesis, we propose a fully automated deep learning-based Computer-Aided Diagnosis (CAD) system that identifies the presence of moles from patient-uploaded images. The entire workflow can be divided into the following two parts:

1. The first part of this thesis mainly concerns the application of a transformer-based mole filtering algorithm that helps to flag moles if they are present in uploaded images so that patients can be directed to book appointments for further procedures. We conduct extensive research and robust experiments [7] on latest state-of-the-art (SOTA) models such as Vision Transformers (ViT) [8], Big Transfer ResNetV2 (BiT) [9], Inception-v4 [10] and NesT [11] since they are found to perform much better than traditional machine learning-based models due to their more complex structures and highly optimized learning capabilities.

If a mole is found in an image, the patient can be immediately warned of it and recommended to consult a dermatologist in the physical setting since moles are treated essentially in person to mitigate any possible risks. This is the first study that leverages AI-assisted systems for screening moles from dermatology images and this can potentially help efficient case triage and expedite condition evaluation, making the entire diagnosis and treatment procedure streamlined and productive in a teledermatology setting.

2. The dataset used in this work predominantly consists of images obtained from people with comparatively lighter skin tones. Therefore, as a secondary contribution, in search of a more versatile algorithm we employed an explicit thresholding technique to detect human skin by classifying each pixel into foreground (skin) and background (non-skin). This is expected to play a vital role in improving the performance of subsequent models down the pipeline by making it more versatile and targeting a wider patient population.

1.4 Outline

This thesis is divided into five chapters:

- Chapter 2 reviews previous work on the application of traditional and CAD-based techniques in dermatology and methodologies developed for mole diagnosis and treatment. We also discuss the growth and scope of teledermatology and its application in treating mole related conditions.

- Chapter 3 consists of two parts: we first propose our fully automated CAD system for identifying moles from macro query images thereby paving the way for efficient triaging of conditions with moles. In the second part, we discuss a more versatile method of segmenting human skin from images.

- Chapter 4 presents the dataset that we used, the evaluation metrics, and the experimental results for the mole identification problem.
- Chapter 5 presents the results and performance of the human skin detection algorithm when applied to real-world images from a publicly available dataset.
- Chapter 6 concludes the thesis with a summary and presents potential future works.

Chapter 2

Literature Review

In this section, we present a detailed review of previous CAD based tools used in dermatology along with commonly used techniques used for the treatment and diagnosis of mole-related skin conditions. We also provide the literature on the advent of telemedicine and its applications in dermatology.

2.1 Dermatology and its significance

Dermatology is an essential and indispensable medical field in today's world. Skin conditions have emerged as one of the most notable public health concerns and one of the primary reasons for seeking medical advice through general practice [12]. Although some skin conditions, such as skin cancer, can be life-threatening, others may still impose a significant burden on patients in terms of poor quality of life and financial costs [13]. Overall, skin diseases rank as the fourth leading cause of non-fatal burden in terms of years lost due to disability [14].

Dermatologists specialize in diagnosing, treating, and preventing various skin conditions, ensuring that individuals maintain healthy skin and general well-being. Neglected skin disorders can have a profound impact on a person's quality of life, self-esteem, and mental health. Unfortunately, skin conditions are extremely prevalent, affecting people of all ages and backgrounds. Common ailments such as acne, eczema, psoriasis, fungal infections, and skin cancer can have significant physical and emo-

tional consequences. Dermatologists are trained to identify and manage these conditions effectively, providing relief to patients and improving their quality of life.

Another crucial aspect of dermatology is the early detection and treatment of skin cancer, a condition whose incidence rates have been steadily increasing worldwide [15][16][17][18][19]. Dermatologists play a central role in the identification and treatment of skin cancer, including the most aggressive form, melanoma. Although melanoma may frequently spread to other parts of the body, it is easily treatable if it is detected early. Regular skin screenings and prompt intervention from dermatologists increase the chances of successful treatment and reduce mortality rates. The key to finding melanoma early is to perform self-exams, the most common of which is the ABCDE method. The ABCD mnemonic was introduced by Friedman *et al.* [20] to educate primary health care physicians and the public about the early detection of melanoma. A stands for asymmetry since one half of the spot should be unlike the other, B for border since the border of the mole is supposed to be irregular, scalloped, or poorly defined, C for color: the mole should have varying colors from one area to the next, such as shades of tan, brown, or black, or areas of white, red, or blue, and D is for diameter as melanomas are usually greater than 6 millimeters, or about the size of a pencil eraser. However, when diagnosed, they can be smaller. The letter E for evolution was added by [21] to adhere to the cases of rapidly changing and appearing moles that can also indicate melanoma. Dermatologists use the ABCDE criteria to diagnose melanoma. These criteria have specific sensitivities and specificities: 57% and 72% for A, 57% and 71% for B, 65% and 59% for C, 90% and 63% for D, and 84% and 90% for E, respectively [22]. When combined, the criteria exhibit higher sensitivity (89.3% for 2 criteria and 65.5% for 3 criteria) but lower specificity (65.3% for 2 criteria and 81% for 3 criteria)[22]. Physicians specializing in cancer detection demonstrate good agreement in recognizing asymmetry, irregular borders, and haphazard color [23]. The high sensitivity and inter-observer concordance support using ABCDEs as a screening tool, and dermatologists consistently show accurate

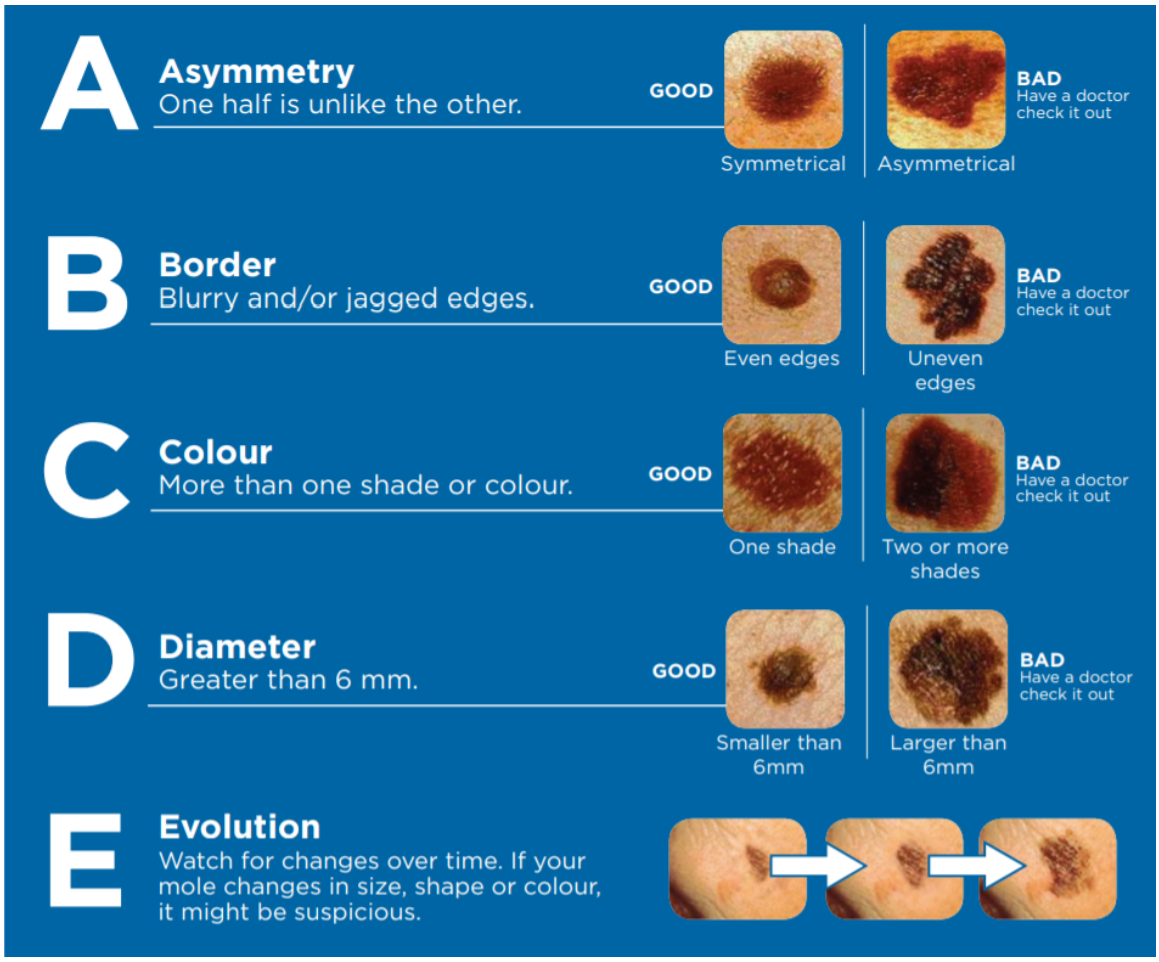


Figure 2.1: The ABCDE method of early detection of melanoma. Image source: Canadian skin cancer foundation [26]

detection of early melanoma [24][25]. Figure 2.1 illustrates the ABCDE method for self-examination of melanoma.

Additionally, dermatologists specializing in cosmetic dermatology offer a variety of treatments aimed at improving the appearance of skin, hair and nails. From injectables like Botox and dermal fillers to laser therapies and chemical peels, these procedures help people achieve their aesthetic goals safely and effectively. Especially with the dramatic increase in cases of body dysmorphic disorder (BDD), the number of patients seeking cosmetic procedures has increased dramatically [27]. BDD is a psychiatric condition in which the individual is under the perception of a certain imaginary physical defect in their body's image [28][29][30]. United States, Brazil,

South Korea, India, and Mexico are the top five countries with the highest number of surgical and nonsurgical procedures[31]. In such cases, dermatologists must ensure that cosmetic procedures are performed only after a proper screening with the utmost care, prioritizing patient safety. On the other hand, cosmetic surgery for aging is a rapidly growing domain in dermatology. Advancements in laser surgery, chemical peeling, liposuction, and phlebology have ushered in new techniques to remove signs of aging that were not possible before. However, the term aging can be defined from both physical and mental perspectives. When aging is considered a pathological process rather than a physiological process, dissatisfaction becomes inevitable keeping in mind the current day's youth promotion. In such a case, the treatment of an aging face [32] is implied with a morality similar to that of the treatment of tuberculosis or hypertension. However, such cases must be handled only after proper screening and following strict guidelines. Alternatively, with increasing accessibility, cosmetic surgery is also used nowadays as an instant alternative to psychotherapy for people with low self-esteem. Studies such as [33][34] reveal considerable psychological improvements in patients, and such cases are cited as proof of the medical value of cosmetic procedures. Hence, as a psychiatric treatment, cosmetic surgery is often used as a legitimate treatment today.

Dermatology also includes the recognition of skin manifestations of systemic diseases. Many systemic conditions can manifest specific skin symptoms that dermatologists are trained to identify. This ability to link skin findings with underlying systemic conditions contributes to a comprehensive approach to patient care, as dermatologists work together with other medical disciplines to diagnose and manage these complex cases.

Studies presented by Thomas[35] show that occupational skin disorders are among the most common occupational diseases. An occupational disease is defined as a disease that a worker develops as a result of their insured work activity. According to German research [36], the rate of newly diagnosed occupational skin diseases in

Germany is 0.67 to 0.68 to 1000 per year. The most common form of occupational skin disease is contact dermatitis, which depends on several factors, such as etiology and irritants. Early detection and rapid start-up of treatment are essential to treat such conditions and prevent the potential onset of severe occupational skin diseases and their associated social and economic consequences. Effective management requires collaboration between dermatologists, occupational and health professionals, and preventive services offered through statutory worker compensation programs. Additionally, specific diagnostic tests play an essential role in the identification and treatment of these occupational skin diseases.

Lastly, the field of dermatology has witnessed significant advancements in research, diagnostic tools, and treatment options. Dermatologists stay up-to-date with the latest developments and incorporate evidence-based practices into their work. This allows them to provide the best possible care to their patients, utilizing innovative therapies and technologies.

In conclusion, dermatology is critical in modern times due to the increasing prevalence of skin conditions, the importance of early detection and treatment of skin cancer, the demand for cosmetic procedures, the recognition of skin manifestations of systemic diseases, and the continuous advancements in dermatological research and technology. Dermatologists play a vital role in promoting skin health, improving quality of life, and addressing the diverse needs of patients.

2.2 Traditional feature-based methods

Over the years, there have been many advances in the development of automatic CAD systems to aid dermatologists in their diagnosis. Older systems employ traditional hand-crafted features such as texture, size, shape, and color to identify skin conditions [37]. For instance, the gray level co-occurrence matrix (GLCM) is used to analyze the skin image in detail [38]. Different combinations of GLCM features such as variance, energy, correlation, homogeneity, contrast, and entropy are used to detect melanoma

skin cancer [39]. Psoriasis detection techniques [40] use color features such as the first-order moment, the mean color, the second-order moment, the standard deviation, and the third-order moment, skewness of color along with the aforementioned textural features. Dermoscopic features such as the asymmetry, border, color, and diameter (ABCD) of skin lesions, along with some texture-based and statistical parameters such as mean, skewness, and kurtosis extracted from GLCM for skin cancer detection [41]. Classical image processing techniques [42] such as discrete cosine transform [43], discrete wavelet transform [44][45] and singular value decomposition to extract feature vectors were employed to detect warts, tinea corporis, acne, vitiligo, nail psoriasis, and eczema.

2.3 AI in Dermatology

During the initial advent of AI, its applications in the field of medicine were unpopular. There were several factors contributing to this problem, such as local extremum and gradient dispersion problems, insufficient hardware problems, and most critical of all, lack of large amounts of labeled data. Despite the revolutionary progress of AI in diverse domains such as image processing, speech recognition, text processing, and other areas in the last decade, its main application was the in vivo classification of benign and malignant lesions [46]. However, in comparison, radiological AI has far surpassed dermatological AI. Different applications of AI in radiology have been widely accepted and incorporated into various healthcare settings. This has greatly helped to bring doctors at different levels closer and has also helped improve the precision of diagnoses. Radiological applications of AI have also greatly benefited and promoted the development and acceptance of AI in dermatology as well.

2.3.1 Classical AI methodologies

In 2017, Esteva *et al.*[47] published a study that applied deep learning to skin tumors. They trained a CNN using a dataset of 129,450 clinical images with 2032 different

diseases. The CNN was trained to learn fine-grained features for pixel-wise disease labeling. The trained model was validated by performing classification tasks such as "keratinocyte carcinomas vs. benign seborrheic keratosis" and "malignant melanomas vs. benign nevi." The predictions were compared with the diagnoses from more than 21 board-certified dermatologists, and the model succeeded in performing on par with all experts tested across both tasks as shown in Figure 2.2, thereby proving that AI assisted tools can be used to classify skin cancer with a level of competence comparable to dermatologists. A meta-analysis of 70 studies conducted by Dick *et al.* [48] revealed that the accuracy of computer-aided diagnostic tools for melanoma is comparable to that of human experts. Consequently, studies [49][50][51] reveal that classification of malignant versus benign melanoma from non-dermoscopic images predominantly curated from East Asian patients reported an area under receiver operating characteristic curve of 0.94 for the detection of malignancy among 134 conditions. This also turned out to be on par with dermatology residents. CNNs have also achieved expert-level diagnosis accuracy in non-pigmented skin cancer [52] and outperformed dermatologists across five disease classes [53]. Even when the modalities are switched from dermoscopic to non-dermoscopic, CNNs managed to achieve dermatologist-level performance in the classification of melanoma [54].

Thus, gradually over the course of time, the implementation of CAD in dermatology, empowered by AI, has been initiated by governmental, industrial, and academic investments. Given the rapid pace of research and development in this domain, nowadays we generally use modern learning-based methods which employ models trained on huge datasets.

2.3.2 Modern deep learning-based methods

Modern systems extensively use deep learning-based algorithms to provide computer-assisted platforms to dermatologists, primarily because their performances are much better and more reliable. Instead of relying on specific hand-crafted features, deep

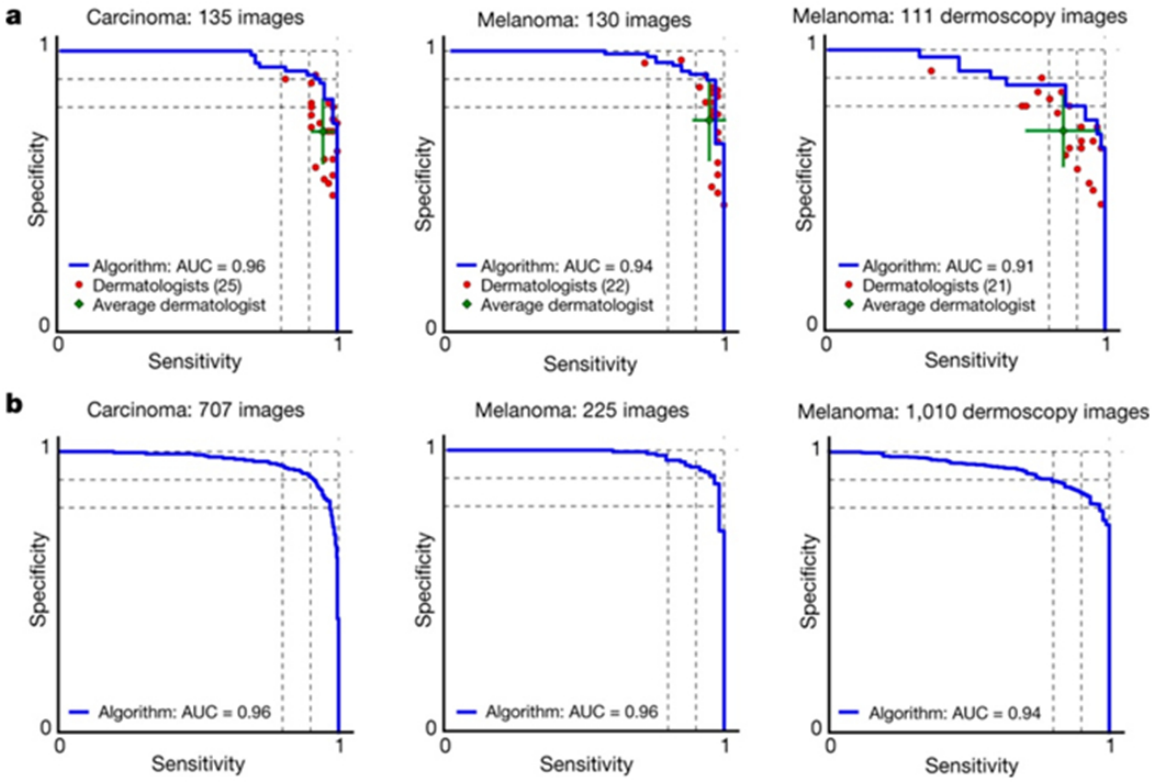


Figure 2.2: The figure depicts the sensitivity-specificity curve of the CNN as well as the dermatologists for different cases of carcinoma classification and melanoma classification. **(a)** A red dot denotes a prediction by a dermatologist per image whereas a green dot represents the mean of the dermatologist for each task, with variation bars denoting one standard deviation (calculated from 25, 22, and 21 dermatologists for keratinocyte carcinoma and melanoma from general and dermoscopy images respectively). Three cases have been experimented upon, Epidermal test with 65 keratinocyte carcinomas and 70 benign seborrheic keratoses, Melanocytic test with 33 malignant melanomas and 97 benign nevi and a second melanocytic test using 71 malignant and 40 benign dermoscopic images. Since the blue curve is above most of the green dots, it reflects the better performance of the model with respect to dermatologists. **(b)** the CNN's performance on a larger dataset has been depicted where the curves are much smoother reflecting robust classification. The above illustration is taken from [47].

learning methods can automatically learn semantic features from large and diverse datasets. As discussed earlier, classical deep CNN based architectures are utilized to develop a skin lesion classification system [47]. MobileNetV2 [55], and LSTM [56] have been employed to develop a system used for skin disease forecasting [57]. Chen *et al.* [58] proposed a self-learning-based artificial intelligence (AI) platform that uses a wide collection of data in a closed-loop flow from users and remote medical data centers for real-time skin disease recognition. In this work, LeNet-5 [59], AlexNet [60], and VGG16 [61] were adopted to carry out the training, classification, and assessment processes. Jiang *et al.* [62] proposed a U-Net [63] inspired model from skin lesion segmentation in which they added a squeeze-and-excitation layer after every convolution to enhance informative features, along with integrating a convolutional LSTM (ConvLSTM) [64] to improve sensitivity and prediction accuracy. SkinVision employed a conditional Generative Adversarial Network [65] together with inpainting [66] for noise reduction. The segmented regions are further explored and their features (shapes, colors, texture) are classified using an SVM classifier. Liu *et al.*[67] developed a deep learning system using inception-v4 modules to generate two types of differential diagnoses. The primary output comprised of the relative likelihood of 27 skin conditions, and the secondary output consisted of the relative likelihood of 419 conditions. The model was trained on a dataset of 16,114 de-identified (photographs and clinical data) cases acquired from a teledermatology organization and the output conditions were chosen based on granularity to assist non-dermatologist clinicians in conducting preliminary clinical care. Cassidy *et al.*[68], on the other hand, discussed in detail the different benchmark algorithms on the International Skin Imaging Collaboration (ISIC) datasets. An in-depth analysis of existing issues in the dataset is provided along with a strategy to remove duplicate images from the dataset to prevent biasing issues.

2.4 Teledermatology

CAD tools for dermatology have become much more relevant, especially in the recent COVID pandemic lockdown. Telemedicine became quite popular as patients turned to online means to receive diagnoses and treatment for specific conditions. The user satisfaction in an AI-assisted teledermatology service to redefine melanoma screening using a pigmented lesion evaluator was found to exceed 75% [69]. Unfortunately, certain conditions, especially moles, require extensive diagnostic procedures and are still preferably treated in person, where dermatologists use specific specialized tools such as dermoscopy for imaging. Melanocytic nevi or moles that persist into old age have elevated chances of malignant degeneration [5]. Therefore, it is highly crucial that newly appearing moles be detected and adequately diagnosed and treated. Birkenfeld *et al.*[70] trained a logistic regression algorithm on the characteristics corresponding to different components of the ABCD criteria to classify skin lesions according to a suspiciousness score. On the other hand, Cavalcanti *et al.*[71] proposed a three-step approach to segment melanocytic skin lesions. Although face-to-face diagnostic accuracy surpasses that of teledermatology settings [72], a lack of accessibility in obtaining reliable dermatological care makes it a viable solution. It has the potential to make services more accessible to patients through referrals and reduced wait times, and to provide diagnostic and triage support for cases under review.

As an extension to teledermatology, AI-assisted tools can also be used to develop mobile applications and personal monitoring devices. The detection of melanoma, one of the deadliest types of cancer, can be made much more accessible through mobile applications, especially for people in rural areas with limited availability of dermatologists. In the present day scenario, there are two types of mobile application services for teledermatology [73]: store-and-forward teledermatology and automated smartphone applications. The collaborating partner of this work, OROHealth[1] Inc.

[1]https://orohealth.me/en

falls in the former category. In their setting, they send pictures of lesions taken by patients through their mobile application to dermatologists located remotely for their diagnosis. On the other hand, automated smartphones are developed to provide a probability of malignancy of melanoma on the spot based on images and other information entered by the user. These tools have been shown to have a significant impact on outreach [74] and provide viable solutions to patients in need.

Chapter 3

Materials and Methods

In this section, we will discuss the proposed end-to-end pipeline of the fully automated mole detection CAD system. We elaborate the different models we experimented with to compare their performance in the search for the best model for the task at hand. Additionally, we will also include a module describing a more versatile, yet classical method of human skin detection that encompasses a wide range of skin-tones.

3.1 CNN vs transformer

Convolutional neural networks (CNN) have always played the most pivotal role in computer vision. For years, CNNs have been the dominant method of processing and classifying image-related data. CNNs were specifically designed for image recognition. They are composed of multiple layers, including convolutional layers, pooling layers, and fully connected layers. When these filters are applied repeatedly to the pixels in an image, they build up a series of channels that help them recognize features. Therefore, a CNN learns the data pixel-by-pixel, identifying features of the image thereby building its way from local to global. They excel at learning spatial patterns in images and extracting features that are relevant for classification.

On the other hand, transformers [76] were originally developed for natural language processing (NLP) tasks, but have been recently adapted for computer vision tasks such as image classification. Transformers use self-attention mechanisms to

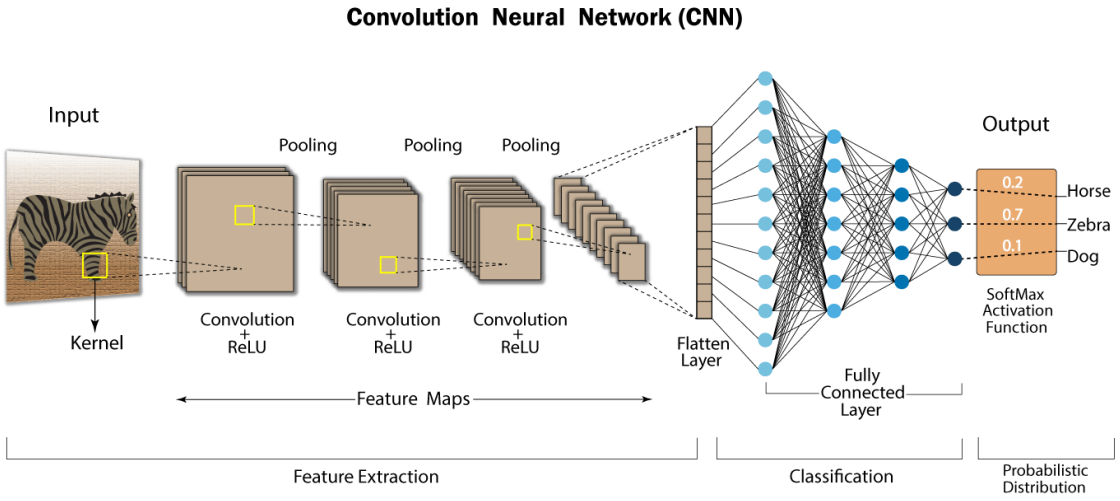


Figure 3.1: A convolutional neural network successively applies a series of filters to an image thereby learning the features of the image, a zebra in this case. The above illustration is taken from Medium [75].

learn global relationships between different parts of input data. They are capable of capturing long-range dependencies in images, which can be useful for recognizing complex patterns. Therefore, unlike CNNs, transformers can make connections between distant image locations (just as in language) with the help of self-attention.

The advent of Vision transformer[8] or ViT ushered in a new set of transformer based models which could be applied confidently in vision related-problems. Its architecture is similar to that of the original transformer (2017) with minor tweaks that allow it to process image patches of a fixed size instead of words. ViT classified images from the ImageNet[60] dataset with over 90% accuracy achieving a SOTA performance in the ImageNet classification challenge.

3.2 Vision transformers (ViT)

Vision Transformer or ViT, as mentioned above, replaces the convolutional layers commonly used in computer vision tasks with the self-attention mechanism used in Transformers. In this section, we discuss the architectural details of the ViT model in detail.

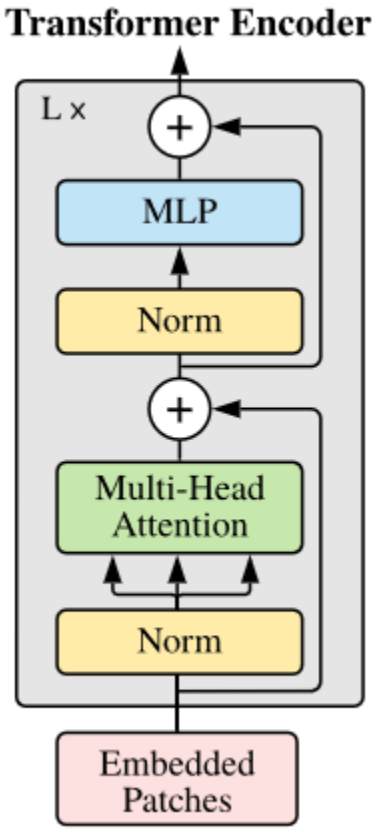


Figure 3.2: A transformer block representation. Norm refers to a normalization layer, Multi-Head Attention is the self-attention layer and MLP is a fully connected layer. The plus signs represent some operation (e.g. concatenation) of an output with a residual connection. Source : Picsellia [77]

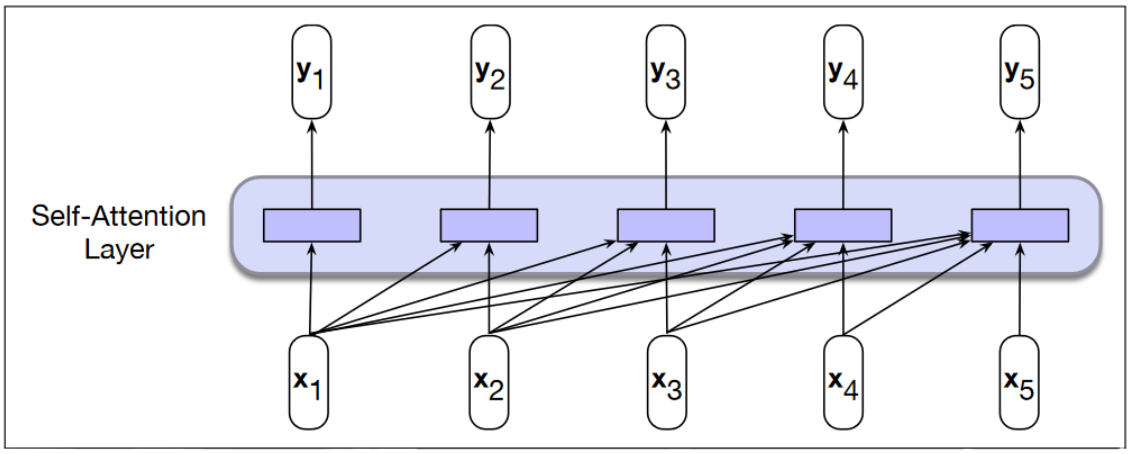


Figure 3.3: Mapping an input sequence $\vec{x}$ to an output sequence $\vec{y}$ through self-attention. All of the preceding inputs $x_i (i \leq j)$ including the one under investigation are used to produce an output y_j. Source : Picsellia [77]

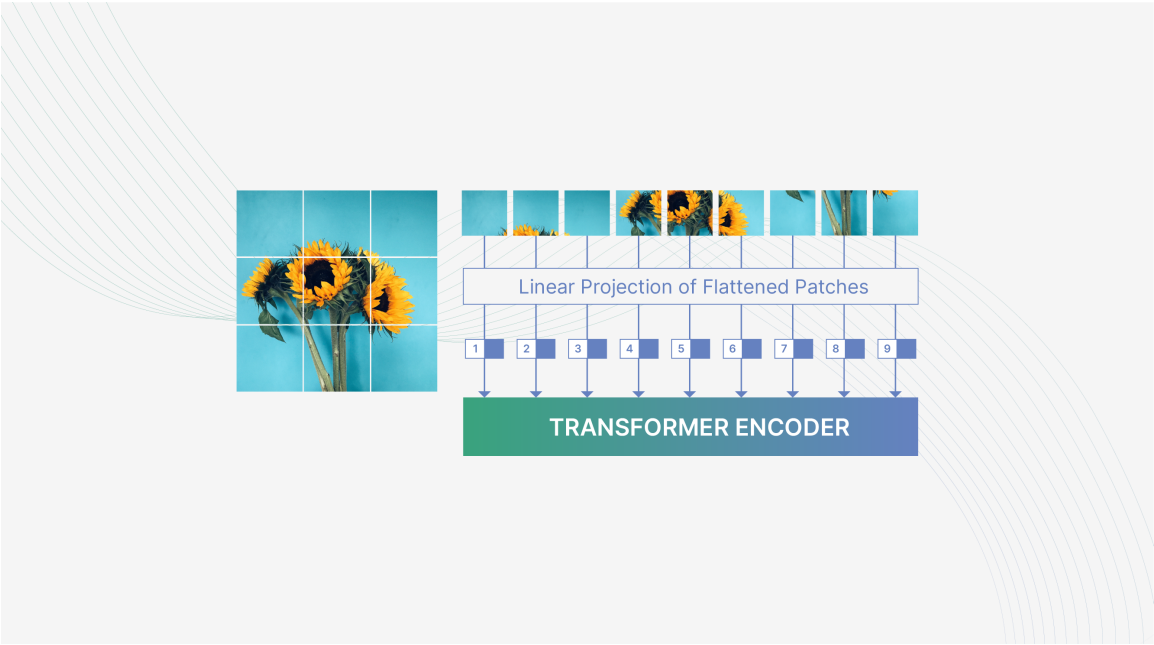


Figure 3.4: From a structured image to sequential data. The image is split into patches, patches are flattened and projected to another space. Finally, a positional embedding is concatenated to keep track of the spatial order, before feeding the input to the transformer. Source : Picsellia [77]

1. Patch Embeddings: The input image is first divided into non-overlapping patches of fixed size, typically 16x16 pixels, and each patch is linearly projected into a lower-dimensional vector representation. This process results in a sequence of 1D patch embeddings.

2. Position Embeddings: Each patch embedding is then augmented with a learnable position embedding, which encodes its relative position within the image. This is similar to the position encoding used in the Transformer model for NLP.

3. Transformer Encoder: The patch embeddings with position embeddings are then fed into a series of Transformer Encoder layers. Each Transformer Encoder layer consists of a Multi-Head Self-Attention (MHSA) mechanism and a Position-wise Feed-Forward Network (FFN). The MHSA mechanism allows the model to attend to different parts of the image patches, while the FFN provides nonlinear transformations.

4. Classification Head: After passing through the Transformer Encoder layers, the final patch embeddings are passed through a linear layer to produce a sequence of class tokens, which are then fed into a classification head to produce the final output.

ViT has achieved SOTA performance on several benchmark image classification datasets, demonstrating that it can compete with or even outperform CNNs on certain tasks. One advantage of ViT is its ability to capture global dependencies in the input image, which is particularly useful for recognizing objects in context or understanding the spatial relationships between objects.

Overall, ViT represents a promising direction in computer vision research, and it has the potential to be applied to a wide range of other computer vision tasks beyond image classification.

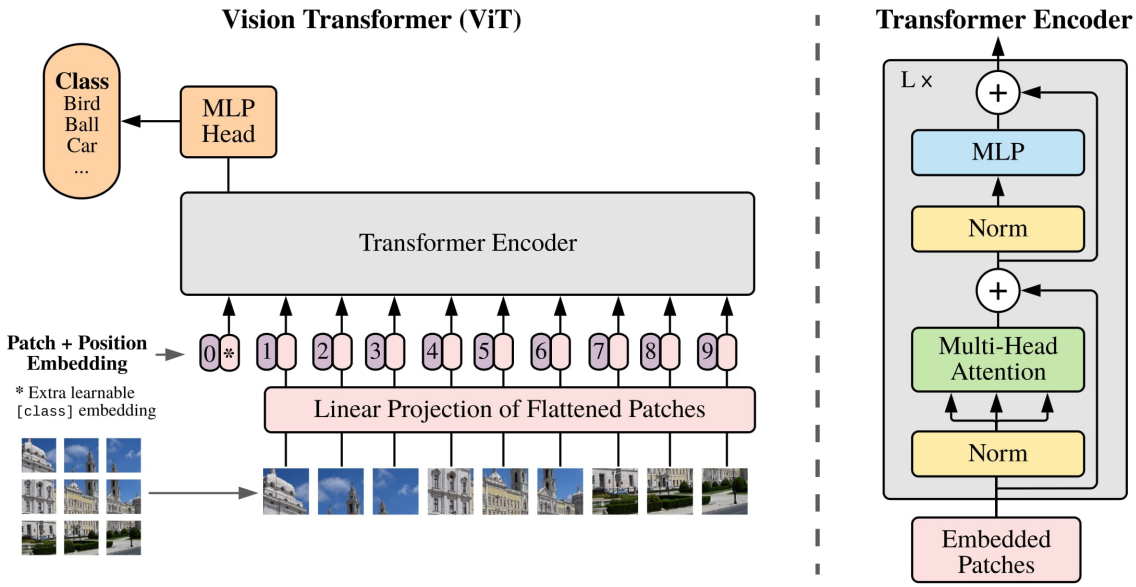


Figure 3.5: Vision transformer (ViT) architecture. Source : original paper [8]

3.3 Nested hierarchical transformer (NesT)

NesT expands to the nested hierarchical transformer, which is essentially an improved version of ViT. It employs the notion of hierarchically nesting local transformers, which operate on non-overlapping image blocks. The proposed block aggregation not only outperforms previously studied self-attention mechanisms but also enables achieving a substantially simplified architecture with superior data efficiency. It also helps in efficient cross-block non-local information communication that helps the model to converge faster and requires much less training data. The overall architecture is illustrated in Figure 3.6.

Given an input image of dimensions $H \times W \times 3$, we divide it into smaller patches of resolution $S \times S$. Each patch is linearly projected into an embedding. All such embeddings are divided into blocks and flattened to generate the inputs $X \in \mathbb{R}^{b \times T_n \times n \times d}$ where b is the batch size, T_n is the total number of blocks at the bottom of the hierarchy, n is the number of embeddings in each block, and d is the dimension length of each embedding in $\mathbb{R}^d$. Thus, $T_n \times n = H \times W/S^2$. Trainable positional embeddings [78] are concatenated to every block embedding before feeding them to the block func-

tion T. Each canonical block houses three layers: multi-head self-attention (MSA) [76] layer followed by feed-forward fully connected layer (FFN) with skip-connections [79] and layer normalization [80]. Since these transformer layers are canonical, all blocks at the same hierarchy level share the same set of parameters, so the MSA layer is applied parallelly as MSA_{NesT} to all blocks at the same hierarchy level.

$$MSA_{NesT}(Q, K, V) = stack(block_1, block_2, ..., block_{T_n}) \tag{3.1}$$

where,

$$block_i = MSA(Q, K, V)W^0 \tag{3.2}$$

Here, $block_i$ is of dimensions $b \times n \times d$. The FFN layer consists of two layers collectively given by:

$$FFN = max(0, xW_1 + b)W_2 + b \tag{3.3}$$

Therefore, the overall block function T is given by:

$$T = multiple\ X \begin{cases} y = x + MSA_{NesT}(x', x', x'), & \text{when } x' = LN(x) \\ x = y + FFN(LN(y)) \end{cases} \tag{3.4}$$

Thus, every block independently processes local information and only shares global information in the block aggregation step. This step shares information from spatially connected blocks in the image plane through simple operations such as convolution, layer normalization (LN), or max pooling. Therefore, the output in hierarchy l is unblocked to the image plane for spatial operations. After sharing spatially local information and down-sampling the feature maps, they are blocked back in hierarchy l + *1* such that the number of sequences n remains the same, but the total number of blocks gets reduced. This process is continued until the total number of blocks reduces to 1 in the top-most hierarchy of the model. Therefore, not only does this architecture solve cross-block communication in local self-attention, but it is also simple due to its canonical stacking of primary transformer layers.

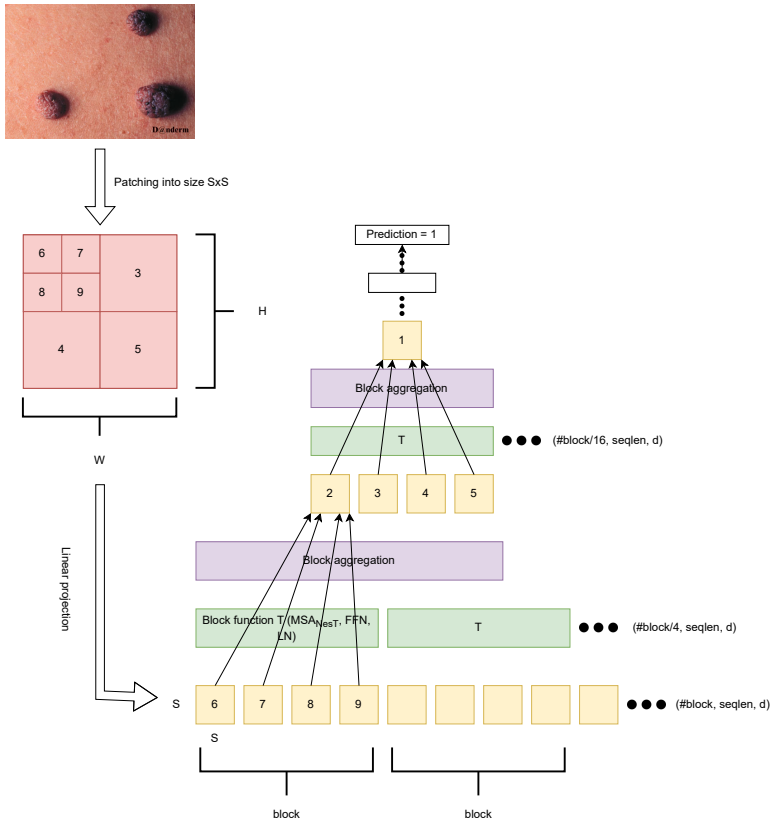


Figure 3.6: NesT illustration with nested transformer hierarchy. Linearly projected patch embeddings from the image are divided into blocks stacked with canonical transformers at every hierarchy. Spatially neighboring blocks are aggregated by simple spatial operations in the image plane, creating the hierarchical structure.

3.4 Big transfer (BiT)

The BiT[9] model is a family of deep neural network models designed for transfer learning in computer vision. The models are based on ResNet (Residual Network) architecture, which is a popular SOTA deep learning model for image classification. The main idea behind BiT is to pretrain a large ResNet model on a large and diverse dataset, such as the ImageNet dataset, and then transfer the learned representations to downstream tasks.

There are 2 main types of components of BiT models: upstream components are those which are used in the pre-training phase, and downstream components are those which are used during fine-tuning the model to a new task.

The BiT architecture consists of a series of building blocks, each of which contains a combination of convolutional layers, batch normalization, ReLU activation functions, and shortcut connections. Building blocks are organized into several stages, with each stage consisting of several building blocks that downsample the spatial dimensions of the feature maps. So essentially it is quite similar to that of ResNetv2, but with two key differences. First, group normalization is used in place of batch normalization in all layers. Second, weight standardization is applied to the convolutional layers. These modifications have been shown by the authors to be beneficial for training models with large batch sizes and have a considerable impact on transfer learning performance.

In addition to the ResNet building blocks, the BiT models also include several modifications that improve their performance on transfer learning tasks. One such modification is the use of *swish* activation functions, which have been shown to outperform ReLU activation functions on several computer vision tasks. Another modification is the use of a *global average pooling* layer at the end of the model, which averages the feature maps across the spatial dimensions to produce a fixed-size representation.

After pretraining, the BiT model can be fine-tuned for downstream tasks such as

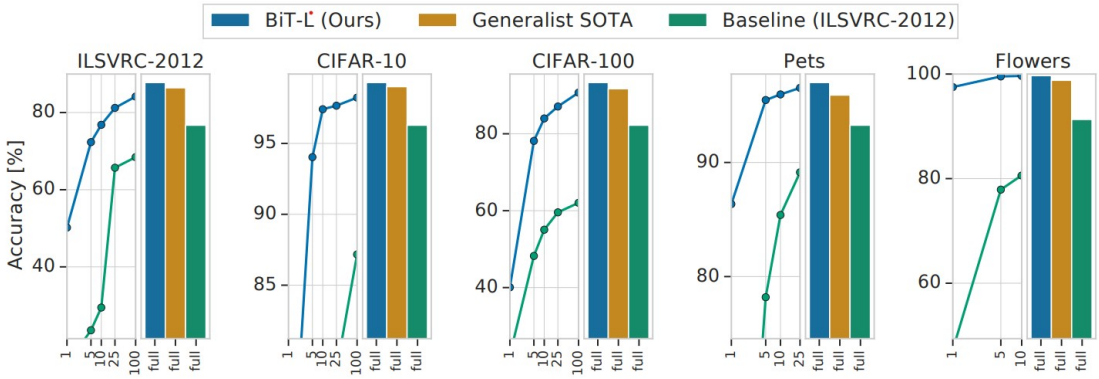


Figure 3.7: Comparison of transfer performance of BiT-L model (curve in blue) vs a ResNet-50 pre-trained on ImageNet (ILSVRC-2012) (curve in green) used as baseline. The x-axis shows the number of images used per class, ranging from 1 to the full dataset.

image classification, object detection, and semantic segmentation. By using transfer learning, BiT models are able to achieve SOTA performance on several benchmark computer vision datasets, such as CIFAR-10, CIFAR-100, and ImageNet, demonstrating their effectiveness for a wide range of tasks.

3.5 Inception-V4

InceptionV4 is an image classification model developed by Google researchers that builds on the Inception architecture with several enhancements to improve its performance. The model consists of Inception modules, which are building blocks that contain convolutional layers with various kernel sizes and strides to extract features at different scales. The model also incorporates techniques such as factorized 7x7 convolutions and shortcut connections to improve computation efficiency and gradient flow.

During training, the model is initialized with weights from a pretrained model such as InceptionV3 or ResNet and fine-tuned on a target dataset using regularization methods such as weight decay and dropout. InceptionV4 has achieved SOTA results on benchmark image classification datasets such as ImageNet and MS COCO, and

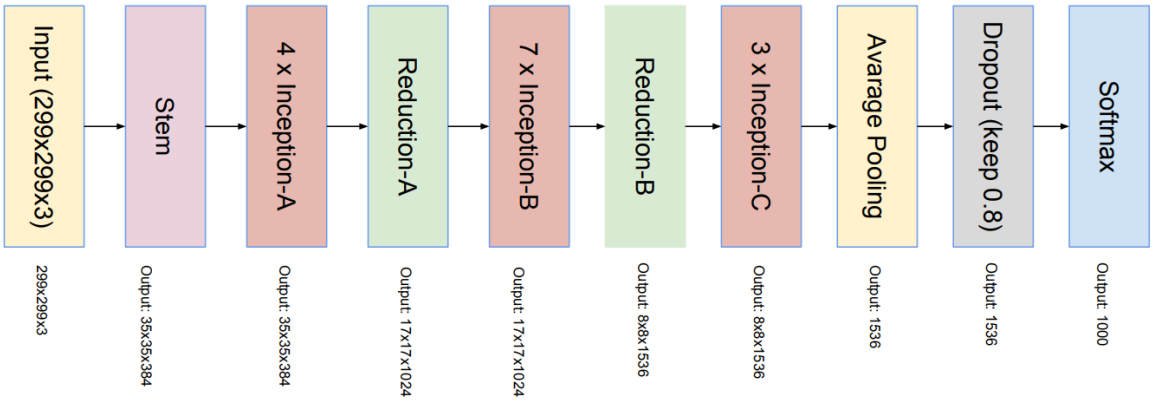


Figure 3.8: Inception-v4 network schema. The details of the individual modules are illustrated in Figures 3.9 and 3.10

its performance can be further improved by using other techniques such as attention mechanisms and adversarial training. We have included this model in our study in order to establish how the transformer-based models perform with respect to SOTA non-transformer based CNNs.

3.6 Human skin detection

In this module we discuss one of the classical human skin detection techniques that we explored as an additive branch of this thesis work. Human skin detection is a widely researched domain of computer vision and is a very essential step in skin disease detection (mole detection in this case). It helps us to localize the part of the image that belongs to the skin or not.

Over the years, many methods of skin detection have been developed that rely on the intensity profile of the sample images to determine whether details conform to skin or non-skin regions. Among these techniques, we have focused our attention to explicit threshold techniques since they are computationally not as expensive as statistical techniques and supervised learning-based methods and are also quite straightforward to implement.

The skin detection method implemented in this work is developed based on the

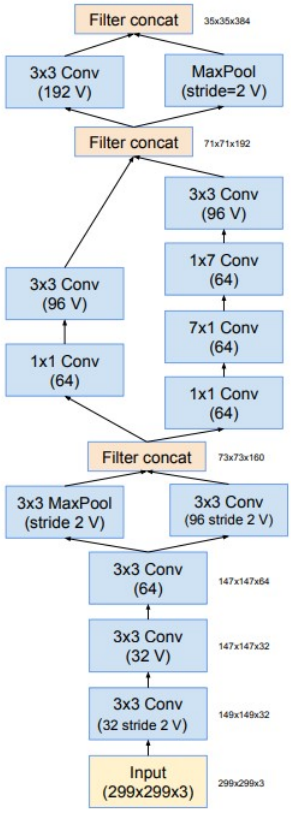


Figure 3.9: Schema of the stem of Inception-v4 model. This acts as the input layer to the network.

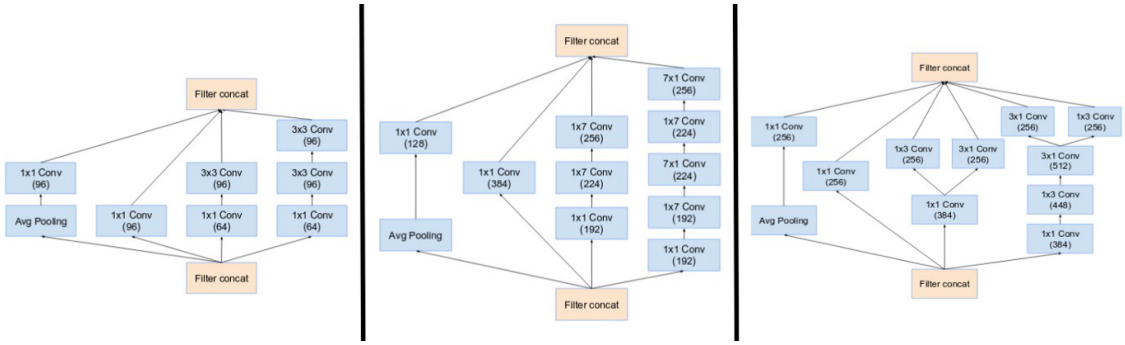


Figure 3.10: (*From left*) Inception A, B and C block in Inception-v4 model

algorithms proposed by [81] and [82]. It determines whether a given pixel in an image belongs to a skin or non-skin region depending on a set of rules applied to its RGB, YCrCb, HSV and CYMK color spaces.

Assuming that R, G, B are the intensity values of a pixel in the RGB color space, Y, Cr, Cb the intensity values in YCrCb space and H, S, V the intensity values in the HSV color space, the following three rules are defined:

- ***RGBA*** check:

  $rule\_1 = (R > 95)\&(G > 40)\&(B > 20)\&(R > G)\&(R > B)\&(abs(R - G) > 15)\&(A > 15)$

  $rule\_2 = (R > 220)\&(G > 210)\&(B > 170)\&(G > B)\&(R > B)\&(abs(R - G) <= 15)$

  $RULE\_RGB = rule\_1 \cup rule\_2$

- ***HSV*** check:

  $rule\_3 = H < 25$

  $rule\_4 = H > 230$

  $RULE\_HSV = rule\_3 \cup rule\_4$

- ***YCrCb*** check:

  $RULE\_YCRCB = (Cr > 135)\&(Cb > 85)\&(Y > 80)\&(Cr <= (1.5862 * Cb) + 20)\&(Cr >= (0.3448 * Cb) + 76.2069)\&(Cr >= (-4.5652 * Cb) + 234.5652)\&(Cr <= (-1.15 * Cb) + 301.75)\&(Cr <= (-2.2857 * Cb) + 432.85)$

In the above constraints, $abs()$ refers to the absolute function. In addition to the aforementioned rules, the authors define a new set of rules based on the CMYK color space as follows:

$$
\begin{aligned}
K &= min(255 - R, 255 - G, 255 - B) \\
C &= (255 - R - K)/(255 - K) \\
M &= (255 - G - K)/(255 - K) \\
Y &= (255 - B - K)/(255 - K)
\end{aligned}
\tag{3.5}
$$

In the above constraints, the $min()$ function refers to the minimum function that returns the minimum value from the given set of arguments. The C, M, Y, and K components are, in turn, used to define a new set of rules as follows.

CMYK check:

$rule\_5 = K < 205$

$rule\_6 = 0 <= C <= 0.05$

$rule\_7 = 0.0909 < Y < 0.945$

$rule\_8 = 0.1 <= Y/M < 4.67$

$RULE\_CMYK = rule\_5 \cap rule\_6 \cap rule\_7 \cap rule\_8$

Following the aforementioned rules, we adopt a hybrid color scheme to explore the combination of these rules:

$RC = RULE\_RGB \cap RULE\_CMYK$

$RH = RULE\_RGB \cap RULE\_HSV$

$RHC = RULE\_RGB \cap RULE\_HSV \cap RULE\_CMYK$

$RHC\_Vote = min\_2vote(RULE\_RGB, RULE\_HSV, RULE\_CMYK)$

In the above equation, min_2vote() refers to a function that returns $True$ if at least two rules from among its parameters vote for a given pixel as a skin pixel.

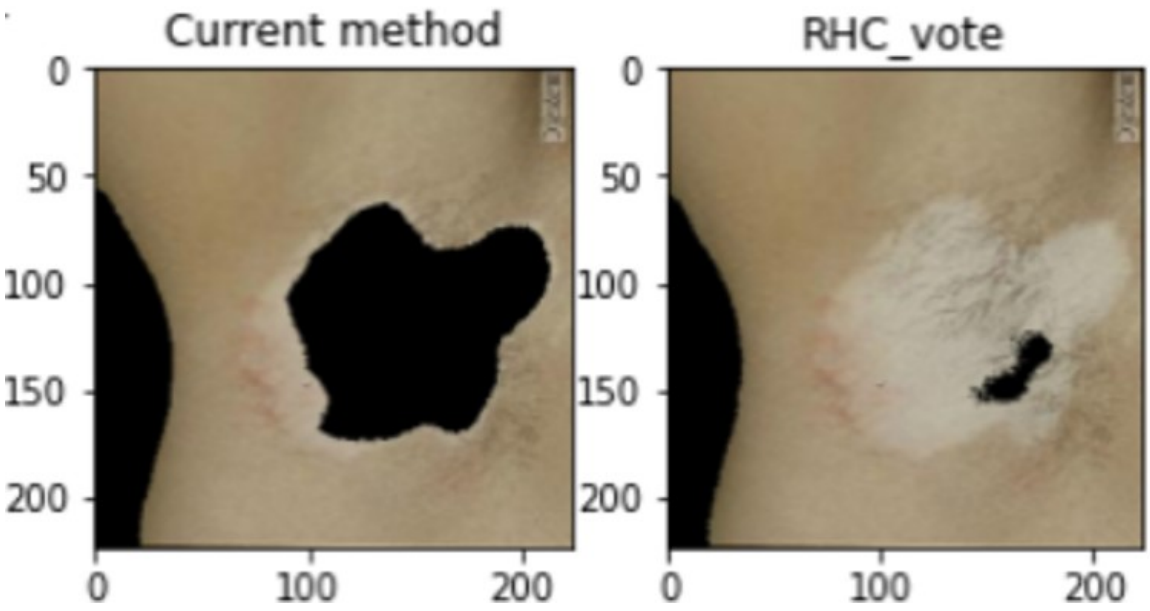


Figure 3.11: Comparison of skin detection performance on a sample image from Danderm dataset

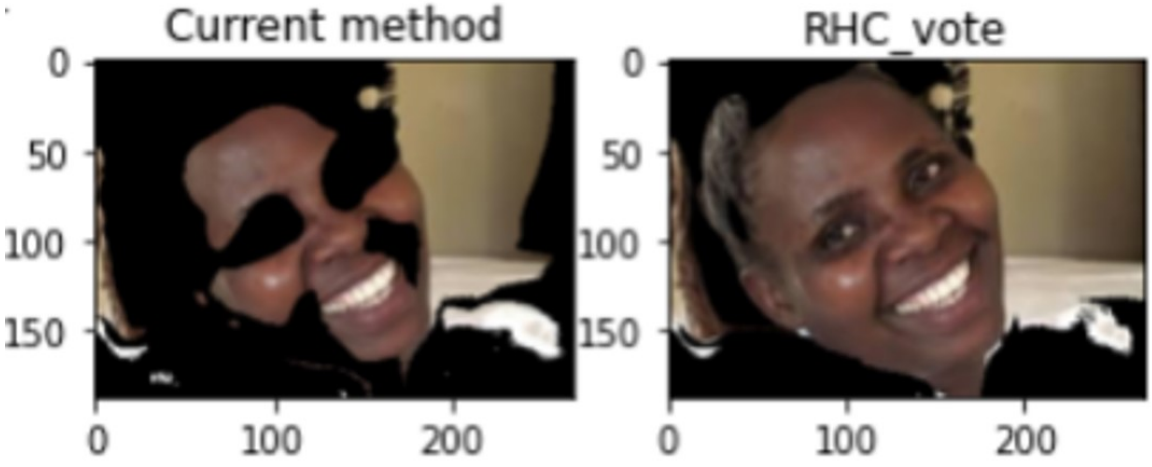


Figure 3.12: Comparison of skin detection performance on a sample Creative commons image

Chapter 4

Experimental results: Mole detection

In this chapter, we present the evaluated performance of the different models proposed earlier in our automatic CAD mole detection system. Multiple datasets, both public and private, have been used to encompass diverse types of skin conditions and robustify our system's classification confidence. We introduce the datasets followed by details of the different performance metrics achieved by performing different experiments.

4.1 Dataset

This project uses a combination of images from two sources: (i) a collection of real-world images curated by OROHealth Inc from different patients and (ii) a publicly available danderm dataset[1]. The former dataset is a private collection of images collected from dermatologists in collaboration with OROHealth Inc. These images were classified into more than 40 different classes of diseases by two Canadian board-certified dermatologists. Out of all these classes of conditions, images belonging to two conditions: nevus and melanoma, are considered moles, and the rest of the images are labeled as non-moles. Care has been taken to prevent data leakage by ensuring that images from the same patient are not included in more than one set (training,

[1]https://danderm-pdv.is.kkh.dk/

validation, or test set). On the other hand, the danderm dataset is a publicly available atlas built using photos taken during the private practice of clinical dermatology. It delivers a broad spectrum of clinical cases, sometimes belonging to the same disease, making it helpful in teaching dermatology. Out of the different clinical cases in this study, the images under nevus, melanoma, and tumors are considered moles, and the rest are non-moles. Figure 4.1 depicts two such cases taken from the public danderm dataset.

(a) Sample image with a nevus

(b) Sample image of melanoma

Figure 4.1: Example images from public danderm dataset showing skin conditions such as nevus and melanoma

4.2 Dataset class imbalance

Initially, we began our experiments with only the private dataset curated by ORO-Health Inc. However, we noticed a massive class imbalance problem in the dataset. Of a total of 24972 images in the dataset, only 451 samples belonged to the mole class (nevus and melanoma) resulting in an extreme class imbalance where the minority class (mole) constitutes just about 1.8% of the entire dataset. In order to tackle this problem, we included the images labeled as mole (nevus, melanoma, and tumours) from the public Danderm dataset along with the private image set. As a result, we managed to increase the number of mole samples from 451 to 1054. Therefore, we minimize the data imbalance problem by modifying the definition of moles or positive samples to include nevus, melanoma as well as tumors. However, this still results in a minority class of 4.12% of the total dataset. This problem was overcome by sub-sampling the training set into a smaller set containing at most 600 samples belonging to each class. Figure 4.2 shows the class distribution of the dataset, which represents about 96% of the dataset comprised of private non-mole images.

4.3 Experimental setup

The five models: BiT, NesT, Inception-v4 and ViT (base version and version trained on ImageNet21k) were subjected to 22 experiments, each using different combinations of hyperparameters such as optimizer, loss function, learning rate, and the total number of steps in the learning rate scheduler keeping all other hyperparameters the same. We relied on the deep learning library *timm*[2] for the implementation of all deep learning methods evaluated in this study. All experiments were executed on NVIDIA Tesla P4 GPUs on virtual machines supported by the Google Cloud Platform (GCP).

Once we had preprocessed and subsampled our image dataset, we applied a series of different augmentation techniques such as image resizing, randomly cropping a

[2]https://timm.fast.ai/

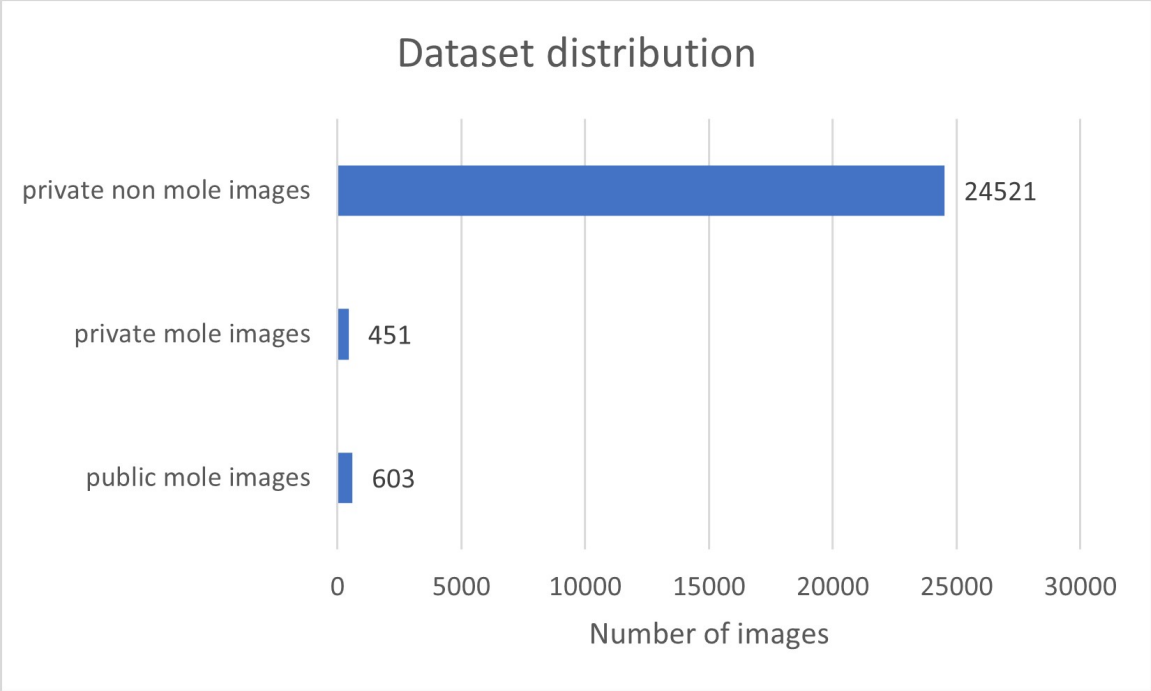


Figure 4.2: Bar chart of dataset distribution

part of an image (the lower bound of the random area of the crop is set to 0.15 times the total image area). This was followed by resizing, random flipping of images both horizontally and vertically, random rotation of an image by 90 degrees, etc. These transforms were followed by normalization and conversion to tensors. These tensors are used to generate data loaders for training, validation, and test sets, respectively. These data loaders are finally subjected to mixed data augmentation [83] in which randomly selected samples are linearly interpolated to generate synthetic training samples. The interpolation coefficient is selected from a beta distribution (fixed to 0.1 for all of our experiments). It must be noted that all the data augmentation procedures mentioned here are implemented online to the data batches to prevent the overhead of calling separate custom functions thereby making it more efficient.

During the construction of torch data loaders, we exploit the inclusive behavior of gradient accumulation [84]. Gradient accumulation essentially helps us when the batch size cannot be increased beyond a limit to account for GPU memory limitations. In this method, instead of updating the model variables after every batch, we

accumulate the gradients for a configured number of steps and then use the sum of those accumulated gradients to update the variables. The number of steps depends on the batch size (denoted by Gradient accumulation batch in Table 4.1) and the number of epochs.

We used three different types of loss function to train our models. The cross-entropy [85] loss was used since this is a binary classification problem, and cross-entropy loss is one of the most popularly used loss functions for such problems. The focal loss [86] had also been used, keeping in mind the considerable class imbalance in the dataset in the number of image samples between the two classes (mole and non-mole). The focal loss has consistently outperformed traditionally used loss functions, especially when there is a substantial dataset class imbalance. The focal loss (FL) is represented as follows:

$$FL(p) = \begin{cases} -\alpha(1-p)^{\gamma} log(p), & \text{if } y = 1 \\ -(1-\alpha)p^{\gamma} log(1-p), & \text{otherwise} \end{cases}$$

Here $y \in \{-1, 1\}$ is the ground truth class, and $p \in [0, 1]$ is the predicted probability of the model for the class being label $y = 1$. The class imbalance issue is addressed by introducing a weighing factor $\alpha \in [0, 1]$ for class $y = 1$ and $(1 - \alpha)$ for class $y = -1$. For our experiments, we have fixed its value at 0.25. A tunable focusing parameter $\gamma \geq 0$ is used to introduce a modulating factor $(1 - p)^{\gamma}$. Its value is set to 1 for all our experiments. Lastly, we have also used an asymmetric single label loss function that essentially decouples the focusing levels of positive and negative samples. Since, in our case, positive samples (mole images) are lesser in number than negative samples ('no mole' images), in order to accentuate the effect of the positive samples, we have set $\gamma_- \geq \gamma_+$ where γ_+ and γ_- are focusing parameters of positive and negative samples, respectively, in the loss function. For our experiments, we have set $\gamma_- = 2$ and $\gamma_+ = 0$, respectively. Additionally, to tackle problems with severe imbalance, probability shifting is employed via an extra tunable hyperparameter that performs hard thresholding of negative samples with very low probability. This value

has been set to 0.05 in all our tests. This modified version of focal loss is called an asymmetric single-label loss function.

In order to optimize our model weights during the training phase, we decided to employ two of the most common optimizers used in the domain of computer vision, which include the stochastic gradient descent (sgd)[87] optimizer and the adam [88] optimizer. Stochastic gradient descent is the most basic form of gradient descent optimizer, which subtracts the product of gradient and learning rate from the weights to optimize them. A momentum value of 0.9 is used to modify the weights, which helps to accelerate the training process. The Adam optimizer, on the other hand, combines the strengths of both the learning rate of RMSprop and the momentum property of stochastic gradient descent. Adam has proven to outperform most other optimizers and is widely regarded as the default optimization algorithm in most fields of machine learning. A weight decay factor is used for regularization and is set to 0.0005 for both sgd and Adam optimizers. Additionally, to make the training process efficient, we decided to include a patience clause of 5 epochs that terminates the trainer when the loss is appreciably low and does not diminish on further training.

In order to strategically control the learning rate while training, we have also employed two types of learning rate schedulers. The one-cycle scheduler anneals the learning rate in three phases, from an initial value to a maximum value. Then it decreases to some minimum value much lower than the initial starting value. The maximum learning rate value for the different experiments (for NesT) is varied, as shown in Table 4.1, and instead of using three phases, we decided to work with only two. The multiplication indices shown in Table 4.2 determine the different learning rate limits for the different models.

On the other hand, cosine annealing begins the learning process with an initial learning rate, where it gradually decreases following the shape of a cosine curve. On reaching a specific minimum value, the learning rate suddenly increases to where it started. This non-linear repetitive nature of the learning rate helps prevent undesir-

able minima in the weight space and takes care of the decay schedule.

Table 4.1 shows the synopsis of the different hyperparameters with which we compared to optimize and arrive at the best combination of values to obtain model weights of production grade. Every model was tested against different scale variations of these 22 sets of hyperparameter combinations.

4.4 Performance metric analysis

By a combined inference of the different box plots in Figures 4.3, 4.4, 4.5, 4.6, 4.7, and 4.8 we can infer that the NesT model gives the best result with a particular combination of hyperparameters (Batch size: 32, optimizer: Stochastic gradient descent, loss function: asymmetric single label loss, maximum learning rate: 0.001). Table 4.3 shows the confusion matrix of the best performing NesT model. Out of all the performance metrics presented, we are most interested in the recall value of test images with moles. This is because, as mentioned earlier, moles are usually not treatable online, and we need to minimize false negatives as much as possible. The high dataset class imbalance also causes the different models to be trained to be more likely to identify images without any moles with higher confidence than those with moles. Since mole detection is a problem where overlooked cases might become more costly than false alarms, we evaluate our models' performance mainly by their recall values for test images with moles.

Figures 4.3, 4.4 and 4.5 represent recall, precision, and overall accuracy box plots when one-cycle scheduler is used to control the learning rate, and figures 4.6, 4.7, and 4.8 represent the same but with cosine annealing. Figures 4.3 and 4.6 and figures 4.5 and 4.8 illustrate that the NesT model scores the best recall as well as overall accuracy values with both schedulers. Although the precision values of NesT are not the best, as illustrated by the figures 4.4 and 4.7, since we are more interested in recall scores, we accept the trade-off. In order to decide which scheduler to choose, we compare their performances by keeping all other hyperparameters the same. Figure 4.9 illustrates

Table 4.1: A table depicting the hyperparameters combination used for different experiments. Here SGD refers to stochastic gradient descent and ASL refers to asymmetric single-label loss.

Gradient accumulation batch	**Optimiser**	**Loss function**	**Learning rate**
16	SGD	ASL	0.01
32	SGD	ASL	0.01
64	SGD	ASL	0.01
32	SGD	ASL	0.001
32	SGD	ASL	0.1
32	adam	ASL	0.01
32	adam	ASL	0.001
32	adam	ASL	0.1
32	adam	cross entropy	0.01
32	adam	cross entropy	0.001
32	adam	cross entropy	0.1
32	SGD	cross entropy	0.01
32	SGD	cross entropy	0.001
32	SGD	cross entropy	0.1
16	SGD	focal	0.01
32	SGD	focal	0.01
64	SGD	focal	0.01
32	SGD	focal	0.001
32	SGD	focal	0.1
32	adam	focal	0.01
32	adam	focal	0.001
32	adam	focal	0.1

the parallel plot of multiple performance metrics from which we can infer that the one-cycle scheduler is more versatile than the cosine-annealing scheduler. Although

Table 4.2: Multiplication indices of the maximum learning rate for different models

NesT	**ViT**	**ViT (ImageNet21k pretrained)**	**BiT**	**InceptionV4**
x1	x0.01	x0.01	x0.3	x0.1

cosine annealing performs better for the recall value with mole images, the one-cycle scheduler performs better in terms of all other metrics such as overall accuracy (test accuracy), recall value of non-mole test images (test-no_mole-recall) and F1-scores for both mole and non-mole test images (test-mole-f1-score and test-no_mole-f1-score respectively). Hence, the model trained with the one-cycle scheduler has been chosen as the best model. Table 4.4 shows the recall values of the best-performing versions of each of the five models with their hyperparameter values.

Table 4.3: Confusion matrix of the best performing NesT model

		Predicted label		
		Mole	Non-mole	Total
True label	Mole	127	9	136
	Non-mole	7	277	284
	Total	145	275	420

4.5 API simulation and experimental tools

In order to simulate the end-to-end integrated application programming interface (API), we have used the Postman API[3] platform to evaluate how our system is expected to work. It provides a platform for developers to design, build, and document APIs. It also enables us to manage our API development workflow's collections, environments, requests, and other aspects.

In our project, most of the data was kept private as it is protected by a fully

[3]https://www.postman.com/api-platform/

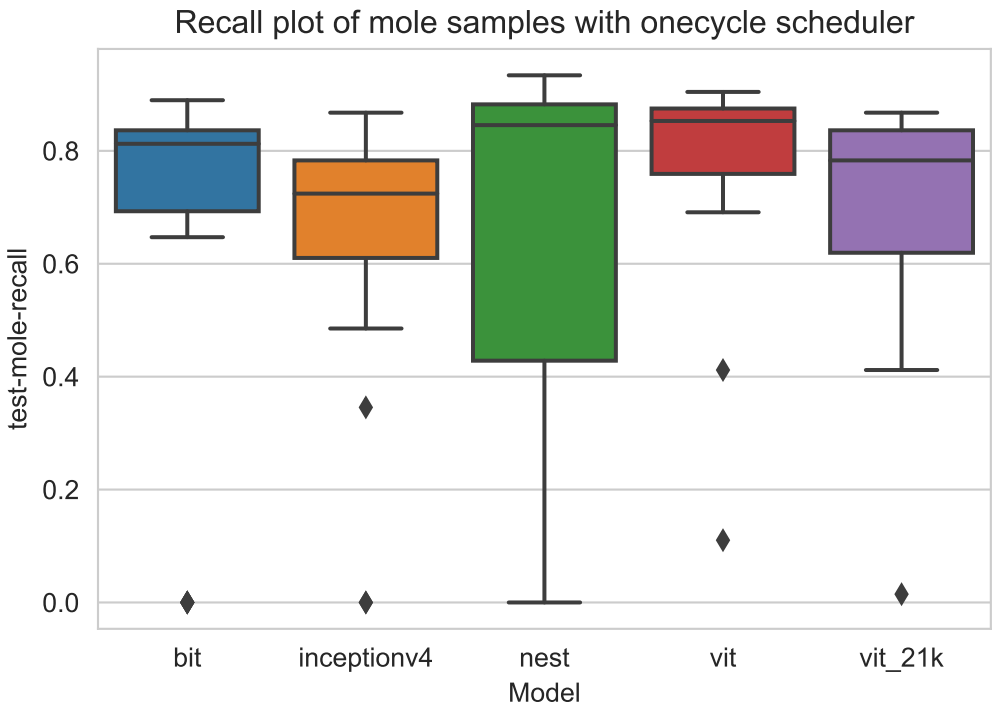


Figure 4.3: Recall box plots for test images with moles with onecycle scheduler

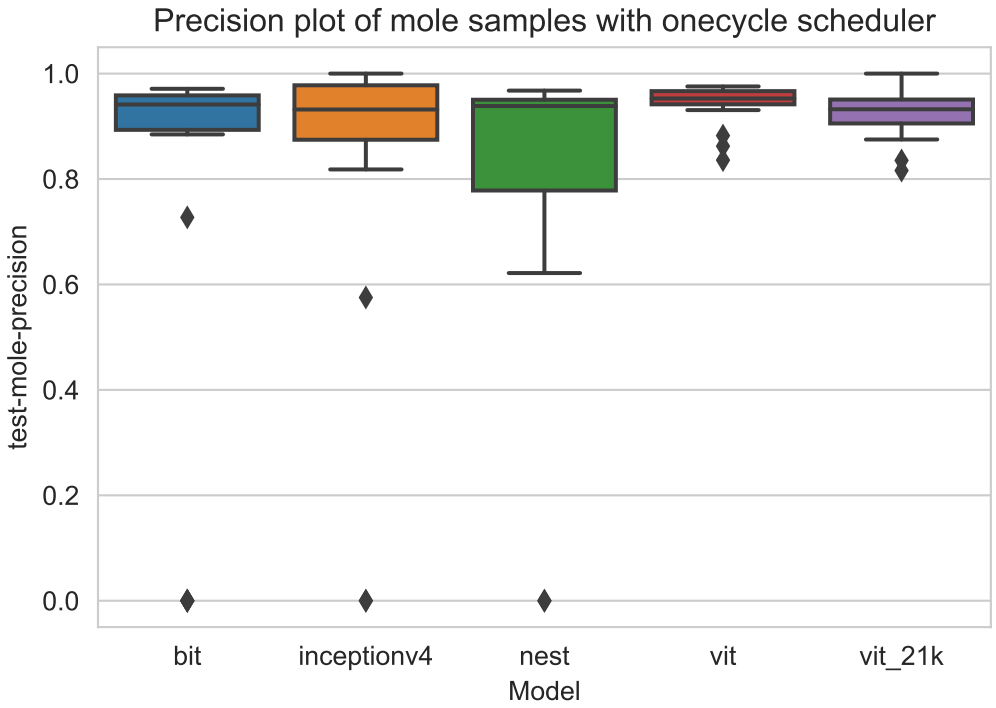


Figure 4.4: Precision box plots for test images with moles with onecycle scheduler

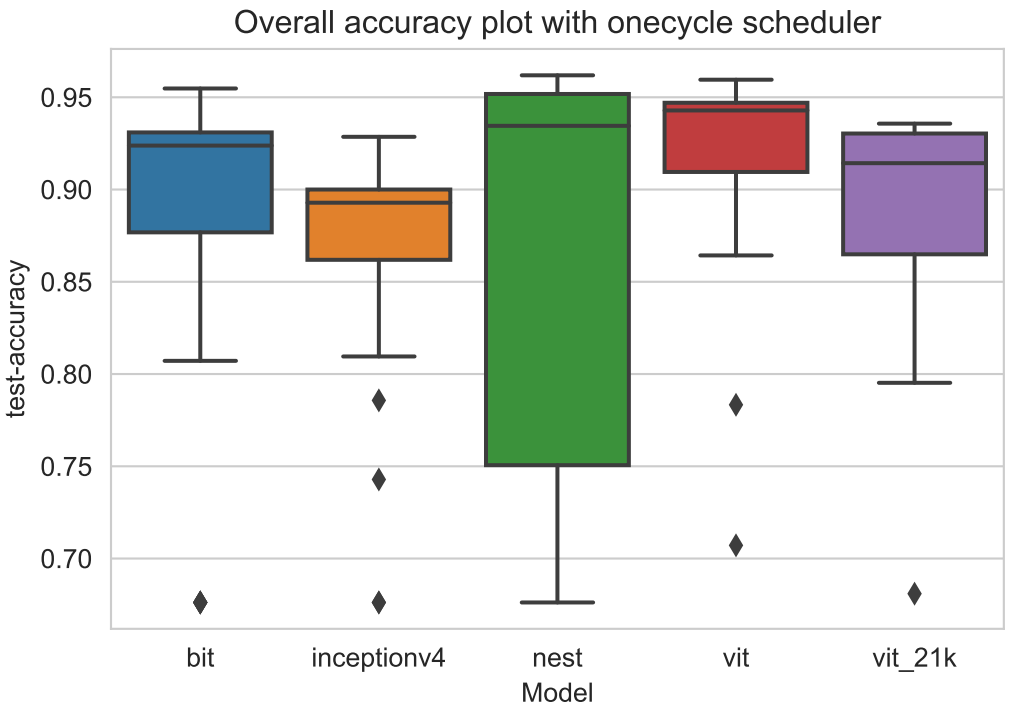


Figure 4.5: Overall accuracy box plots with onecycle scheduler

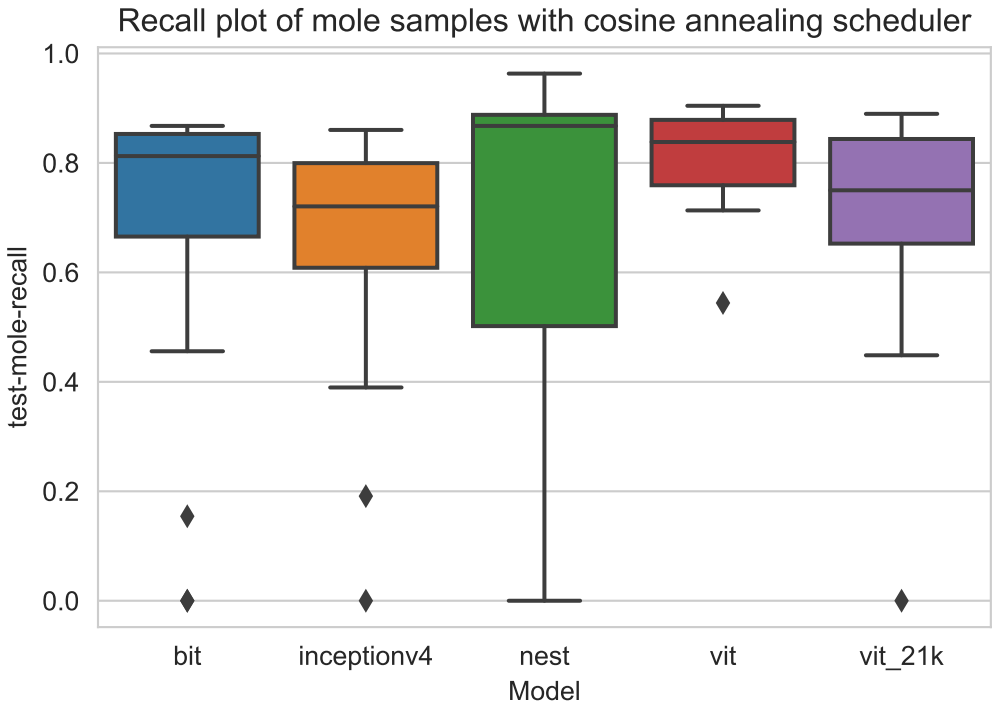


Figure 4.6: Recall box plots for test images with moles with cosine annealing scheduler

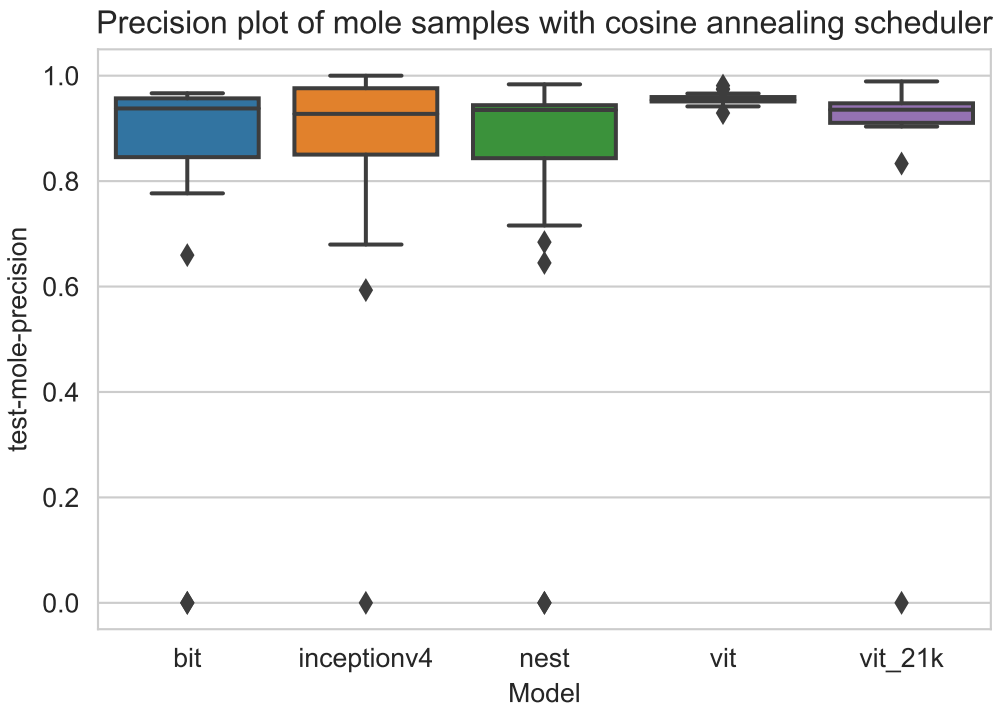


Figure 4.7: Precision box plots for test images with moles with cosine annealing scheduler

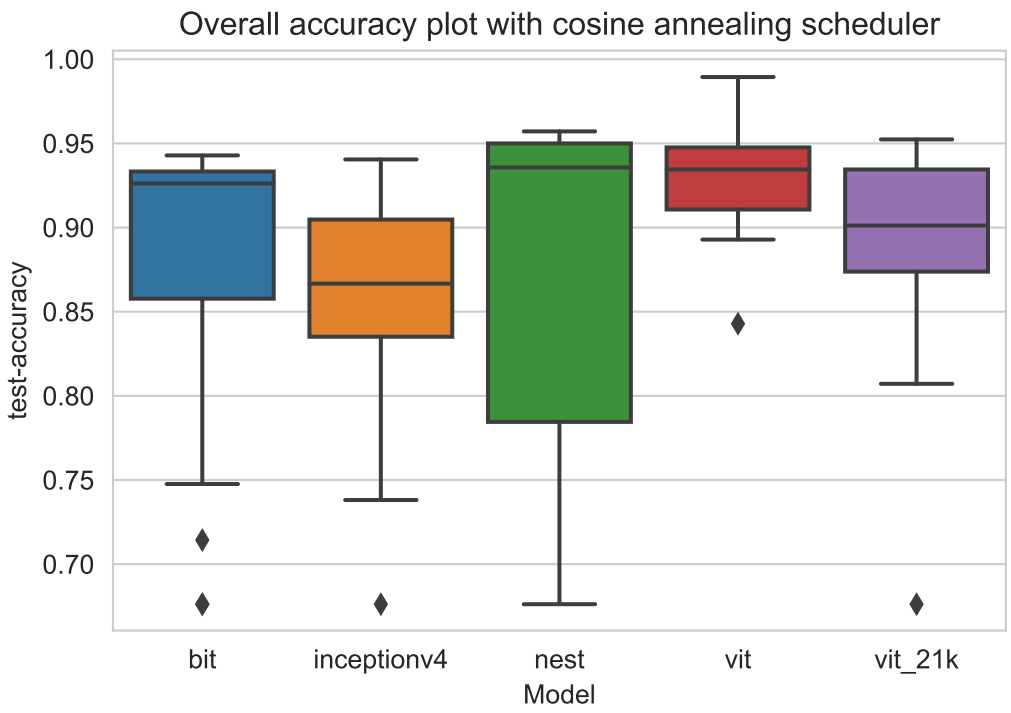


Figure 4.8: Overall accuracy box plots with cosine annealing scheduler

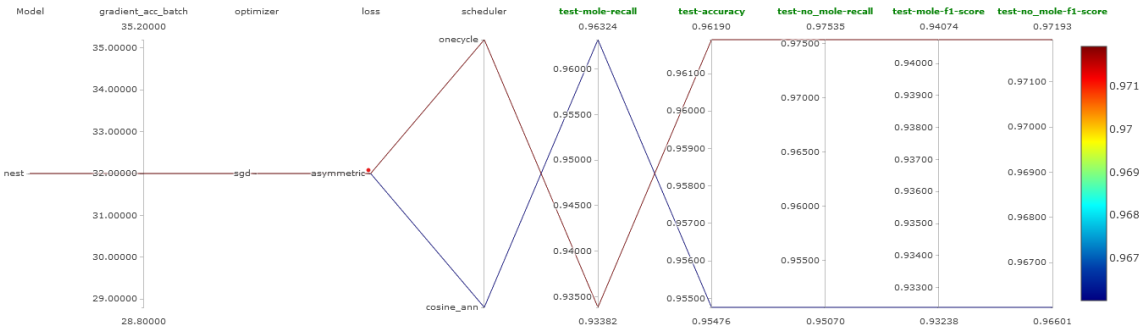


Figure 4.9: Parallel plot for metric comparison between onecycle scheduler vs cosine annealing scheduler keeping other parameters fixed. Since onecycle scheduler outperforms cosine annealing with respect to most metrics such as recall on images without moles (test-no_mole-recall), F1 score on images with and without moles (test-mole-f1-score, test-no_mole-f1-score), we accept the trade-off of a slightly inferior recall on images with moles (test-mole-recall).

Table 4.4: Best performing experiments of each model with onecycle scheduler with regards to recall in test images with moles

Model	**Gradient accumulation batch**	**Optimiser**	**Loss function**	**Learning rate**	**Mole recall**
BiT	32	sgd	asymmetric single label	0.03	0.8897
InceptionV4	32	sgd	asymmetric single label	0.01	0.8676
NesT	32	sgd	asymmetric single label	0.001	0.9338
ViT	32	sgd	cross entropy	0.0001	0.9044
ViT_21k	32	sgd	cross entropy	0.0001	0.8676

encrypted database that complies with HIPAA, GDPR, PIPEDA, and all other legal compliance requirements designed to protect confidential data. Every patient file or physician case is fully encrypted using the user's unique code such that only users have access to open their data using security keys. Such security measures are crucial for entities involved in handling public health information (PHI), such as healthcare providers, health plans, and their business associates. Hence we containerized our mole detection system by creating a docker[4] image of it. The container image was then sent to the GCP virtual machine with privileges to access the data in order to train the model, and the trained weights were stored externally. The experiments were automated and documented via MlFlow[5] which is an open-source toolkit to manage the machine learning life cycle and to follow experiments, manage models and reproduce results.

The docker image that ended up generating the best results was chosen to test in the postman API environment. Figures 4.10a, 4.10b, 4.11a, and 4.11b show the four

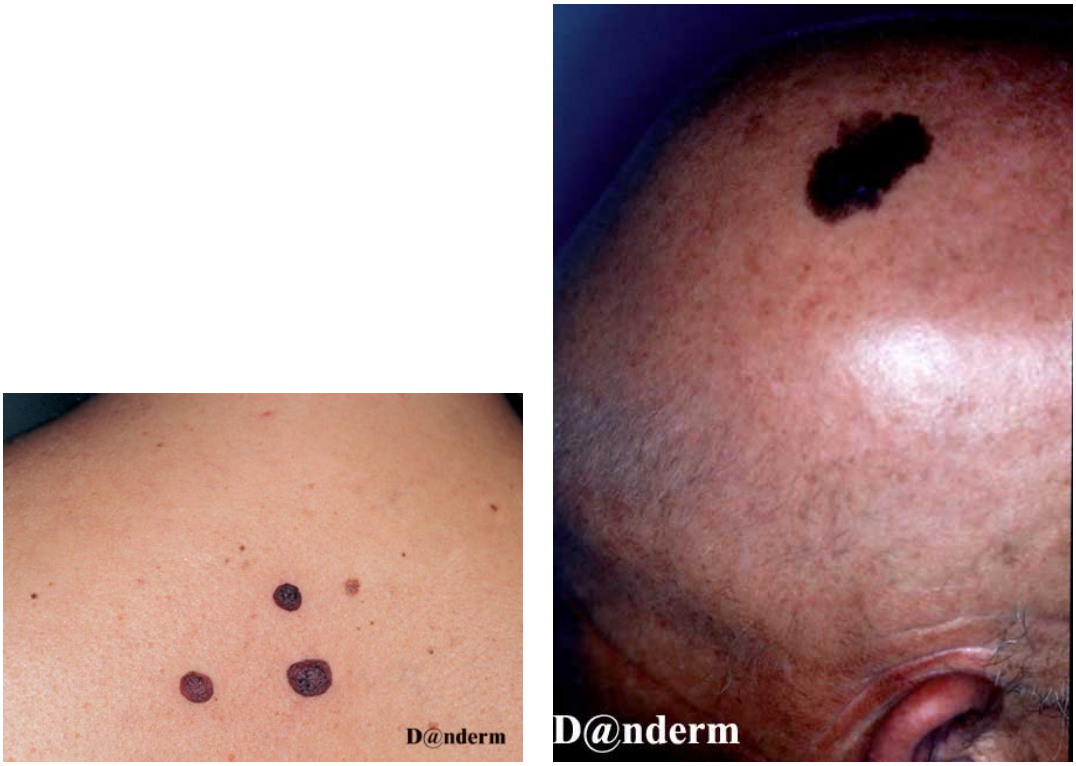
[4]https://www.docker.com/
[5]https://mlflow.org/toolkit

sample images taken from the danderm dataset. It returns Mole if the system detects the presence of a mole in the image and No_mole if it doesn't find any along with a confidence probability, i.e. probability that the image belongs to the predicted class. It is a measure of how confident the system is in its prediction.

Figure 4.10a shows a sample image of nevus taken from the danderm dataset. The system correctly classifies it as a mole with a confidence of about 94%. Such high confidence is reflective of the fact that the moles in the image are well-defined in a clear background with characteristic nevi features.

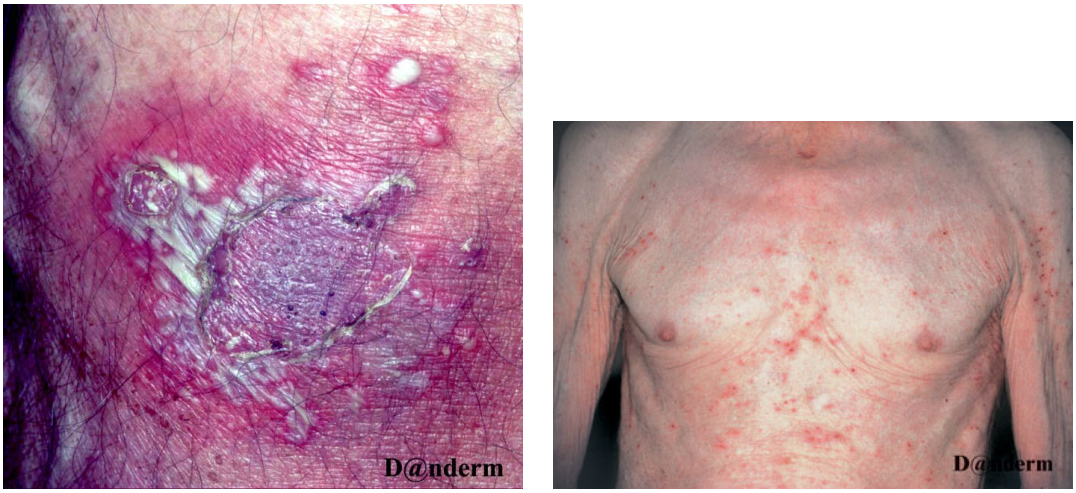
Figure 4.10b is that of a sample image of melanoma from the danderm dataset. Its prediction turns out to be a correct classification of Mole, although with a low confidence of 50%. This is due to the fact that the patch of mole in this case is located in an unclear background and the image has been taken under poor lighting conditions.

Figures 4.11a and 4.11b show sample images of pustular psoriasis, which is a type of genodermatosis and parasitic infestation, respectively. Their label prediction turns out to be correct as No_mole with high confidence values of 90% and 96% respectively. The high confidence in No_mole prediction is reflective of the fact that the dataset has a high imbalance between the two classes with considerably more images without moles than with moles.



(a) (b)

Figure 4.10: Sample (a) nevus image and (b) melanoma from the danderm dataset. These images were correctly classified by the system as images with moles with probability values of 94.9% and 50.5% respectively.



(a) (b)

Figure 4.11: Sample image of (a) pustular psoriasis, a type of genodermatosis and (b) parasite infestation from danderm dataset. These images were correctly classified by the system as images without any moles with probablility values of 89.8% and 95.9% respectively.

Chapter 5

Experimental results: Human skin detection

The primary objective of exploring a superior algorithm to detect regions of human skin from any given image is to make our study more versatile with respect to different skin tones. The datasets used in this thesis: the danderm dataset, as well as the private dataset from OROHealth Inc., predominantly consist of image samples taken from people with light skin tones. As a result, our model is trained on data limited to a certain group of people. This might result in faulty predictions if you find cases from patients with comparatively darker skin tones. Therefore, a robust human skin detection algorithm becomes vital to prevent such unforeseen data cases. In this chapter, we present and compare the performance of our skin detection algorithm against a classical SOTA algorithm. First, we discuss the different sources of our test data, followed by a qualitative analysis of masks obtained by implementing the aforementioned algorithms.

5.1 Data curation

In order to include skin tones of different shades, we searched for images of people of different ethnicities under the Creative Commons license. Images of natives from five different geographical units were selected: Indian subcontinent, the Middle East, Africa, Asia and Latin America. On the other hand, we have also tested the algo-

rithms against sample images taken from the Danderm dataset and the Pratheepan dataset [89].

5.2 Experimental setup and results

The different images curated are at first manually segmented using Segments.ai [90]. This segmenting tool is used to delineate regions of human skin as the foreground and the rest as the background, thereby generating a ground truth mask for reference.

Figures 5.1, 5.2 and 5.3 are 3 such cases illustrated in this section. Of these figures 5.1 is taken from the Danderm dataset, figure 5.2 is a creative commons image and figure 5.3 is taken from the Pratheepan dataset. The ground truths shown alongside the query images are the masks generated from the segments.ai API followed by a version showing the ground truth overlapped on the image. All the following images in each figure correspond to the images overlapped with masks generated from different thresholding rules. Each of these images is accompanied by their dice score with respect to the ground truth. The current method refers to the method currently being used by the data science team at OROHealth Inc [82]. Additionally, we have also included masks generated by intensity thresholding rules *RHC_vote, RC, RH* and *RHC* as described earlier in Chapter 3. Lastly, two more combinations of intensity thresholding rules, namely $RHC\_vote|YCrCb\&HSV$ and $RGB\&YCrCb$ were also tested.

Figure 5.1 shows a case of vitiligo from the danderm dataset. Quantitatively speaking, we see that the $RHC\_vote|YCrCb\&HSV$ method generates the mask with the highest dice score. Even visual inspection of the results reveals that the method in question generates a much better mask as it covers very little of the discolored area of the skin, unlike other methods.

Figure 5.2 is an image of a person of African descent with a Creative Commons license. Visual inspection reveals that the $RHC\_vote|YCrCb\&HSV$ method does a pretty good job of identifying the foreground (human skin) and leaving it all un-

masked. However, it does not fare that well when it comes to masking the background (non-skin) area. From a quantitative point of view, the dice scores of the masks generated by the $RHC\_vote|YCrCb\&HSV$ as well as the *RHC_vote* methods far exceed the ones generated by other methods.

Figures 5.3 are sample images taken from the Pratheepan dataset, which generate quite interesting results. All methods provide quite similar results masking the hair and eyes correctly but failing to mask the dress. Quantitatively, $RHC\_vote|YCrCb\&HSV$ and *RHC_vote* generate the masks with maximum Dice score, but do not correctly mask the eyes.

Tables 5.1, 5.2 and 5.3 show the dice scores of the segmentation masks generated by the different methods with respect to ground truth.

By carefully analyzing the results generated by the methods in hand, we can state that the current method generates masks in a more conservative manner compared to methods that include *RHC_vote*. The upside of this is the fact that the former (current) method covers more area of the background (non skin regions), whereas the *RHC_vote* based methods end up leaving too much of the background area unmasked. Consequently, the former method ends up masking too much of the skin in the process (more false negative) whereas the *RHC_vote* based methods perform quite well in avoiding areas of false negative and does not cover too much of the skin. As a result, we can say that it finally comes to the question of the trade-off of how much of the skin region we are interested in to be correctly masked at the cost of the background.

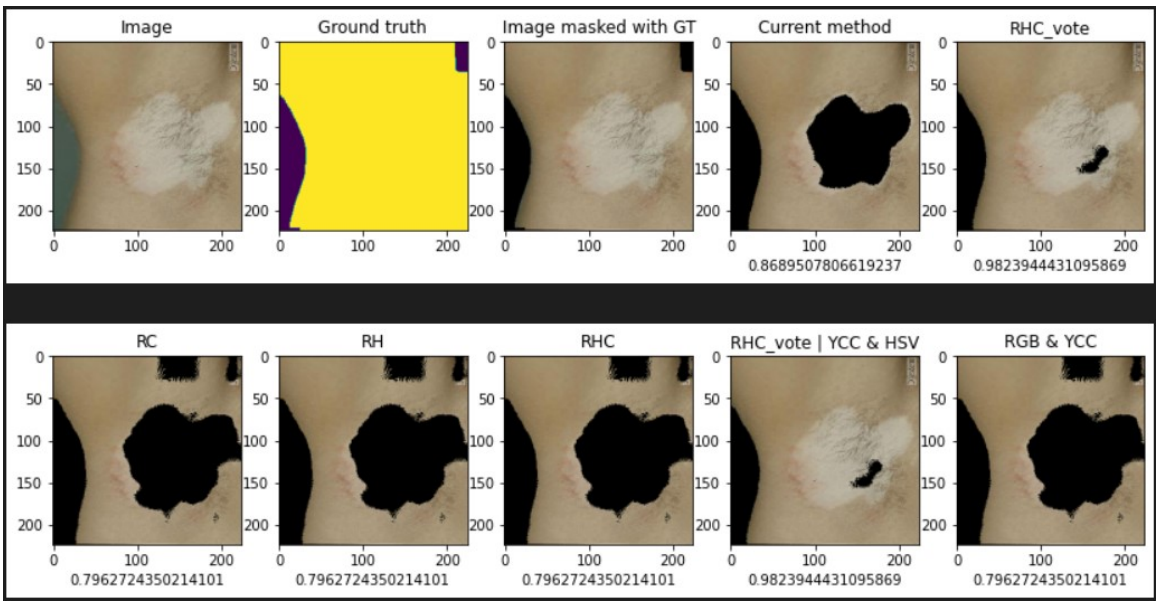


Figure 5.1: Different masking methods applied to a sample image from Danderm dataset. The illustrations represent (in clockwise manner) the query image, the ground truth generated manually from Segments.ai, ground truth overlapped on the query image, the masked images by former method, *RHC_vote* method, *RGB&YCC* method, *RHC_vote|YCC&HSV* method, *RHC* method, *RH* method and *RC* method respectively. Each of the masked images is accompanied by the dice score of their mask with respect to the ground truth written underneath.

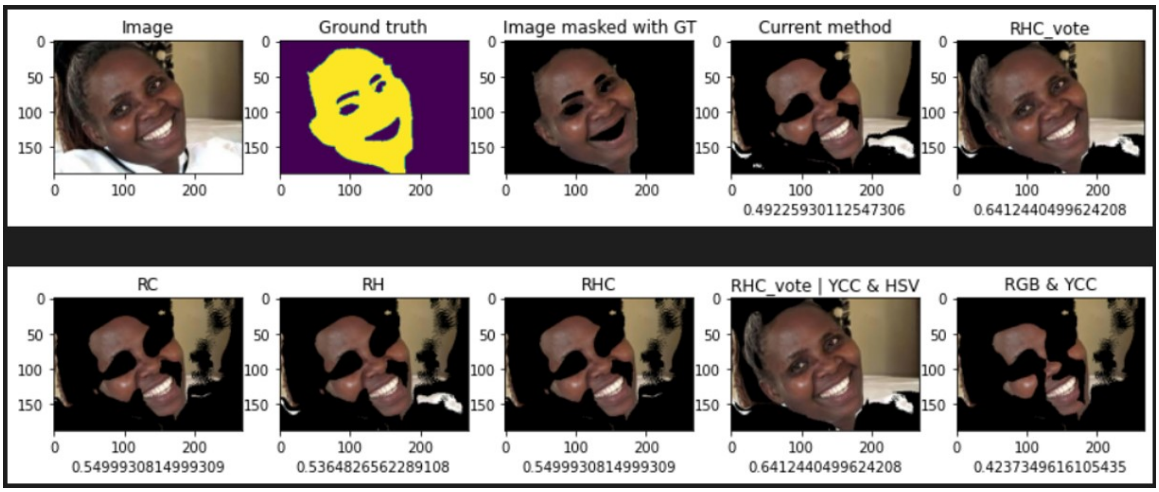


Figure 5.2: Different masking methods applied to a sample image from Creative Commons. Illustrations are presented in the same order as described in Figure 5.1

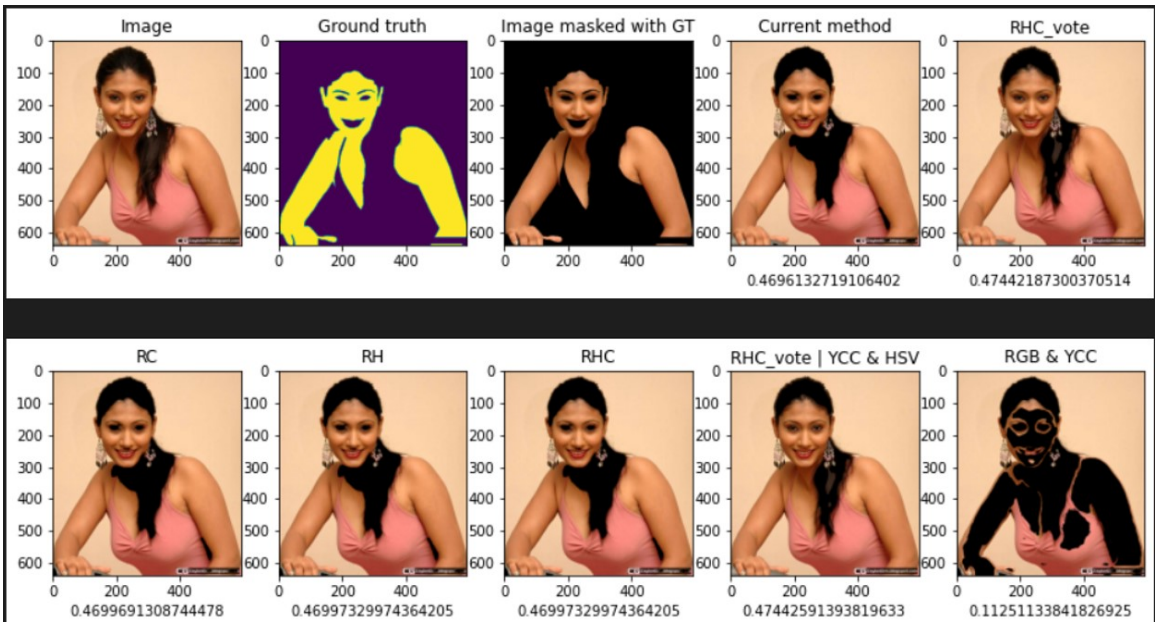


Figure 5.3: Different masking methods applied to a sample image from the Pratheepan dataset. Illustrations are presented in the same order as described in Figure 5.1

Table 5.1: Dice scores of the skin segmentation masks generated by the different methods on the Danderm image from figure 5.1

Segmentation method	**Dice score**
Current method	0.87
RHC_vote	0.98
RC	0.80
RH	0.80
RHC	0.80
RHC_vote|YCC&HSV	0.98
RGB&YCC	0.80

Table 5.2: Dice scores of the skin segmentation masks generated by the different methods on the Creative commons image from figure 5.2

Segmentation method	**Dice score**
Current method	0.87
RHC_vote	0.98
RC	0.80
RH	0.80
RHC	0.80
RHC_vote|YCC&HSV	0.98
RGB&YCC	0.80

Table 5.3: Dice scores of the skin segmentation masks generated by the different methods on the image from Pratheepan dataset shown in figure 5.3

Segmentation method	**Dice score**
Current method	0.87
RHC_vote	0.98
RC	0.80
RH	0.80
RHC	0.80
RHC_vote|YCC&HSV	0.98
RGB&YCC	0.80

Chapter 6

Conclusion & Future Work

6.1 Conclusion

Our aim for this project was to present a solution to the problem of detecting moles from images uploaded by patients on a telemedicine platform. The model was trained on a large dataset of real-world images and achieved accurate results in detecting moles in new images. The results of this research demonstrate the great potential of deep learning in classification in the early stages of dermatological conditions and the need for further research in this area.

The NesT model achieved a recall value of 93.4% on test images with moles and a macro average recall value of 95.46% on the entire test dataset. This presents one of the many possible avenues in which deep learning can be confidently applied to revolutionize the field of telemedicine, especially in dermatology, making treatment more accessible and efficient.

However, it is essential to note that this research is limited because most images have a lighter skin complexion and the number of image samples with moles that have been used to train and test the model is small. Further research should focus on increasing the diversity of the dataset and incorporating more diverse skin types and moles to improve the model's generalizability.

We have explored the aforementioned diversity issue from a different angle by testing two different classical methods of human skin detection. However, it is found that

the higher detection of skin tones of darker complexions comes at the cost of inferior performance in terms of segmenting the foreground or skin from the background. Hence future research for this problem should either focus on employing models with better abilities to detect human skin or on implementing models which work better for the specific task such as mole detection without segmenting out skin regions.

In general, the results of this research provide a promising first step toward using deep learning to successfully assess patients to determine their urgency of treatment. This demonstrates the potential of this technology to improve the lives of many people, as it enables the patient to understand the nature of the treatment required and the appropriate place to go for the best treatment.

6.2 Future Work

Considering the future scope of research in applications of artificial intelligence in telemedicine, the future direction of research is significantly optimistic. Our work can be extended to various problems of classifying and analyzing specific conditions such as alopecia, eczema, and genitopathology. Moving beyond the sole use of images, other aspects of machine learning can be used to develop tools to maintain patient history and screen patients, thereby recommending follow-up appointments or treatment procedures. Since the application of AI in teledermatology is still in its very early stage of development in business, problems are abundant and there is great potential for innovation.

Bibliography

[1] J. A. Segre, "Epidermal barrier formation and recovery in skin disorders," *The Journal of Clinical Investigation*, vol. 116, no. 5, pp. 1150–1158, May 2006. DOI: 10.1172/JCI28521. [Online]. Available: https://www.jci.org/articles/view/28521.

[2] H. L. Weiner, L. P. Levitt, and A. Rae-Grant, *Neurology*. Williams & Wilkins, 1999.

[3] H. W. Lim *et al.*, "The burden of skin disease in the united states," *Journal of the American Academy of Dermatology*, vol. 76, no. 5, 958–972.e2, 2017, ISSN: 0190-9622. DOI: https://doi.org/10.1016/j.jaad.2016.12.043.

[4] A. M. Goldstein and M. A. Tucker, "Dysplastic Nevi and Melanoma," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 22, no. 4, pp. 528–532, Apr. 2013, ISSN: 1055-9965. DOI: 10.1158/1055-9965.EPI-12-1346.

[5] H. Tsao, C. Bevona, W. Goggins, and T. Quinn, "The Transformation Rate of Moles (Melanocytic Nevi) Into Cutaneous Melanoma: A Population-Based Estimate," *Archives of Dermatology*, vol. 139, no. 3, pp. 282–288, Mar. 2003, ISSN: 0003-987X. DOI: 10.1001/archderm.139.3.282.

[6] C. Liddy *et al.*, "How long are canadians waiting to access specialty care? retrospective study from a primary care perspective," *Canadian family physician Medecin de famille canadien*, vol. 66, pp. 434–444, Jun. 2020.

[7] D. Das, E. Ergin, B. Morel, M. Noga, D. Emery, and K. Punithakumar, "AI-assisted mole detection for online dermatology triage in telemedicine settings," *Informatics in Medicine Unlocked*, vol. 41, p. 101 311, 2023, ISSN: 2352-9148. DOI: https://doi.org/10.1016/j.imu.2023.101311.

[8] A. Dosovitskiy *et al.*, *An image is worth 16x16 words: Transformers for image recognition at scale*, 2021. arXiv: 2010.11929 [cs.CV].

[9] A. Kolesnikov *et al.*, *Big transfer (bit): General visual representation learning*, 2020. arXiv: 1912.11370 [cs.CV].

[10] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, *Inception-v4, inception-resnet and the impact of residual connections on learning*, 2016. arXiv: 1602.07261 [cs.CV].

[11] Z. Zhang, H. Zhang, L. Zhao, T. Chen, S. O. Arik, and T. Pfister, *Nested hierarchical transformer: Towards accurate, data-efficient and interpretable visual understanding*, 2021. arXiv: 2105.12723 [cs.CV].

[12] J. Schofield, D. Fleming, D. Grindlay, and H. Williams, "Skin conditions are the commonest new reason people present to general practitioners in England and Wales," *British Journal of Dermatology*, vol. 165, no. 5, pp. 1044–1050, Nov. 2011, ISSN: 0007-0963. DOI: 10.1111/j.1365-2133.2011.10464.x. eprint: https://academic.oup.com/bjd/article-pdf/165/5/1044/47505999/bjd1044.pdf. [Online]. Available: https://doi.org/10.1111/j.1365-2133.2011.10464.x.

[13] M. K. Basra and M. Shahrukh, "Burden of skin diseases," *Expert Review of Pharmacoeconomics & Outcomes Research*, vol. 9, no. 3, pp. 271–283, 2009, PMID: 19527100. DOI: 10.1586/erp.09.23. eprint: https://doi.org/10.1586/erp.09.23. [Online]. Available: https://doi.org/10.1586/erp.09.23.

[14] R. J. Hay *et al.*, "The global burden of skin disease in 2010: An analysis of the prevalence and impact of skin conditions," *Journal of Investigative Dermatology*, vol. 134, no. 6, pp. 1527–1534, 2014, ISSN: 0022-202X. DOI: https://doi.org/10.1038/jid.2013.446. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022202X15368275.

[15] J. Ferlay *et al.*, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012," *International journal of cancer*, vol. 136, no. 5, E359–E386, 2015.

[16] H. Sung *et al.*, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[17] M. R. Donaldson and B. M. Coldiron, "No end in sight: The skin cancer epidemic continues," in *Seminars in cutaneous medicine and surgery*, WB Saunders, vol. 30, 2011, pp. 3–5.

[18] C. Smittenaar, K. Petersen, K Stewart, and N Moitt, "Cancer incidence and mortality projections in the uk until 2035," *British journal of cancer*, vol. 115, no. 9, pp. 1147–1155, 2016.

[19] D. E. Godar, "Worldwide increasing incidences of cutaneous malignant melanoma," *Journal of skin cancer*, vol. 2011, 2011.

[20] R. J. Friedman, D. S. Rigel, and A. W. Kopf, "Early detection of malignant melanoma: The role of physician examination and self-examination of the skin.," *CA: a cancer journal for clinicians*, vol. 35, no. 3, pp. 130–151, 1985.

[21] N. R. Abbasi *et al.*, "Early diagnosis of cutaneous melanoma: Revisiting the abcd criteria," *Jama*, vol. 292, no. 22, pp. 2771–2776, 2004.

[22] L Thomas, P Tranchand, F Berard, T Secchi, C Colin, and G Moulin, "Semiological value of abcde criteria in the diagnosis of cutaneous pigmented tumors," *Dermatology*, vol. 197, no. 1, pp. 11–17, 1998.

[23] R. L. Barnhill, G. C. Roush, M. S. Ernstoff, and J. M. Kirkwood, "Interclinician agreement on the recognition of selected gross morphologic features of pigmented lesions: Studies of melanocytic nevi v.," *Journal of the American Academy of Dermatology*, vol. 26, no. 2, pp. 185–190, 1992.

[24] M. D. Corbo and J. Wismer, "Agreement between dermatologists and primary care practitioners in the diagnosis of malignant melanoma: Review of the literature," *Journal of cutaneous medicine and surgery*, vol. 16, no. 5, pp. 306–310, 2012.

[25] S. C. Chen, D. M. Bravata, E. Weil, and I. Olkin, "A comparison of dermatologists' and primary care physicians' accuracy in diagnosing melanoma: A systematic review," *Archives of dermatology*, vol. 137, no. 12, pp. 1627–1634, 2001.

[26] *Abcde method for early detection fo melanoma*, www.canadianskincancerfoundation.com/early-detection/the-abcdes-of-early-detection, Accessed: 2023-06-25.

[27] S. Higgins and A. Wysong, "Cosmetic surgery and body dysmorphic disorder – an update," *International Journal of Women's Dermatology*, vol. 4, no. 1, pp. 43–48, 2018, Self-Esteem: The Emotional Burden and Psychological Devastation of Skin Disease, ISSN: 2352-6475. DOI: https://doi.org/10.1016/j.ijwd.2017.09.007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352647517300801.

[28] R. V. E. Ribeiro, "Prevalence of body dysmorphic disorder in plastic surgery and dermatology patients: A systematic review with meta-analysis," *Aesthetic plastic surgery*, vol. 41, pp. 964–970, 2017.

[29] K. França *et al.*, "Body dysmorphic disorder: History and curiosities," *Wiener Medizinische Wochenschrift*, vol. 167, pp. 5–7, 2017.

[30] M. Alavi, Y. Kalafi, G. R. Dehbozorgi, and A. Javadpour, "Body dysmorphic disorder and other psychiatric morbidity in aesthetic rhinoplasty candidates," *Journal of plastic, reconstructive & aesthetic surgery*, vol. 64, no. 6, pp. 738–741, 2011.

[31] A. Valikhani and M. A. Goodarzi, "Contingencies of self-worth and psychological distress in iranian patients seeking cosmetic surgery: Integrative self-knowledge as mediator," *Aesthetic plastic surgery*, vol. 41, pp. 955–963, 2017.

[32] H. H. Roenigk Jr, "Treatment of the aging face.," *Dermatologic clinics*, vol. 13, no. 2, pp. 245–261, 1995.

[33] M. K. Goin and T. D. Rees, "A prospective study of patients' psychological reactions to rhinoplasty.," *Annals of plastic surgery*, vol. 27, no. 3, pp. 210–215, 1991.

[34] L. Linn and R. Blacher, "Cosmetic surgery, with particular reference to rhinoplasty," *The Psychological Experience of Surgery. Ed. Richard S. Blacher. New York: John Wiley and Sons*, pp. 194–206, 1987.

[35] T. L. Diepgen, "Occupational skin diseases," *JDDG: Journal der Deutschen Dermatologischen Gesellschaft*, vol. 10, no. 5, pp. 297–315, 2012. DOI: https://doi.org/10.1111/j.1610-0387.2012.07890.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1610-0387.2012.07890.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1610-0387.2012.07890.x.

[36] H. Dickel, T Bruckner, C Bernhard-Klimt, T Koch, R Scheidt, and T. L. Diepgen, "Surveillance scheme for occupational skin disease in the saarland, frg: First report from bkh-s," *Contact Dermatitis*, vol. 46, no. 4, pp. 197–206, 2002.

[37] S. K. Singh and A. S. Jalal, "A robust approach for automatic skin cancer disease classification," in *2016 1st India International Conference on Information Processing (IICIP)*, 2016, pp. 1–4. DOI: 10.1109/IICIP.2016.7975301.

[38] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using GLCM approach," 2013.

[39] V. J. Ramya, J. Navarajan, R. Prathipa, and L. A. Kumar, "Detection of melanoma skin cancer using digital camera images," 2015.

[40] N. K. El Abbadi, N. S. Dahir, M. A. AL-Dhalimi, and H. Restom, "Psoriasis detection using skin color and texture features," *Journal of Computer Science*, vol. 6, no. 6, pp. 648–652, 2010. DOI: 10.3844/jcssp.2010.648.652.

[41] S. Achakanalli and G Sadashivappa, "Skin cancer detection and diagnosis using image processing and implementation using neural networks and abcd parameters," *International Journal of Electronics, Communication & Instrumentation Engineering Research and Development (IJECIERD)*, vol. 4, pp. 85–96, Jun. 2014.

[42] A. Ajith, V. Goel, P. Vazirani, and M. M. Roja, "Digital dermatology: Skin disease detection model using image processing," *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 168–173, 2017.

[43] W.-H. Chen, C. Smith, and S. Fralick, "A fast computational algorithm for the discrete cosine transform," *IEEE Transactions on Communications*, vol. 25, no. 9, pp. 1004–1009, 1977. DOI: 10.1109/TCOM.1977.1093941.

[44] M. Sifuzzaman, "Application of wavelet transform and its advantages compared to fourier transform," 2009.

[45] S. Kulkarni and A. R. Shelke, *Multiresolution analysis for medical image segmentation using wavelet transform*, 2014.

[46] J.-M. Renders and T. Simonart, "Role of Artificial Neural Networks in Dermatology," *Dermatology*, vol. 219, no. 2, pp. 102–104, Jun. 2009, ISSN: 1018-8665. DOI: 10.1159/000225933. eprint: https://karger.com/drm/article-pdf/219/2/102/2655641/000225933.pdf. [Online]. Available: https://doi.org/10.1159/000225933.

[47] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.

[48] V. Dick, C. Sinz, M. Mittlböck, H. Kittler, and P. Tschandl, "Accuracy of Computer-Aided Diagnosis of Melanoma: A Meta-analysis," *JAMA Dermatology*, vol. 155, no. 11, pp. 1291–1299, Nov. 2019, ISSN: 2168-6068. DOI: 10.1001/jamadermatol.2019.1375. eprint: https://jamanetwork.com/journals/jamadermatology/articlepdf/2736374/jamadermatology_dick_2019_oi_190033.pdf. [Online]. Available: https://doi.org/10.1001/jamadermatol.2019.1375.

[49] Y. Fujisawa *et al.*, "Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis," *British Journal of Dermatology*, vol. 180, no. 2, pp. 373–381, Feb. 2019, ISSN: 0007-0963. DOI: 10.1111/bjd.16924. eprint: https://academic.oup.com/bjd/article-pdf/180/2/373/47586788/bjd0373.pdf. [Online]. Available: https://doi.org/10.1111/bjd.16924.

[50] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park, and S. E. Chang, "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *Journal of Investigative Dermatology*, vol. 138, no. 7, pp. 1529–1538, 2018, ISSN: 0022-202X. DOI: https://doi.org/10.1016/j.jid.2018.01.028. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022202X18301118.

[51] S. Han, I. Park, S. Chang, and J. Na, "991 deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for general skin disorders," *Journal of Investigative Dermatology*, vol. 139, no. 5, Supplement, S171, 2019, Society for Investigative Dermatology (SID) 2019 Meeting Abstract Supplement, ISSN: 0022-202X. DOI: https://doi.org/10.1016/j.jid.2019.03.1067. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022202X19312588.

[52] P. Tschandl *et al.*, "Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks," *JAMA Dermatology*, vol. 155, no. 1, pp. 58–65, Jan. 2019, ISSN: 2168-6068. DOI: 10.1001/jamadermatol.2018.4378. eprint: https://jamanetwork.com/journals/jamadermatology/articlepdf/2716294/jamadermatology_tschandl_2018_oi_180064.pdf. [Online]. Available: https://doi.org/10.1001/jamadermatol.2018.4378.

[53] R. C. Maron *et al.*, "Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks," *European Journal of Cancer*, vol. 119, pp. 57–65, 2019, ISSN: 0959-8049. DOI: https://doi.org/10.1016/j.ejca.2019.06.013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959804919303818.

[54] T. J. Brinker *et al.*, "A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task," *European Journal of Cancer*, vol. 111, pp. 148–154, 2019, ISSN: 0959-8049. DOI: https://doi.org/10.1016/j.ejca.2019.02.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959804919301443.

[55] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," Jun. 2018, pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474.

[56] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–80, Dec. 1997. DOI: 10.1162/neco.1997.9.8.1735.

[57] P. R. Kshirsagar, H. Manoharan, S. Shitharth, A. M. Alshareef, N. Albishry, and P. K. Balachandran, "Deep learning approaches for prognosis of automated skin disease," *Life*, vol. 12, 2022.

[58] M. Chen, P. Zhou, D. Wu, L. Hu, M. M. Hassan, and A. Alamri, "AI-skin: Skin disease recognition based on self-learning and wide data collection through a closed-loop framework," *Information Fusion*, vol. 54, pp. 1–9, 2020, ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2019.06.005.

[59] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. DOI: 10.1109/5.726791.

[60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25, Curran Associates, Inc., 2012.

[61] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, 2014. DOI: 10.48550/ARXIV.1409.1556.

[62] X. Jiang, J. Jiang, B. Wang, J. Yu, and J. Wang, "SEACU-Net: Attentive ConvLSTM U-Net with squeeze-and-excitation layer for skin lesion segmentation," *Computer Methods and Programs in Biomedicine*, vol. 225, p. 107 076, 2022, ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2022.107076.

[63] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional networks for biomedical image segmentation*, 2015. DOI: 10.48550/ARXIV.1505.04597.

[64] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, *Convolutional LSTM network: A machine learning approach for precipitation nowcasting*, 2015. DOI: 10.48550/ARXIV.1506.04214.

[65] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, *Image-to-image translation with conditional adversarial networks*, 2018. arXiv: 1611.07004 [cs.CV].

[66] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of Graphics Tools*, vol. 9, Jan. 2004. DOI: 10.1080/10867651.2004.10487596.

[67] Y. Liu *et al.*, "A deep learning system for differential diagnosis of skin diseases," *Nature medicine*, vol. 26, no. 6, pp. 900–908, 2020.

[68] B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska, and M. H. Yap, "Analysis of the isic image datasets: Usage, benchmarks and recommendations," *Medical Image Analysis*, vol. 75, p. 102 305, 2022, ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2021.102305. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841521003509.

[69] Y. Po (Harvey) Chin *et al.*, "User satisfaction with a smartphone-compatible, artificial intelligence-based cutaneous pigmented lesion evaluator," *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105 649, 2020, ISSN: 0169-2607.

[70] J. S. Birkenfeld, J. M. Tucker-Schwartz, L. R. Soenksen, J. A. Avilés-Izquierdo, and B. Marti-Fuster, "Computer-aided classification of suspicious pigmented lesions using wide-field images," *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105 631, 2020, ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2020.105631.

[71] P. G. Cavalcanti and J. Scharcanski, "A coarse-to-fine approach for segmenting melanocytic skin lesions in standard camera images," *Computer Methods and Programs in Biomedicine*, vol. 112, no. 3, pp. 684–693, 2013, ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2013.08.010.

[72] A. Finnane, K. Dallest, M. Janda, and H. P. Soyer, "Teledermatology for the Diagnosis and Management of Skin Cancer: A Systematic Review," *JAMA Dermatology*, vol. 153, no. 3, pp. 319–327, Mar. 2017, ISSN: 2168-6068. DOI: 10.1001/jamadermatol.2016.4361. eprint: https://jamanetwork.com/journals/jamadermatology/articlepdf/2588699/jamadermatology_finnane_2016_rv_160006.pdf. [Online]. Available: https://doi.org/10.1001/jamadermatol.2016.4361.

[73] S. Chan, V. Reddy, B. Myers, Q. Thibodeaux, N. Brownstone, and W. Liao, "Machine learning in dermatology: Current applications, opportunities, and limitations," *Dermatology and therapy*, vol. 10, pp. 365–386, 2020.

[74] S Cazzaniga *et al.*, "Mobile teledermatology for melanoma detection: Assessment of the validity in the framework of a population-based skin cancer awareness campaign in northern italy," *JOURNAL OF THE AMERICAN ACADEMY OF DERMATOLOGY*, vol. 81, no. 1, pp. 257–260, 2019.

[75] *Pixelwise learning in convolution*, https://www.analyticsvidhya.com/blog/2022/03/basics-of-cnn-in-deep-learning/, Accessed: 2023-05-27.

[76] A. Vaswani *et al.*, *Attention is all you need*, 2017. arXiv: 1706.03762 [cs.CL].

[77] *Transformer block*, https://www.picsellia.com/post/are-transformers-replacing-cnns-in-object-detection#5-How-and-Why-Are-ViTs-Different-From-CNNs-, Accessed: 2023-05-27.

[78] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, *Training data-efficient image transformers distillation through attention*, 2021. arXiv: 2012.12877 [cs.CV].

[79] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, 2015. arXiv: 1512.03385 [cs.CV].

[80] J. L. Ba, J. R. Kiros, and G. E. Hinton, *Layer normalization*, 2016. arXiv: 1607.06450 [stat.ML].

[81] A. Kumar and P. Shanmugavadivu, "Skin detection using hybrid colour space of rgb-h-cmyk," in *Progress in Advanced Computing and Intelligent Engineering*, B. Pati, C. R. Panigrahi, S. Misra, A. K. Pujari, and S. Bakshi, Eds., Singapore: Springer Singapore, 2019, pp. 75–83, ISBN: 978-981-13-1708-8.

[82] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia, "Human skin detection using rgb, hsv and ycbcr color models," in *Proceedings of the International Conference on Communication and Signal Processing 2016 (ICCASP 2016)*, Atlantis Press, 2016/12, pp. 324–332, ISBN: 978-94-6252-305-0. DOI: 10.2991/iccasp-16.2017.51. [Online]. Available: https://doi.org/10.2991/iccasp-16.2017.51.

[83] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, *Mixup: Beyond empirical risk minimization*, 2018. arXiv: 1710.09412 [cs.LG].

[84] J. Hermans, G. Spanakis, and R. Möckel, *Accumulated gradient normalization*, 2017. arXiv: 1710.02368 [stat.ML].

[85] Z. Zhang and M. R. Sabuncu, *Generalized cross entropy loss for training deep neural networks with noisy labels*, 2018. arXiv: 1805.07836 [cs.LG].

[86] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, *Focal loss for dense object detection*, 2018. arXiv: 1708.02002 [cs.CV].

[87] S. ichi Amari, "Backpropagation and stochastic gradient descent method," *Neurocomputing*, vol. 5, no. 4, pp. 185–196, 1993, ISSN: 0925-2312. DOI: https://doi.org/10.1016/0925-2312(93)90006-O.

[88] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017. arXiv: 1412.6980 [cs.LG].

[89] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell, "A fusion approach for efficient human skin detection," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 138–147, 2011.

[90] *Segments.ai api*, https://segments.ai/, Accessed: 2022-08-01.