Automated Diagnosis of Skin Lesions for Detection of Skin Cancer

A Report Submitted in
Partial Fulfillment of the
Requirements for the Degree of
Bachelor of Technology
in
Biotechnology

by
Abhinav Aggarwal (20200003)

under the supervision of
Dr. Ashutosh Mani

to the

DEPARTMENT OF BIOTECHNOLOGY
MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY ALLAHABAD,
PRAYAGRAJ, UTTAR PRADESH
May, 2024

UNDERTAKING

I declare that the work presented in this report titled "Automated Diagnosis of Skin Lesions for Detection of Skin Cancer" submitted to the Biotechnology Department, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, for the award of the ***Bachelor of Technology*** degree in ***Biotechnology***, is my original work. I have not plagiarized or submitted the same work for the award of any other degree. In case this undertaking is found incorrect, I accept that my degree may be unconditionally withdrawn.

May, 2024
Allahabad

[Abhinav Aggarwal]
20200003

CERTIFICATE

Certified that the work contained in the report titled “Automated Diagnosis of Skin Lesions for Detection of Skin Cancer” has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

Dr. Ashutosh Mani
Biotechnology Dept. MNNIT Allahabad

May, 2024

Preface

Skin cancer has emerged as a substantial global health concern due to the escalating incidence of detrimental ultraviolet radiation in the Earth's atmosphere and the ongoing surge in skin cancer cases. In order to tackle this problem, the project "Automated Diagnosis of Skin Lesions for Detection of Skin Cancer" was undertaken.

The primary objective of this project is to develop an automated, fast and accurate model for diagnosing skin cancer. I used transfer learning to fine-tune a MobileNet model that had been pretrained on images from the 2014 ImageNet Challenge on the HAM10000 dataset, a dataset consisting of 10015 dermatoscopic images of skin lesions. Various deep learning techniques like improved evaluation metrics and data augmentation have been used to improve the model. The model has demonstrated good performance and achieved great accuracy, as detailed in the study. This approach has the ability to help dermatologists in quick decision-making during early or critical stages of skin cancer. A web application has also been developed using the model for use by healthcare professionals.

The results of this study carry significant importance for the identification and diagnosis of skin cancer. By using an accurate and fast model for predicting skin lesion types indicative of potential skin cancer, medical practitioners can intervene promptly, thereby enhancing patient outcomes. Additionally, this research adds to the existing body of knowledge and literature surrounding skin cancer detection, laying the groundwork for future advancements and inquiries in this vital domain.

It is my aspiration that this endeavour will illuminate the efficacy of a fast, accurate, and efficient model for diagnosis of skin lesions for detection of skin cancer. I anticipate that the findings of this study will supplement healthcare, ultimately benefiting individuals at risk of skin cancer and the broader healthcare community.

Acknowledgement

I would like to express my deepest gratitude and appreciation to all those who have supported and guided us throughout the completion of this project. This project would not have been possible without the assistance and encouragement of numerous individuals whom I would like to acknowledge:

My heartfelt thanks to my supervisor, Dr. Ashutosh Mani, for his invaluable guidance, expertise, and unwavering support.

I am immensely grateful to the faculty members of the Biotechnology Department at MNNIT Allahabad, Prayagraj, for their exceptional teaching and mentorship. Their knowledge, expertise, and passion for the subject matter have been a constant source of inspiration throughout my academic journey.

Thank you all once again for your invaluable contributions.

Table of Contents

Chapter 1

Introduction

A skin lesion refers to any abnormal growth or appearance on the skin that deviates from the surrounding tissue. There are two main classifications of skin lesions: primary and secondary. Primary lesions are unusual skin conditions that may emerge over time or be present from birth, while secondary lesions can arise from pre-existing primary lesions that have been exacerbated or altered. For instance, when a mole is scratched until it bleeds, the resulting crust forms a secondary skin lesion. The majority of skin lesions like acne, birthmarks, and moles are benign, meaning they are non-cancerous and pose no harm and thus do not require treatment. However, a small subset may indicate malignancy and thus be cancerous. Despite the generally harmless nature of most skin lesions, it is important to note that a select few can be significantly detrimental to the patient's health, which means that they are malignant and cancerous.

Skin cancers are cancers that arise from the skin. They are due to the development of abnormal cells that can invade or spread to other parts of the body. Skin cancer is the most commonly diagnosed form of cancer in humans. There are three main types of skin cancers: Melanoma, basal-cell skin cancer (BCC) and squamous-cell skin cancer (SCC). The latter two, along with a number of less common skin cancers, are known as nonmelanoma skin cancer (NMSC). Basal-cell cancer grows slowly and can damage the tissue around it but is unlikely to spread to distant areas or result in death. It often appears as a skin lesion characterised with a painless raised area of skin that may be shiny with small blood vessels running over it or may present as a raised area with an ulcer. Squamous-cell skin cancer is more likely to spread. It usually presents as skin lesion with a hard lump and a scaly top which may also form an ulcer. Melanomas are the most aggressive. Signs of melanoma include a skin mole lesion that has changed in size, shape, colour, has irregular edges, has more than one colour, is itchy or bleeds.

Non-invasive skin cancer detection methods used for initial diagnosis include photography, dermatoscopy or confocal microscopy. Dermatologist primarily utilizes visual inspection to diagnose skin cancer, which is a challenging task considering the visual similarity among skin cancers.

Dermatoscopy is also generally used for the diagnosis of skin cancer recently considering its ability to accurately visualize the skin lesions not discernible with the naked eye of a dermatologist. However, the accuracy of diagnosis through these methods depends on a lot of factors and has frequently been found to be poor.

Machine Learning methods could be used to automate the analysis, resulting in a system and framework in the medical field that would aid in providing contextual relevance, improving clinical reliability, assisting physicians in communicating objectively, reducing errors related to human fatigue, lowering mortality rates, lowering medical costs, and more easily identifying diseases. A machine learning method that can categorize both malignant and benign pigmented skin lesions is a step toward achieving these goals. In the proposed work, Convolutional Neural Networks (CNN) and Machine Learning algorithms are used to accurately classify pigmented skin lesions in dermoscopic images to detect malignant skin lesions as early as feasible.

In summary, skin cancer stands as a considerable health concern, carrying the potential for serious health consequences if left inadequately treated. Accurate and quick prediction of skin cancer in its early stages holds paramount importance in averting severe complications and improving patient outcomes. Recent research endeavors have showcased the utility of machine learning algorithms as valuable aids in forecasting the likelihood of developing skin cancer. These algorithms play a pivotal role in aiding healthcare professionals in pinpointing individuals at heightened risk. Leveraging machine learning algorithms for the diagnosis of skin cancer holds immense promise in substantially elevating the accuracy, effectiveness and efficiency of skin cancer management and prevention efforts.

1.1 Motivation

Ultraviolet radiation from sun exposure is the primary environmental cause of skin cancer. More than 90% of cases are caused by exposure to ultraviolet radiation from the Sun. This exposure increases the risk of all three main types of skin cancer. Exposure has increased, partly due to a thinner ozone layer. The researchers had discovered a further 10 percent depletion of the ozone layer will intensify the problem of skin cancer with an additional 300,000 non-melanoma

and 4,500 melanoma cases each year. People with lighter skin are at higher risk. Age and smoking tobacco also increase the risk of skin cancer. HPV infections increase the risk of squamous-cell skin cancer. People with poor immune function such as from medications or HIV/AIDS also have an increased risk of skin cancer.

Skin cancers are the most common groups of cancers diagnosed worldwide. More than 1.5 million new cases of skin cancer were diagnosed in 2022. In 2022, an estimated 330 000 new cases of melanoma were diagnosed worldwide and almost 60 000 people died from the disease. Around 30,00,000 non-melanoma cases are recorded worldwide. The recent study on the prevention of skin cancer reports 90 percent of non-melanoma and 86 percent of melanoma cases induced by excessive exposure of ultraviolet rays.

The most straightforward and effective solution to control the mortality rate for skin cancer is the timely diagnosis of skin cancer as the survival rate for melanoma patients in a five-year timespan is greater than 95 percent when diagnosed and screened at the early stage. Moreover, the most mundane skin cancer types BCC and SCC are highly treatable when early diagnosed and treated adequately. Dermatologist primarily utilizes visual inspection to diagnose skin cancer, which is a challenging task considering the visual similarity among skin cancers. However, dermoscopy has been popular for the diagnosis of skin cancer recently considering the ability of dermoscopy to accurately visualize the skin lesions not discernible with the naked eye. Reports on the diagnostic accuracy of clinical dermatologists have claimed 80 percent diagnostic accuracy for a dermatologist with experience greater than ten years, whereas the dermatologists with experience of 3-5 years were able to achieve diagnostic accuracy of only 62 percent, the accuracy further dropped for less experienced dermatologists. The studies on Dermoscopy imply a need to develop an automated efficient, and robust system for the diagnosis of skin cancer since the fledgling dermatologists may deteriorate the diagnostic accuracy of skin lesions.

Deep learning algorithms have shown exceptional performance in visual tasks and object recognition, which has lead to conduct the research on automated screening of skin cancers. Several studies have been done to compare the dermatologist-level, and Deep Learning based automated classification of skin cancer. Recently, Deep Neural Networks (DNNs) have been found to have an appealing impact on medical image classification. The results of this study carry substantial implications for clinical practice, indicating that machine learning algorithms can provide valuable assistance in predicting the risk of skin cancer

in its early stages. By employing these methods, healthcare professionals can effectively identify individuals who are at a higher risk, facilitating timely interventions to prevent the onset of the disease. Moreover, the utilization of such algorithms enables healthcare providers to allocate resources more efficiently for the prevention and management of skin cancer.

Robust models which are fast, accurate and efficient can be created to identify patients at risk of developing skin cancer. By harnessing the power of these technologies, preventive measures can be implemented proactively. The effective deployment of such models holds the promise of substantially alleviating the burden of skin cancer on individuals, healthcare systems, and economies alike. Ultimately, this endeavour has the potential to enable a healthier and more prosperous society for all.

1.2 Problem Statement

Skin cancer is among the most prevalent and potentially lethal forms of cancer worldwide, with melanoma being particularly aggressive and deadly if not detected and treated early. The prognosis for patients with melanoma is highly dependent on the stage at which the cancer is detected, with early detection significantly improving treatment outcomes and survival rates. However, the increasing incidence of skin cancer poses a significant challenge to healthcare systems, particularly in regions with limited access to specialized dermatological care.

One of the primary barriers to early detection is the shortage of dermatologists and the uneven distribution of these specialists, leading to delays in diagnosis and treatment. This situation underscores the urgent need for effective, accessible, and automated diagnostic tools that can assist healthcare providers in identifying skin cancers early and accurately.

Training of neural networks for automated diagnosis of pigmented skin lesions was hampered by the small size and lack of diversity of available dataset of dermatoscopic images. In order to solve this problem, the HAM10000 ("Human Against Machine with 10000 training images") dataset provides a diverse and comprehensive collection of dermatoscopic images of various skin lesions, including both benign and malignant cases. It includes a representative collection of all important diagnostic categories in the realm of pigmented lesions. This dataset is invaluable for developing machine learning models aimed at skin

cancer detection and classification, offering a robust foundation for training and validating such models.

The primary objective of this project is to develop a robust, accurate, and reliable machine learning model for the detection and classification of skin cancer using the HAM10000 dataset. The model aims to classify different types of skin lesions, effectively distinguishing between benign and malignant cases. By integrating advanced image processing techniques and leveraging the wealth of data provided by the HAM10000 dataset, the project aspires to create a model that can aid dermatologists and general practitioners in making more accurate and timely diagnoses.

The project aims to use this model to create an online tool - a web application which will find the three highest probability diagnoses for a given skin lesion. Using this diagnosis, doctors can quickly and accurately identify high priority patients and speed up their workflow. The app should produce a result in less than 5 seconds. To ensure privacy, the images must be pre-processed and analysed locally and never be uploaded to an external server.

This initiative seeks to contribute to the advancement of automated skin cancer screening tools, thereby improving patient outcomes by enabling earlier detection and treatment. Additionally, such tools can help reduce the workload on healthcare systems and ensure that high-quality diagnostic services are accessible even in regions with limited specialist availability. By harnessing the power of artificial intelligence in medical diagnosis, this project aims to make significant strides in the fight against skin cancer.

Chapter 2

Literature Review

Over the past few years, there has been a noticeable upward trend focus on developing skin cancer diagnosis models to aid in early detection and prevention efforts. This section presents a summary of the key findings from previous studies that have conducted comparative analyses of these models.

Tschandl et al. were responsible for the release of HAM10000 dataset. The HAM10000 dataset is a large collection of multi-source dermatoscopic images of common pigmented skin lesions. The paper outlines how these images are useful for research in dermatology, particularly in the field of skin cancer diagnosis. The dataset includes images of both benign and malignant lesions, covering a wide range of common pigmented skin lesions. The paper also mentions one notable aspect of the HAM10000 dataset as its diversity, since it contains images sourced from multiple institutions and clinical settings. This diversity enhances the dataset's utility for training and evaluating machine learning algorithms for skin lesion classification. Each image in the HAM10000 dataset is accompanied by metadata, including clinical information such as diagnosis, patient age, and lesion location. This metadata provides valuable context for researchers and helps ensure the dataset's usefulness for various research purposes. Overall, the paper outlines how HAM10000 dataset serves as a valuable resource for researchers, clinicians, and data scientists interested in advancing the understanding and diagnosis of skin diseases, particularly melanoma and other forms of skin cancer.

Further research has been conducted which use the HAM10000 dataset for model building and evaluation. These research uses a range of techniques for model training, such as various supervised algorithms and tree-based models like Decision trees and Random forest. Dhivyaa et al. have proposed a model that can

produce feature maps of high resolution that can be used to assist in the preservation of the image's spatial information. On two separate datasets, the authors proposed Random Forest and Decision Tree algorithms for skin class categorization. Polat, Kemal, and Kaan Onur Koc have proposed a system that uses no filtering and feature extraction. Authors claim to have obtained very encouraging results in the identification of skin lesions.

Kumar et al. have proposed a system with RGB color-space, GLCM, and Local Binary Pattern (LBP) methods for pre-processing and image segmentation that uses fuzzy-c clustering and obtained encouraging results in the identification of skin lesions. Adegun and Viriri have proposed a system with an encoder-decoder Conditional Random Field (CRF) module for contour refining and lesion boundary localization that uses a linear combination of Gaussian kernels for paired edge potentials. A Fully Convolutional Network (FCN) with hyper-parameters tuning gave good results.

Data augmentation strategies have been proposed by Srinivasu et al. to balance various forms of lesions to the same range of images. The proposed model, which is predicated on the LSTM and MobileNet V2 approaches, was found to be effective in classifying and detecting skin diseases with little effort and computational resources. When using CNN transfer learning, Mahbod et al. confirmed that image size affects skin lesion categorization performance. They also showed that image cropping outperforms image resizing in terms of performance. Finally, the best classification performance is demonstrated using a simple ensembling strategy that merges the findings from images clipped at six scales and three fine-tuned CNNs. Zhang et al. presented an efficiency result of CNN that was optimized using an upgraded version of the whale optimization technique. This technique is used to find the best weights and biases in the network to reduce the difference between the network output and the desired output.

Hameed et al. proposed a Multi-Class Multilevel (MCML) classification technique inspired by the "divide and conquer" strategy. The proposed classification algorithm combines machine learning and deep learning methods. Hasan et al. proposed the DSNet, an automatic semantic segmentation network for skin lesions. To reduce the number of parameters and make the network lighter, they used a separable depth-wise convolution. Hosny et al. used pre-trained AlexNet with transfer learning. As initial values, the parameters from the original model were used. and the weights of the last three replaced layers were randomly initialized. ISIC 2018, the most recent public dataset, was used to test

the suggested technique. Chatterjee et al. used SVM with RBF to identify three lesions by extracting form, fractal dimension, texture, and color variables using fractal-based regional texture analysis (FRTA). Pereira et al. proposed integrating gradients with the local binary patterns (LBP) technique to increase the performance of skin lesion classification algorithms and to further exploit the border-line properties of the lesion segmentation mask.

A kernel sparse coding approach for both segmentation and classification of skin lesions. A skin lesion detection system was proposed by Garcia-Arroyo et al.[17] with fuzzy histogram thresholding, lesions were segmented. Using the ABCD rule, Zaqout has developed a model capable of partitioning and classifying skin pictures. In the majority of these efforts, the ABCD rule is followed, which includes image pre-processing, segmentation to locate the lesion, feature extraction, and classification. TDS is a dermoscopy score that assists in the diagnosis of the lesion condition. Khan et al. have proposed a system that uses the deep learning model MASK-RCNN for segmentation and pre-trained DenseNet for classification. Using YOLOv5, Shelatkar et al. describe a deep learning-based method for classifying and identifying brain tumours. To extract the features, a transfer learning concept is used which is then exposed to selection and classification stages. Khan et al. developed an automated method for skin lesion categorization that used pre-trained RESNET-50 and RESNET-101 deep neural network (DCNN) with transfer learning was employed for feature extraction and optimal feature selection based on the kurtosis-controlled principle component (KcPCA). The information was then combined and the best features were chosen, which were then fed into a supervised learning algorithm—SVM of kernel function radial basis function (RBF) for classification.

Khan et al. developed a segmentation and classification framework based on deep learning. For skin lesion segmentation, a MASK R-CNN-based architecture with a Resnet50 feature pyramid network (FPN) is used. The final mask is then generated by mapping connected layer-based features. A 24-layer convolutional neural network architecture is built during the classification phase, with activation based on the display of higher characteristics. Finally, the best CNN features are delivered to softmax classifiers for final classification. Harris Hawks Optimization with Deep Learning Model for Detection of Diabetic Retinopathy was proposed by Gundluru et al., Tajeddin et al. used highly discriminative characteristics to classify skin melanoma. For lesion segmentation, the authors started with contour propagation. To extract features, lesions were mapped in log-polar space using Daugman's transformation based on the peripheral area. Finally, the various classifiers used in the proposed work were evaluated.

To construct a dermoscopic skin image recognition system, Yu et al. recommended CNN and the local descriptor encoding approach. To extract skin lesion features from images, the authors utilized ResNet101 and ResNet50. Using a Fisher vector (FV) and the collected ResNet features, a global image representation was generated. Finally, a Chi-squared kernel was applied in an SVM for classification. Melanoma was categorized into three groups based on the thickness of the lesion. The researchers utilized two categorization schemata: one classified lesion as thin or thick, and the other separated them into thin, moderate, and thick categories. To categorize the lesion data, a combination of ANN and logistic regression algorithms is used. Using a cloud computing system, Rajput et al. suggested diabetes diagnosis to Indians living in rural locations. To construct a breast cancer detection system, Abbas et al. proposed the Extremely Randomized Tree and Whale Optimization Algorithm (WOA) to present a unique approach called BCD-WERT for efficient feature selection and classification. To detect ships from satellite imaginary, deep CNN with YOLOv3 for object detection with SHA-256 hashing for security to the detected images. For the purpose of diagnosing heart disease, Reddy et al. suggested a hybrid genetic algorithm and a fuzzy logic classifier.

To conclude, many researchers have contributed to classifying the skin lesion categories using distinct machine learning and deep learning approaches. Also, they have worked on a variety of datasets. The performance of these models depends on various factors, including the dataset used for training, the machine learning or deep learning technique used and the metrics and objectives used for model development. Performance of the models developed so far has been severely limited mostly due to the small datasets used. Models developed using HAM10000 have been limited and also show limited performance due to poor training and evaluation metrics.

Chapter 3

Material & Methods

In this chapter, the material and methodology used for implementing this project is discussed. The goal of this project was to build an efficient and automated machine learning model for skin cancer diagnosis using the HAM10000 dataset, which is a large collection of 10015 multi-source dermatoscopic images of common pigmented skin lesions. It includes a representative collection of all important diagnostic categories in the realm of pigmented lesions. The lesions span across both sexes, various ages, and various locations on the body. More than 50% of lesions are confirmed through histopathology. The dataset includes lesions with multiple images, which can be tracked by the lesion_id-column within the HAM10000_metadata file. MobileNet model, pretrained on the ImageNet Challenge for image classification, has been further finetuned on the HAM10000 dataset using transfer learning, for developing an efficient, fast and accurate model for skin cancer diagnosis. The model was further used to develop a web application for easy use by healthcare specialists. The code is available at https://github.com/agg-geek/skincancer.

3.1 Data Collection

In this study, data collection for skin cancer diagnosis model was conducted by retrieving data from the HAM10000 ("Human Against Machine with 10000 training images") dataset. The dataset was used in the "The International Skin Imaging Collaboration (ISIC) 2018" classification challenge hosted by the annual MICCAI conference in Granada, Spain. It was later released publicly and is available for research on the Harvard Dataverse. The decision to utilize this repository was based on the need for a comprehensive and well-curated dataset that would enable the development and evaluation of an accurate skin cancer diagnosis model.

The HAM10000 dataset contains 10015 high quality, close-up dermatoscopic images of skin lesions aimed for training programs for skin cancer detection. The dataset contains 7 classes of skin lesions that can be comprehensively used for skin cancer diagnosis. These seven classes provide a more robust foundation for the development of the skin cancer diagnosis model. The dataset also contained data related to identification of the lesion images as well as patient data like age, sex, etc in metadata files.

The dermatoscopic images in HAM10000 dataset were collected over a period of 20 years from two different sites, the Department of Dermatology at the Medical University of Vienna, Austria, and the skin cancer practice of Cliff Rosendahl in Queensland, Australia. The Australian site stored images and meta-data in PowerPoint files and Excel databases. The Austrian site started to collect images before the era of digital cameras and stored images and metadata in different formats during different time periods.

The images and the metadata were processed through a data-preprocessing pipeline including the Extraction of images and meta-data from PowerPoint files, Digitization of diapositives, Extraction of data from a digital dermatoscopy system, Filtering of dermatoscopic images, Unifying pathologic diagnoses and Manual quality review. The use of an established dataset like HAM10000 contributed to the credibility and validity of the data, as it was sourced from a trusted and reputable platform, the ISIC.

In order to make the dataset suitable for research, analysis and model development purposes, the original data in PowerPoint, Excel and different image formats were processed by researchers using a combination of automation, machine learning and manual review to extract all the images into a standard JPG format and the associated metadata into a tab file. In addition, researchers also filtered away bad data, such as low-quality images, images with poor or no annotations, and non-dermatoscopic images. Furthermore, differences in the images were addressed and converted such that each image had similar luminosity / hue with a centered lesion. Proper copyrighting and licensing were produced to make the dataset public.

Summing up, the data collection process for the skin cancer diagnosis model involved retrieving data from HAM10000 dataset from the Harvard Dataverse. The dataset was carefully chosen to align with the objectives of the study and included a comprehensive range of variables related to skin cancer. Through thorough preprocessing, the collected data were made suitable for analysis and

ensured the privacy and confidentiality of the individuals represented in the dataset. The utilization of the HAM10000 dataset repository provided a robust foundation for the development of an accurate and reliable skin cancer diagnosis model, facilitating informed decision-making in skin cancer management and prevention.

3.2 Description of Dataset

The HAM10000 dataset contains 10015 close-up, high quality dermatoscopic images of skin lesions. Skin lesions are perfectly centered in the image, which reduces the data augmentation step. All the images have colour. The images have height 450 pixels and width 600 pixels, which is very convenient since the images don't need initial standardizing to make them of the same size for modelling.

The 7 diagnostic categories of skin lesions present in HAM10000 used for skin cancer diagnosis are:

1. **Actinic Keratoses (Solar Keratoses) and Intraepithelial Carcinoma (Bowen's disease) – `akiec`**
   They are common non-invasive, variants of squamous cell carcinoma that can be treated locally without surgery. Some authors regard them as precursors of squamous cell carcinomas and not as actual carcinomas. There is, however, agreement that these lesions may progress to invasive squamous cell carcinoma – which is usually not pigmented. Both neoplasms commonly show surface scaling and commonly are devoid of pigment. Actinic keratoses are more common on the face and Bowen's disease is more common on other body sites. Because both types are induced by UV-light the surrounding skin is usually typified by severe sun damaged except in cases of Bowen's disease that are caused by human papilloma virus infection and not by UV. Pigmented variants exist for Bowen's disease and for actinic keratoses, and both are included in this set. The dermatoscopic criteria of pigmented actinic keratoses and Bowen's disease are described in detail by Zalaudek et al. and by Cameron et al.

2. **Basal Cell Carcinoma - `bcc`**
   Basal cell carcinoma is a common variant of epithelial skin cancer that rarely metastasizes but grows destructively if untreated. It appears in different morphologic variants (flat, nodular, pigmented, cystic), which are described in more detail by Lallas *et al*.

3. **Benign Keratoses – `bkl`**
   "Benign keratosis" is a generic class that includes seborrheic keratoses ("senile wart"), solar lentigo - which can be regarded a flat variant of seborrheic keratosis - and lichen-planus like keratoses (LPLK), which corresponds to a seborrheic keratosis or a solar lentigo with inflammation and regression. The three subgroups may look different dermatoscopically, but were grouped together because they are similar biologically and often reported under the same generic term histopathologically. From a dermatoscopic view, lichen planus-like keratoses are especially challenging because they can show morphologic features mimicking melanoma and are often biopsied or excised for diagnostic reasons. The dermatoscopic appearance of seborrheic keratoses varies according to anatomic site and type.

4. **Dermatofibroma - `df`**
   Dermatofibroma is a benign skin lesion regarded as either a benign proliferation or an inflammatory reaction to minimal trauma. The most common dermatoscopic presentation is reticular lines at the periphery with a central white patch denoting fibrosis.

5. **Melanoma - `mel`**
   Melanoma is a malignant neoplasm derived from melanocytes that may appear in different variants. If excised in an early stage it can be cured by simple surgical excision. Melanomas can be invasive or non-invasive (in situ). All variants of melanoma including melanoma in situ were included, but non-pigmented, subungual, ocular or mucosal melanoma were excluded. Melanomas are usually, albeit not always, chaotic, and some melanoma specific criteria depend on anatomic site.

6. **Melanocytic Nevi - `nv`**
   Melanocytic nevi are benign neoplasms of melanocytes and appear in a myriad of variants, which all are included in our series. The variants may differ significantly from a dermatoscopic point of view. In contrast to

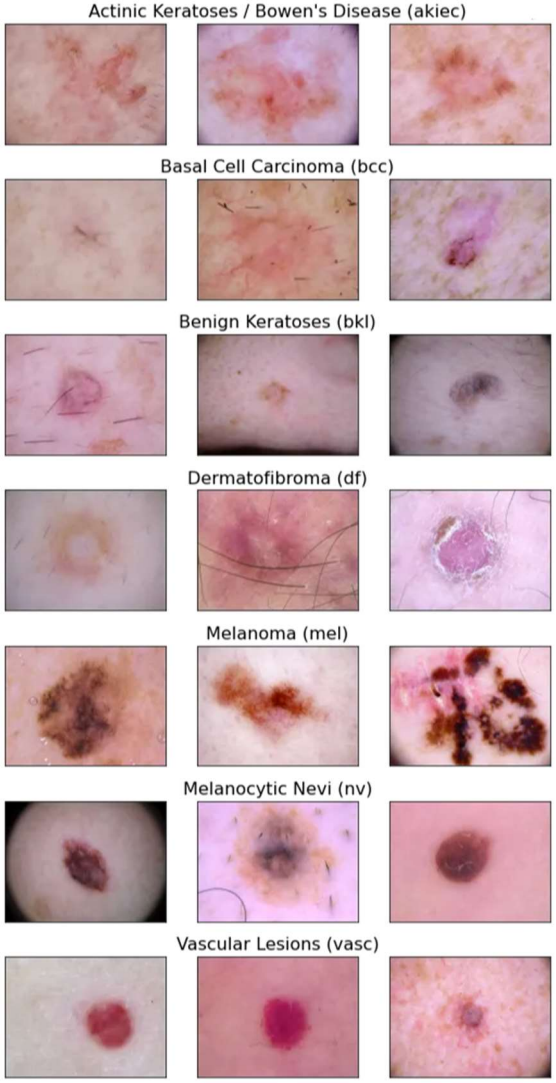
melanoma, they are usually symmetric with regard to the distribution of colour and structure.

7. **Vascular Lesions - `vasc`**
   Vascular skin lesions in the dataset range from cherry angiomas to angiokeratomas and pyogenic granulomas. Hemorrhage is also included in this category.

The malignant forms of skin lesions are:

1. Melanoma - mel
2. Basal Cell Carcinoma - bcc
3. Actinic Keratoses / Bowen's Disease - akiec

Below are some sample images for each of the 7 diagnostic classes:



The diagnosis of HAM10000 images were validated and verified in four ways:

1. **Histopathology – `histo`**
   Histopathologic diagnoses of excised lesions have been performed by specialized dermatopathologists. Researchers scanned all available histopathologic slides of the current ViDIR image set for later review. Researchers manually reviewed all images with the corresponding histopathologic diagnosis and checked for plausibility. If the histopathologic diagnosis was implausible, they checked for sample mismatch and reviewed the report and reexamined the slide if necessary.

2. **Confocal - `confocal`**
   Reflectance confocal microscopy is an in-vivo imaging technique with a resolution at near-cellular level, and some facial benign keratoses were verified by this method. Most cases were included in a prospective confocal study conducted at the Department of Dermatology at the Medical University of Vienna that also included follow-up for one year.

3. **Follow-up – `follow-up`**
   Individuals with nevi monitored via digital dermatoscopy were deemed to exhibit biologic benignity if they displayed no alterations over the course of three follow-up visits or within a span of 1.5 years. This classification was exclusively applied to nevi, as opposed to other benign diagnoses, given that dermatologists typically do not conduct monitoring procedures for dermatofibromas, seborrheic keratoses, or vascular lesions. Evaluation of changes was carried out by authors possessing over 20 years of experience in digital dermatoscopic follow-up examinations.

4. **Consensus - `consensus`**
   In cases of typical benign instances lacking histopathological analysis or follow-up assessments, an expert-consensus rating was offered by experienced authors. This consensus designation was assigned only when all authors independently provided an unequivocal benign diagnosis. Lesions categorized under this ground-truth classification were typically photographed for educational purposes and did not necessitate additional follow-up or biopsy procedures for confirmation.

The dataset also contains metadata associated with each image in a metadata file. It contains data for identification of the lesion image and patient data.

1. **`lesion_id`**: refers to the patient associated with the skin lesion. It is in the form of HAM_X where X is a 7-digit unique identifier number (multiple skin lesions can come from the same patient)
2. **`image_id`**: refers to the image name in the HAM10000 data folders. It is in the form ISIC_X where X is a 7-digit unique identifier number (each image has its own ISIC number). This will be useful for finding the image in the data directory.
3. **`dx`**: refers to the diagnosis. The column uses the following abbreviation scheme:
    1. **`akiec`**: Actinic keratoses / Bowen's disease
    2. **`bcc`**: basal cell carcinoma
    3. **`bkl`**: benign keratosis-like lesions
    4. **`df`**: dermatofibroma
    5. **`mel`**: melanoma
    6. **`nv`**: melanocytic nevi
    7. **`vasc`**: vascular lesions
4. **`dx_type`**: refers to how the skin lesion is verified. It will be either
    1. **`histo`**: confirmed via histopathology
    2. **`confocal:`** confirmed via confocal microscopy
    3. **`consensus:`** confirmed via consensus of experts
    4. **`follow-up`**: confirmed via follow-up
5. **`age`**: refers to the age of the patient associated with the skin lesion image
6. **`sex`**: refers to the sex of the patient associated with the skin lesion image (will be `male`, `female`, or `unknown`)
7. **`localization`**: refers to the location of the skin lesion on the body (e.g., back, hand, scalp, etc.)
8. **`dataset`**: refers to the origin dataset for the skin lesion (the site / group it was collected from)

Since the aim of this project is to diagnose skin cancer from the skin lesion, the key features of interest to analyze would be the images, the labels (dx), and the demographic / patient information (age, sex, and localization). The other variables, while insightful, provide mostly information on how the data was collected, verified, and organized, which is not useful in our current analysis.

	lesion_id	image_id	dx	dx_type	age	sex	localization
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear

Figure. The first five rows of the metadata file of the dataset.

The dataset's size of 10015 images which represent a comprehensive and representative collection of all important diagnostic categories in the realm of pigmented lesions, along with the associated features useful for skin cancer diagnosis, provided a robust foundation for training and evaluating the diagnostic models. This extensive dataset allowed for building a fast, accurate and efficient model for skin cancer diagnosis.

3.3 Data Preprocessing

The pre-processing of skin lesion images was done by using Keras ImageDataGenerator. The 57 null `age` entries in the dataset were filled using the 12 mean filling method (figure 3.3a). The Dermoscopy images in the dataset were downscaled to 224 × 224 pixel resolution from 600 × 450 pixel resolution to make images compatible with the MobileNet model. The 10015 images in the dataset were split into the training set (9077 images) and validation set (938 images). The dataset images with no duplication in training data were selected for the validation set so that the authenticity in the validation process can be maintained.

```
df_eda.isna().sum()
```

```
lesion_id         0
image_id          0
dx                0
dx_type           0
age              57
sex               0
localization      0
cell_type         0
cell_type_idx     0
dtype: int64
```

```
# df_eda['age'].fillna((df_eda['age'].mean()), inplace=True)
df_eda['age'] = df_eda['age'].fillna(df_eda['age'].mean())
```

```
df_eda.isna().sum().sum()
```

```
0
```

Figure 3.3a. Filling null values in age.

	lesion_id	image_id	dx	dx_type	age	sex	localization	duplicates
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp	has_duplicates
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp	has_duplicates
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp	has_duplicates
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp	has_duplicates
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear	has_duplicates

```
df_data['duplicates'].value_counts()
```

```
no_duplicates     5514
has_duplicates    4501
Name: duplicates, dtype: int64
```

```
# Filter out images that don't have duplicates
df = df_data[df_data['duplicates'] == 'no_duplicates']
df.shape
```

```
(5514, 8)
```

Figure 3.3b. Removal of duplicated images in the dataset used for further image augmentation.

3.4 Explorative Data Analysis

Extensive Exploratory Data Analysis (EDA) was conducted to gain comprehensive insights into the HAM10000 dataset. The EDA involved both univariate analysis, which examines each feature individually, and bivariate analysis, which explores the relationships between pairs of features. This thorough analysis enabled the identification of hidden patterns and key characteristics within the dataset. The insights obtained from the EDA were instrumental in refining the model. These insights guided the modification and training of the model, enhancing its performance and reducing the likelihood of overfitting. Additionally, the EDA facilitated the selection of appropriate evaluation metrics, ensuring a robust assessment of model performance.

A few important EDA plots have been highlighted below, along with the relevant observations that can be interpreted from the plot.

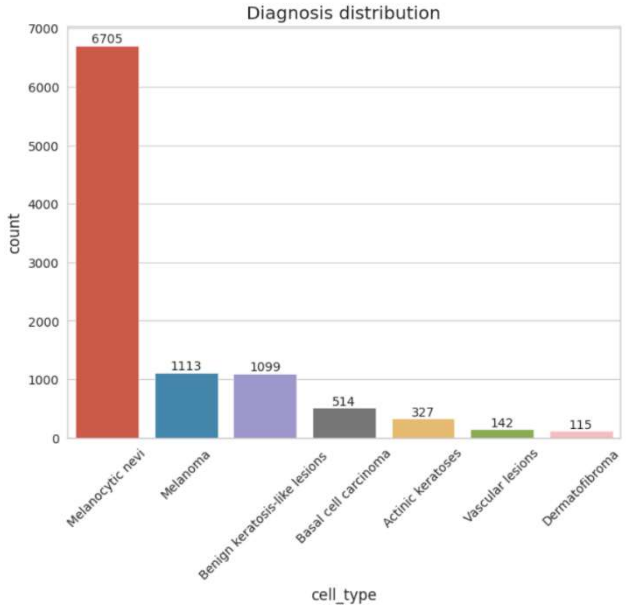


Figure 3.4a Count of diagnoses

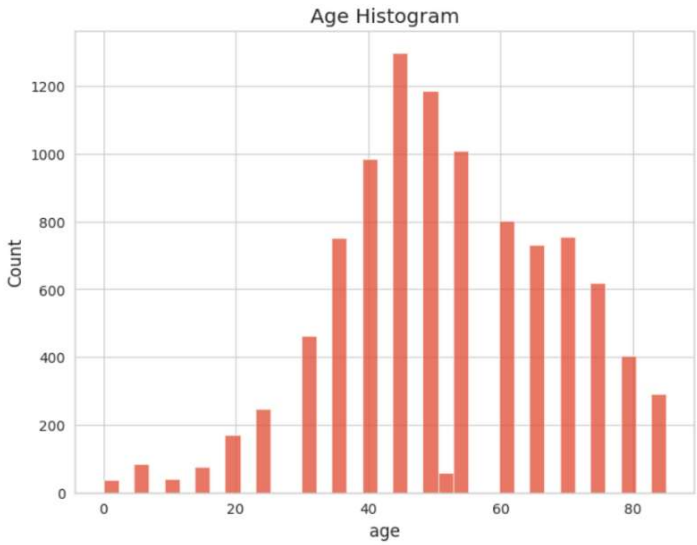


Figure 3.4b Age histogram

Observations for ***figure 3.4a***

1. The dataset is extremely unbalanced.
2. There are vast number of cases of Melanocytic nevi as compared to others. This `nv` category has 6705 images in the dataset, which means it has more representation than all the other categories combined.

3. Melanoma and Benign keratosis-like lesions are quite less wide spread as compared to Melanocytic nevi, which is important as Melanoma is cancerous.
4. Model should be evaluated while considering this skewness, especially for Melanoma.

Observations for ***figure 3.4b***

1. The range of ages in the HAM10000 dataset is from 0 to 85 years.
2. The distribution appears to be bimodal, with a large number of patients in the 35–50 age range and 60–75 age range.
3. Most of the patients in the dataset are generally older. This distribution matches the general trend about skin cancer, in that it typically affects older individuals.
4. There doesn't seem to be any anomalies in the ages.

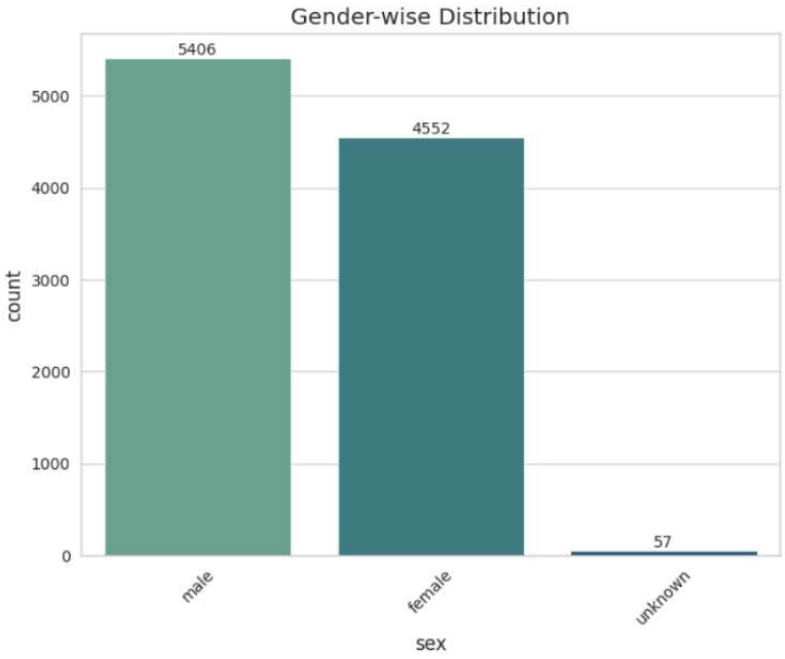


Figure 3.4c. Gender-wise distribution

1. The male category has the most representation and there are roughly 1000 more images (10% more representation) for the male category than the female category.
2. Ideally, there should be an equal representation of both sexes, but since the imbalance is relatively small, it requires no investigation.

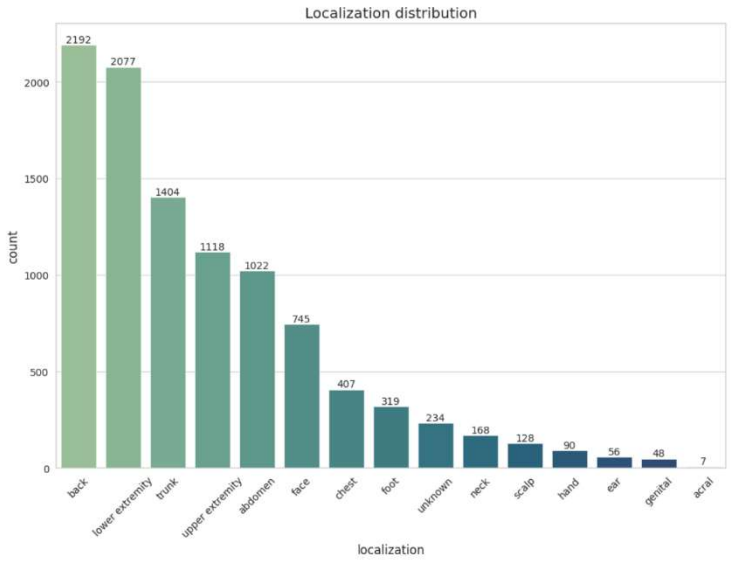


Figure 3.4d Localization

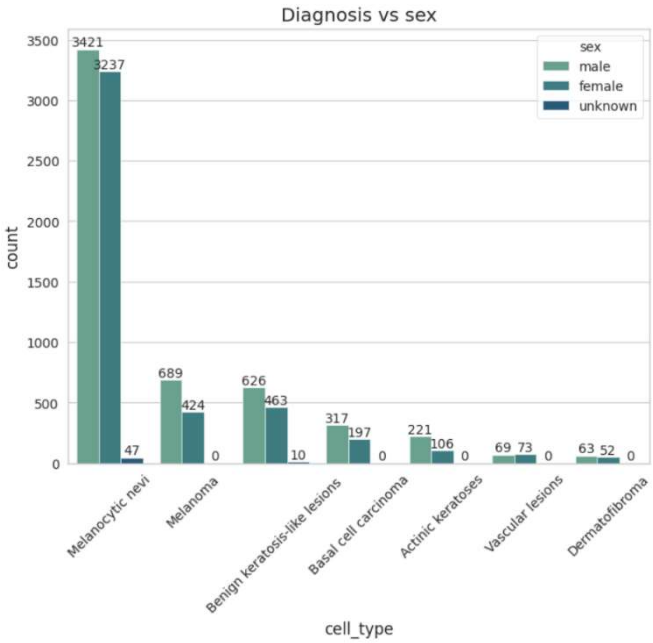


Figure 3.4e Diagnosis vs sex

Observations for ***figure 3.4d***

1. The localization of the skin lesion is very imbalanced.
2. A large number of skin lesions are located on the back, lower extremity, and the trunk
3. In contrast, the number of lesions in acral areas, genital, and ears combined barely cross 100 examples (~1% of the dataset).
4. Since skin lesions can look very different depending on the part of the body, this imbalance can ultimately affect model's performance and its ability to predict correctly on skin lesions found in less represented locations in the body.
5. This can be a future area of improvement for developing better models.

Observations for ***figure 3.4e***

1. Some diagnoses `akiec`, `bcc`, `bkl` and `mel` have a greater difference between male and female which might indicate that these types of skin cancer is more prevalent in males than females.
2. This conclusion should be taken with a pinch of salt as this might be a reflection of distribution imbalance in the broader dataset due to the following possible biases:
    - There are more skin lesions in males than females, so naturally, there will be more male examples.

- There isn't enough data of a certain diagnosis.

3. There can be use of statistical tests like proportion tests to see if there's significant differences in sex representation across the different diagnoses.
4. Overall, the EDA played a crucial role in understanding the dataset, guiding model development, and ensuring effective performance evaluation.

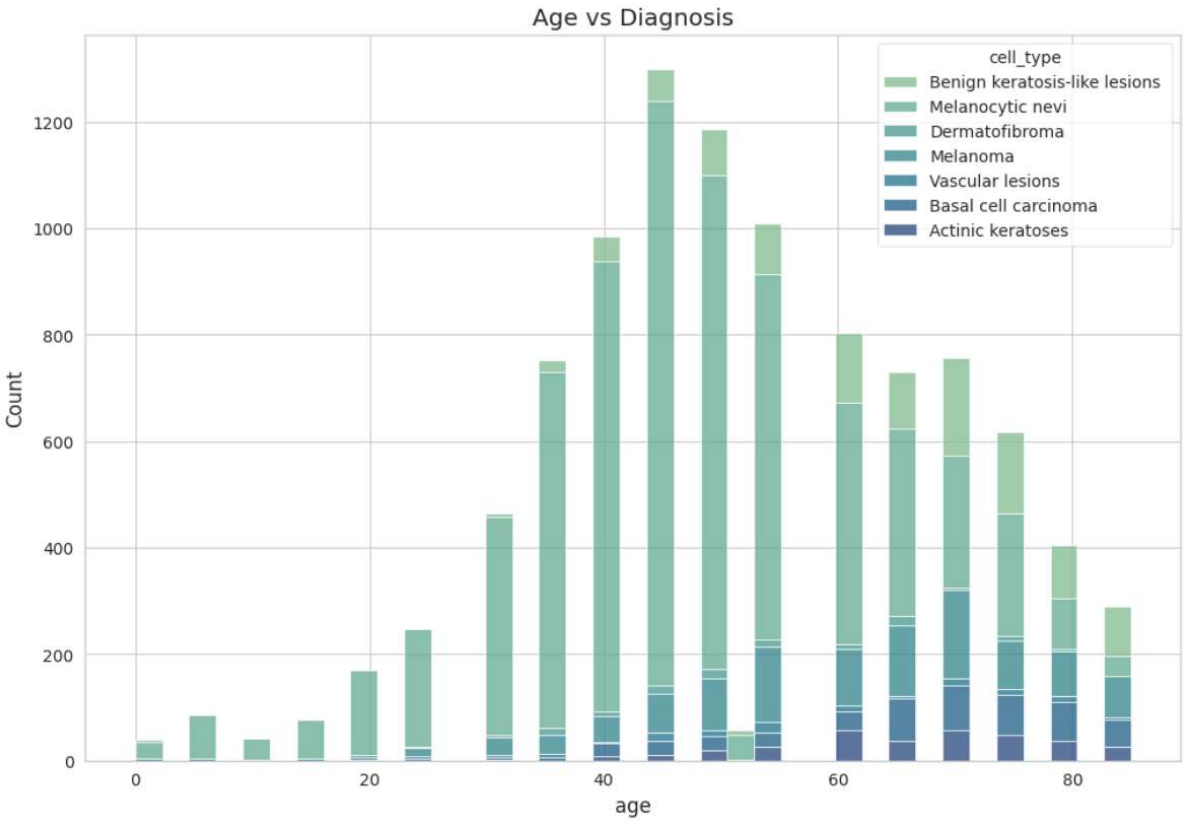


Figure 3.4f Age vs diagnosis

Observations for ***figure 3.4f***

1. Some diagnoses only have people over a certain age, such as `akiec` where the youngest patients recorded was around 30 years old.
2. This is very different from `vasc` and `nv` which have very young patients.
3. This indicates that some cancer diagnoses are more likely to affect older patients.
4. Similar to age vs diagnosis, problems in data distribution and bias may be present.

5. For future investigation, statistical tests like ANOVA test can be used to check for significant differences in age across the different diagnoses.
6. vasc and nv are non-cancerous / benign diagnosis which occur in young patients. Thus, another conclusion could be that cancerous / malignant diagnosis could affect older ages.
7. This can be tested this by differentiating diagnoses into a non-cancerous and cancerous group and then using statistical tests like t-test for comparison and verification.

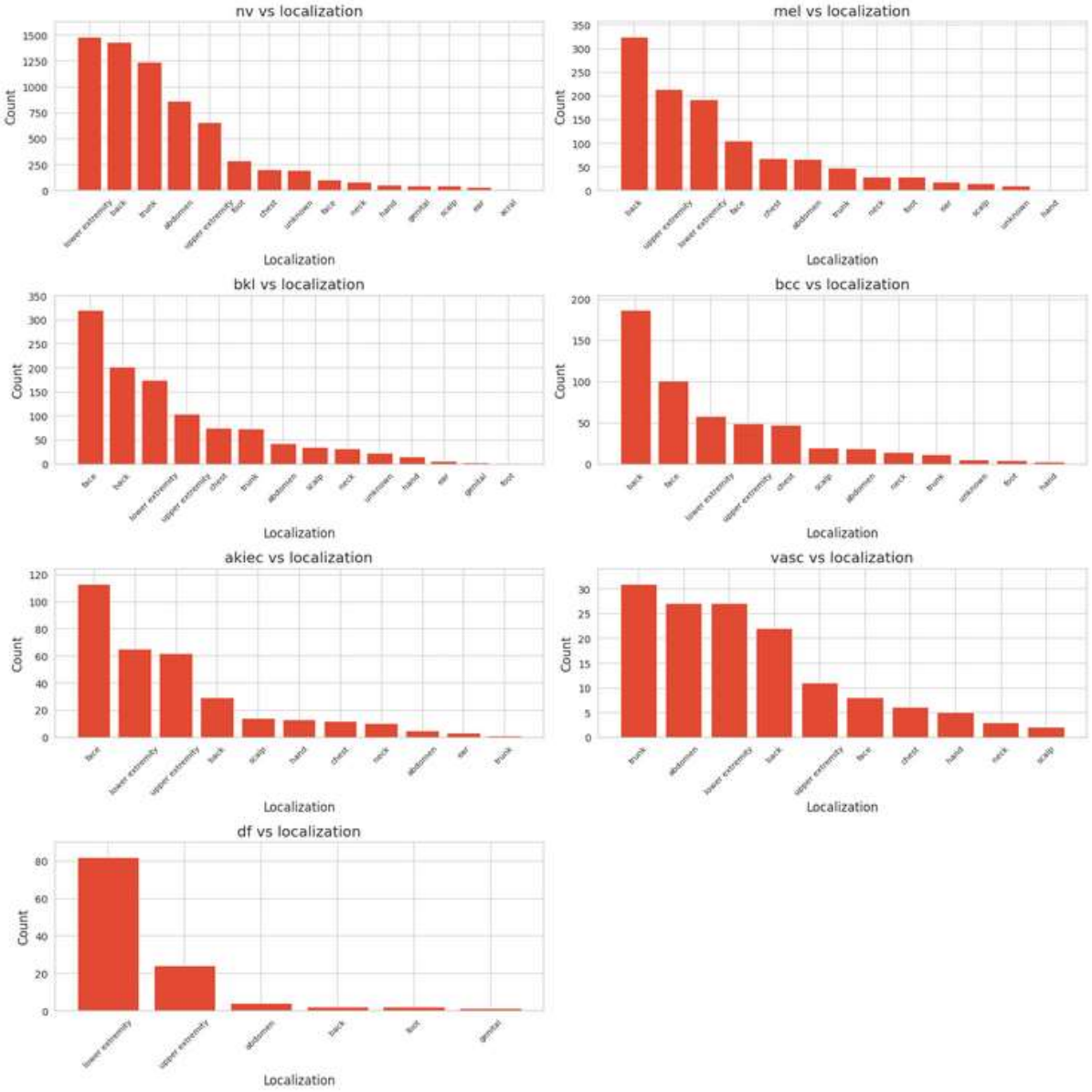


Figure 3.4g Diagnosis vs localization

Observations for ***figure 3.4g***

1. `df` seems to exclusively happen on the lower and upper extremity.
2. The back is home to many of the diagnoses, except for `df` and `akiec`, both of which have a relatively smaller percentage of examples located there
3. The face seems to be the most prevalent location for `akiec` and `bkl`

Overall, the EDA was pivotal in understanding the dataset, guiding model development, and ensuring effective performance evaluation. Insights from the EDA informed the definition of appropriate evaluation metrics and highlighted critical issues such as class imbalance among the seven categories of skin lesions. A significant finding from the EDA was the underrepresentation of Melanoma images in the dataset. Given that Melanoma, despite being less common, is responsible for the majority of skin cancer-related deaths compared to non-melanoma skin cancers like Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC), addressing this imbalance was essential. To mitigate this, the class weight for Melanoma was adjusted to be three times that of the other classes, making the model more sensitive and improving its ability to accurately detect Melanoma. By incorporating these adjustments obtained through EDA ensured that the model was better equipped to handle the dataset's specific characteristics, ultimately enhancing its predictive accuracy and reliability.

3.5 Data Augmentation

HAM10000 dataset has an unbalanced distribution of images among the seven classes. Data Augmentation brings an opportunity to rebalance the classes in the dataset, alleviating other minority classes. Data Augmentation is an effective means to expand the size of training data by randomly modifying several parameters of training data images like rotation range, zoom range, horizontal and vertical flip, fill_mode, etc. I conducted data augmentation of minority classes in the dataset: Melanoma, Benign Keratosis, Basal Cell Carcinoma, Actinic Keratosis, vascular lesion, and dermatofibroma to generate

approximately 6000 images in each class giving a total of 38,569 images in the training set.

```
# Create a data generator
datagen = ImageDataGenerator(
    rotation_range=180,
    width_shift_range=0.1,
    height_shift_range=0.1,
    zoom_range=0.1,
    horizontal_flip=True,
    vertical_flip=True,
    #brightness_range=(0.9,1.1),
    fill_mode='nearest')

batch_size = 50

aug_datagen = datagen.flow_from_directory(path,
                                          save_to_dir=save_path,
                                          save_format='jpg',
                                          target_size=(224,224),
                                          batch_size=batch_size)

# Generate the augmented images and add them to the training folders
num_aug_images_wanted = 6000 # total number of images we want to have in each class
```

Figure 3.5a Parameters used for Image Data Augmentation

The ImageDataGenerator class from the Keras library is used to create a data generator that performs data augmentation on the images in a specified directory. Data augmentation parameters used for the current work is given below:

1. `rotation_range=180`: Randomly rotates the images within a range of 0 to 180 degrees.
2. `width_shift_range=0.1`: Shifts the images horizontally by up to 10% of the total width.
3. `height_shift_range=0.1`: Shifts the images vertically by up to 10% of the total height.
4. `zoom_range=0.1`: Zooms in or out on the images by up to 10%.
5. `horizontal_flip=True`: Randomly flips the images horizontally.
6. `vertical_flip=True`: Randomly flips the images vertically.
7. `fill_mode='nearest'`: Specifies the strategy for filling in new pixels that may appear after a transformation. Here, it uses the nearest pixel value.

The image are resized to 224 × 224 pixels and processed in batches of 50.

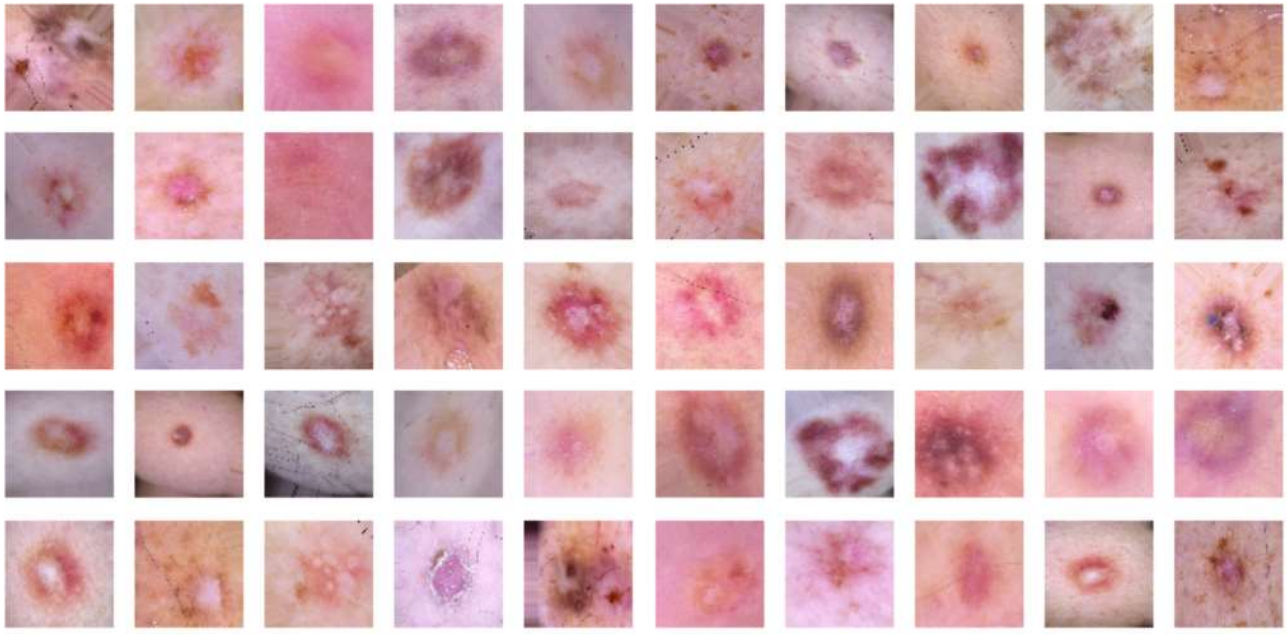


Figure 3.5b Fifty sample images after performing augmentation

3.6 Model Training

The MobileNet model is ideal for mobile and embedded vision applications as they have lightweight DNN architecture. I used MobileNet convolutional neural network pretrained on 12,80,000 images containing 1,000 object classes from the 2014 ImageNet Challenge. The 25 layered MobileNet architecture was constructed for the current study, which employs four Conv2D layers, seven BatchNormalization layers, seven ReLU layers, three ZeroPadding2D layers, and single DepthwiseConv2D, GlobalAveragePooling, Dropout, and Dense layers as shown below. The training of the model was done on a training set of 38,569 images using Transfer Learning with batch size and epochs as 10 and 50 respectively. The Categorical Crossentropy loss function, Adam optimizer and metric function Accuracy, Top2 accuracy, and Top3 accuracy were used to evaluate MobileNet model performance.

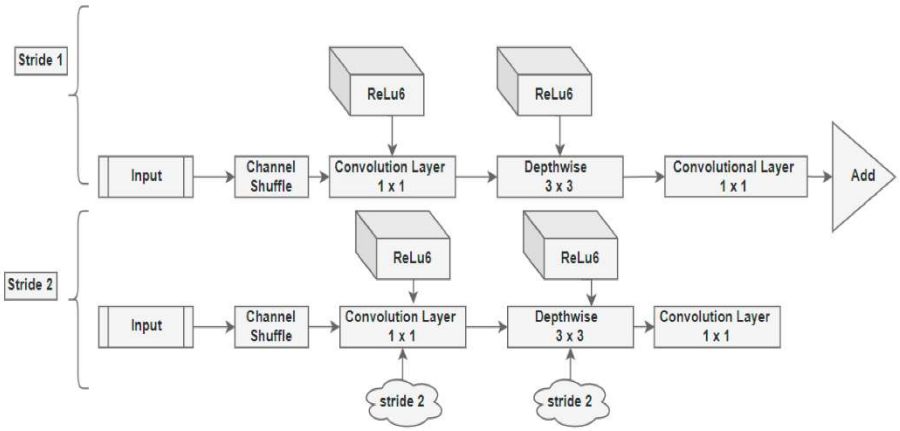


Figure 3.6a The architecture of MobileNet V2 model

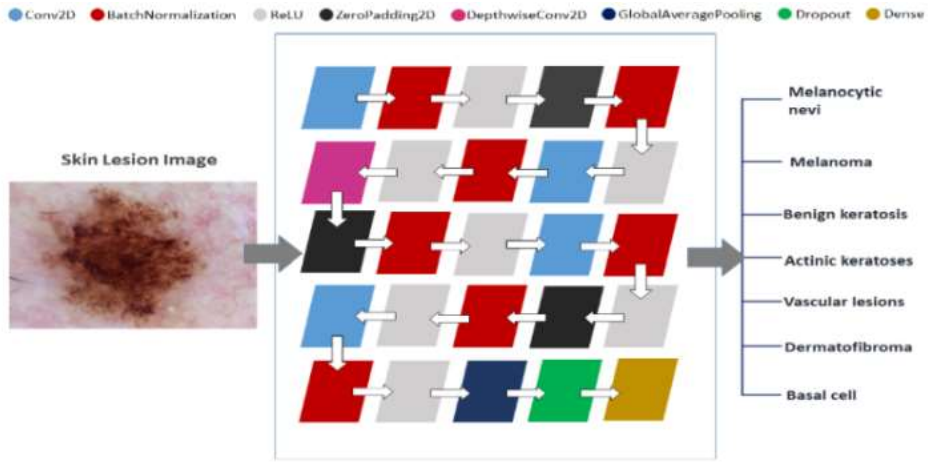


Figure 3.6b MobileNet V2 architecture for current seven categories of classification

Categorical Crossentropy Loss Function

The categorical crossentropy loss function is commonly used in classification tasks, particularly when the target variable is categorical and the model outputs probabilities for each class. It measures the difference between the true labels and the predicted probabilities, providing a quantifiable way to assess the performance of a classification model.

For a multi-class classification problem with N samples and C classes, the categorical crossentropy loss is defined as:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(p_{i,c})$$

1. **Predicted Probabilities**: The model outputs a probability distribution across all classes for each input sample. This is usually achieved using a softmax activation function in the output layer of the neural network.
2. **True Labels**: The true labels are one-hot encoded vectors where the correct class for each sample is marked as 1, and all other classes are marked as 0.
3. **Logarithmic Penalty**: The logarithm of the predicted probability of the correct class is taken. If the predicted probability for the correct class is high, the logarithm will be closer to zero, indicating a lower penalty. Conversely, if the probability is low, the logarithm will be more negative, indicating a higher penalty.
4. **Sum and Average**: The sum of these penalties across all classes and samples is taken, and then averaged to get the final loss value.

Adam Optimizer

The Adam optimizer is an advanced optimization algorithm used for training deep learning models. It combines the best features of the AdaGrad and RMSProp algorithms to provide an efficient and effective approach to gradient-based optimization. The name "Adam" is derived from "Adaptive Moment Estimation."

1. **Adaptive Learning Rate**: Adam computes individual adaptive learning rates for different parameters. This is achieved through estimates of first and second moments of the gradients.
2. **Momentum**: Adam incorporates the concept of momentum by using moving averages of the gradients to smooth the optimization process and accelerate convergence.
3. **Bias Correction**: Adam includes bias correction terms to counteract the biases introduced during the early stages of training, which helps improve performance and stability.

Accuracy

Accuracy is the most common metric for evaluating classification models. It measures the proportion of correct predictions out of all predictions made. For a multi-class classification problem, accuracy is calculated by comparing the predicted class label to the true class label for each instance and counting the number of correct predictions.

Top2 Accuracy

Top-2 Accuracy is a metric that measures the proportion of times the true class label is among the top two predicted class labels with the highest predicted probabilities. Top-2 accuracy is useful in scenarios where the model is allowed some leeway in its predictions. If the correct class label is within the top two predicted probabilities, it counts as a correct prediction.

Top3 Accuracy

Top-3 Accuracy is similar to Top-2 Accuracy but extends to the top three predicted class labels. It measures the proportion of times the true class label is among the top three predicted class labels with the highest probabilities. Top-3 accuracy further relaxes the evaluation criteria, allowing the model to be considered correct if the true label is within its top three predicted classes.

Dropout

Dropout is a regularization technique used to prevent overfitting in neural networks by randomly dropping out (setting to zero) a proportion of the input units (neurons) during training. This forces the network to learn redundant representations of features, making it more robust and less likely to overfit to the training data.

Dense

A Dense layer, also known as a fully connected layer, is a type of layer in a neural network where each neuron is connected to every neuron in the previous layer. In other words, each neuron in a Dense layer receives input from all the neurons in the previous layer. Dense layers are commonly used for mapping input features to output predictions in neural networks.

```
_________________________________________________________________
global_average_pooling2d (Gl (None, 1024)              0
_________________________________________________________________
reshape_1 (Reshape)          (None, 1, 1, 1024)        0
_________________________________________________________________
dropout (Dropout)            (None, 1, 1, 1024)        0
_________________________________________________________________
conv_preds (Conv2D)          (None, 1, 1, 1000)        1025000
_________________________________________________________________
act_softmax (Activation)     (None, 1, 1, 1000)        0
_________________________________________________________________
reshape_2 (Reshape)          (None, 1000)              0
=================================================================
Total params: 4,253,864
Trainable params: 4,231,976
Non-trainable params: 21,888
_________________________________________________________________
```

Figure 3.6c Training layers in MobileNet V2

3.7 Evaluation Metrics

The overall performance of the model was evaluated with several evaluation metrics: Accuracy, Weighted Average of Precision (WAP), Weighted Average of Recall (WAR), and Weighted Average of F1-score (WAF).

Accuracy

Accuracy measures the proportion of correctly classified instances out of the total number of instances. It is calculated as:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Accuracy gives an overall measure of how often the model makes correct predictions.

Precision

Precision measures the proportion of true positive predictions among all positive predictions made by the model. It is calculated as:

$$Precision = \frac{TP}{(TP + FP)}$$

Precision measures the model's ability to correctly identify positive instances, while minimizing false positives.

Recall (Sensitivity)

Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive instances that are correctly identified by the model. It is calculated as:

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

Recall measures the model's ability to identify all positive instances, minimizing false negatives.

F1-score

F1 Score is the harmonic mean of precision and recall, providing a balanced measure of both metrics. It is calculated as:

$$F1 - score = 2 \cdot \frac{Precision * Recall}{Precision + Recall}$$

F1 Score provides a balance between precision and recall, useful when there is an uneven class distribution or when both false positives and false negatives are important.

Top-2 and Top-3 accuracy were also used during evaluation. They are metrics used to assess the performance of classification models, particularly when dealing with multi-class classification problems. Top-2 accuracy measures the proportion of predictions where the correct class label is among the top two predicted classes. In other words, if the model predicts the correct class as the top choice or the second most likely choice, it is considered a correct prediction. This metric provides a more lenient evaluation compared to traditional accuracy, as it allows for some level of uncertainty in the prediction. Top-3 accuracy measures the proportion of predictions where the correct class label is among the top three predicted classes. This metric offers an even broader evaluation, allowing for more flexibility in the prediction. It considers a prediction correct if the correct class label is ranked within the top three predicted classes by the model.

These metrics are particularly useful when dealing with datasets where there may be some ambiguity or overlap between classes, or when the model's task is to provide a ranked list of potential class labels rather than a single prediction. They provide a more nuanced understanding of the model's performance beyond traditional accuracy.

The evaluation process included assessing performance metrics and employing cross-validation techniques to ensure the model’s robustness. By extending the models and evaluating their performance, the study ensured that the most accurate model for used for prediction.

Chapter 4

Results

The calculations were performed in a Google Colab environment. The CPU used was an Intel Xeon CPU with 2 vCPUs (virtual CPUs) and 13GB of RAM. The GPU used was a NVIDIA Tesla K80 with 12GB of VRAM (Video Random-Access Memory). Model Evaluation was performed by calculating categorical accuracy, top2 accuracy, top3 accuracy, classification report, and confusion matrix. Further, the loss and accuracy curves were plotted to validate the model's performance for the optimization and prediction phase.

4.1 Model Validation

The validation of the model was conducted on 938 unknown sample images from the validation set. I evaluated weighted average for precision, recall, and f1- score to evaluate the MobileNet model performance on unknown images of the validation set. The Weighted Average of 86%, 81%, 81% was recorded for Precision, Recall, and F1-score. The MobileNet model shows best precision, recall, and F1-score value for Melanocytic Nevi. The Multi-Class Classification Report showing weighted average for Precision, Recall, and F1-Score are represented in figure 4.1b.

```
# Here the the last epoch will be used.
val_loss, val_cat_acc, val_top_2_acc, val_top_3_acc = \
model.evaluate_generator(test_batches,
                         steps=len(df_val))

print('val_loss:', val_loss)
print('val_cat_acc:', val_cat_acc)
print('val_top_2_acc:', val_top_2_acc)
print('val_top_3_acc:', val_top_3_acc)
```

```
val_loss: 0.8585962670087299
val_cat_acc: 0.7547974413646056
val_top_2_acc: 0.8827292110874201
val_top_3_acc: 0.9424307036247335
```

Figure 4.1a Validation statistics for categorical accuracy, top2 accuracy and top3 accuracy.

	precision	recall	f1-score	support
akiec	0.50	0.42	0.46	26
bcc	0.41	0.87	0.55	30
bkl	0.77	0.13	0.23	75
df	0.10	0.50	0.17	6
mel	0.28	0.49	0.35	39
nv	0.94	0.91	0.92	751
vasc	0.67	0.73	0.70	11
avg / total	0.86	0.81	0.81	938

Figure 4.1b Multi-Class Classification Report showing weighted average for Precision, Recall and F1-Score

Categorical accuracy: 75.47%
top2 accuracy: 88.27%
top3 accuracy: 94.24%

The comparison of the current study with other related previous work is represented in ***table 4.1a***. The majority of previous work is done on two or three classes, and their accuracies and recall vary between approximately 66 percent to 81 percent and 60 percent to 76 percent, respectively. In a previous study, it was reported 48.9 percent and 55.4 percent classification accuracy evaluated for nine classes using CNN models.

Source	Year	Classifier	No. of classes	Accuracy %
[1]	2016	Multi-tract CNN	Ten	75.1
[2]	2017	CNN	Three	69.4
		CNN-PA		72.1
		CNN	Nine	48.9
		CNN-PA		55.4
[3]	2019	InceptionResnetV2	Seven	70.0
		PNASNet-5-Large		76.0
		SENet154		74.0
		InceptionV4		67.0
	2024	MobileNet	Seven	75.47% (cat) 88.27% (top2) 94.24% (top3)

Table 4.1a Comparison results of the current study with previous related work

In another study, classification accuracy for ten classes using Multi-track CNN was reported to be 75.1 percent. Another study reported accuracy as 70 percent, 76 percent, 74 percent, and 67 percent for seven classes using InceptionResnetV2, PNASNet-5-Large, SENet154, and InceptionV4, respectively. In this study, I achieved categorical accuracy of 75.47 percent, top2 accuracy of 88.27 percent and top3 accuracy of 94.24 percent using MobileNet. My seven-way skin cancer classification method has performed better than previously proposed computer-aided diagnosis systems in terms of both accuracies and recall. Additionally, the proposed method is more efficient considering the faster processing capability and lightweight architecture of MobileNet.

4.2 Confusion Matrix

The confusion matrix for our model was evaluated for seven classes. Each element of confusion matrix shows the comparison between the True label and Predicted label for each image in the validation set. Our model showed the best result for Melanocytic nevi by making a correct prediction for 696 images out of 751. Basal cell carcinoma and Melanoma were correctly determined for 26 images out of 30 and for 27 images out of 39, respectively. The diagnosis of Benign keratosis was most challenging due to their similar appearance with Melanoma and Melanocytic nevi. Only ten correct predictions were recorded for Benign keratosis.

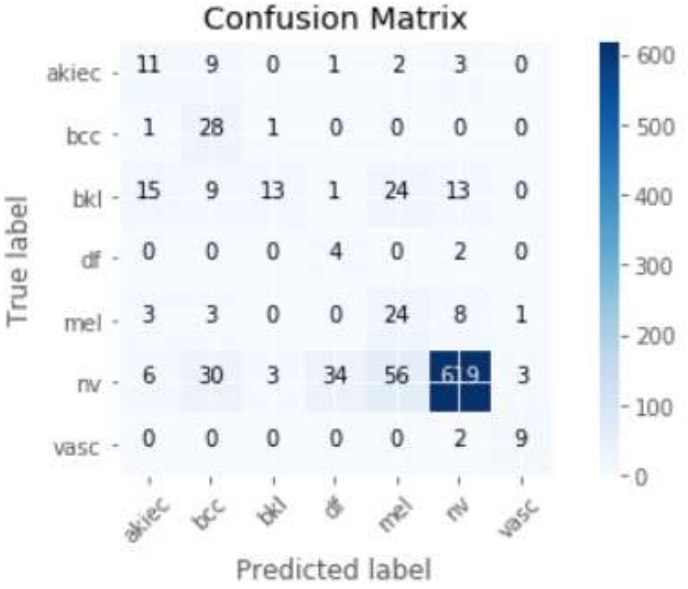
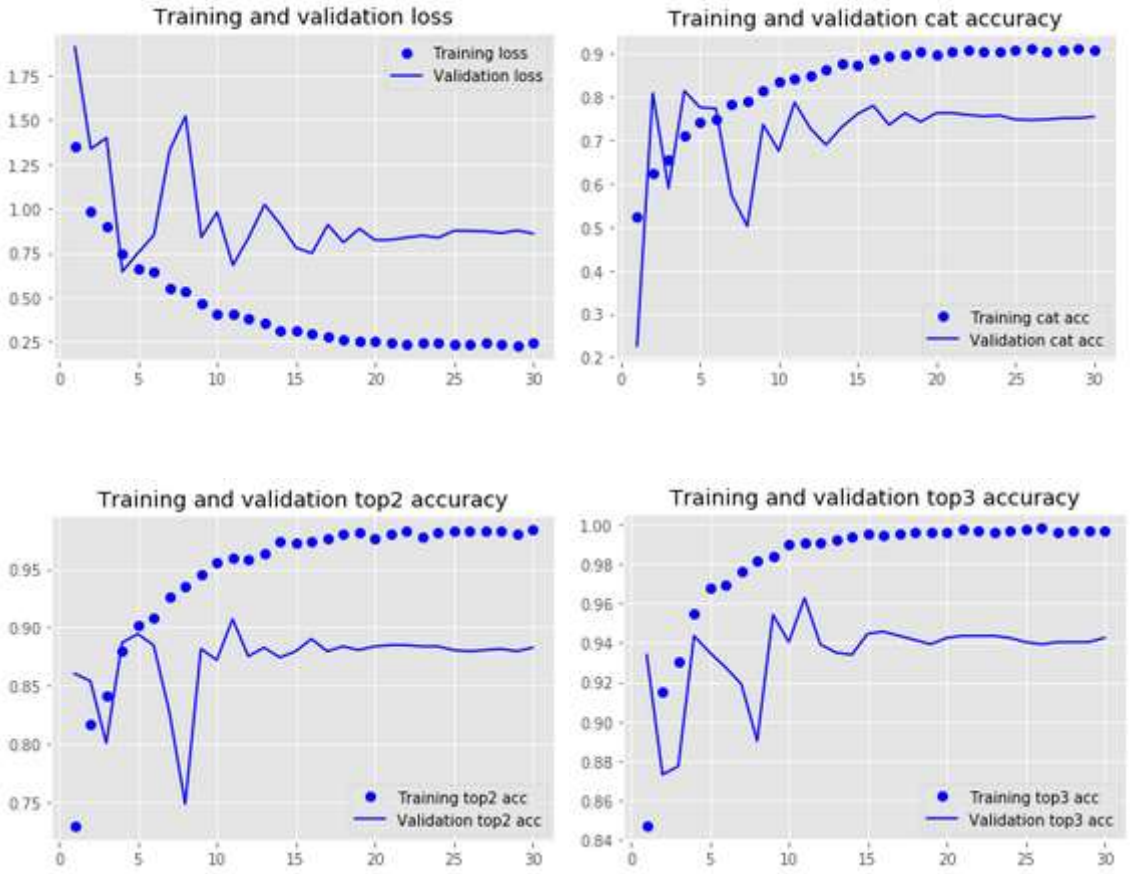


Figure 4.2 Confusion matrix for seven classes along with the true and predicted labels.

4.3 Loss and accuracy curves

In order to examine learning, generalizing and performance of the model, I computed training-validation loss curve, training-validation accuracy curves for categorical, top2 and top3 accuracies which are given below. The model shows a good learning rate as the training accuracy increase with the number of iterations along with symmetric downward sloping of the training loss curve. The small gap between training and validation curves represents a good-fit, showing model can generalize well on unknown images.



Skin cancer classification performance curves of MobileNet model
(a) Training and Validation loss
(b) Training and Validation categorical accuracy
(c) Training and validation top2 accuracy
(d) Training and Validation top3 accuracy

4.4 Live Web Application

I have developed a web application to provide an effective automated online tool for the multi-class classification of dermoscopy skin lesion images. This web application is available for use at https://agg-geek.github.io/skincancer/.

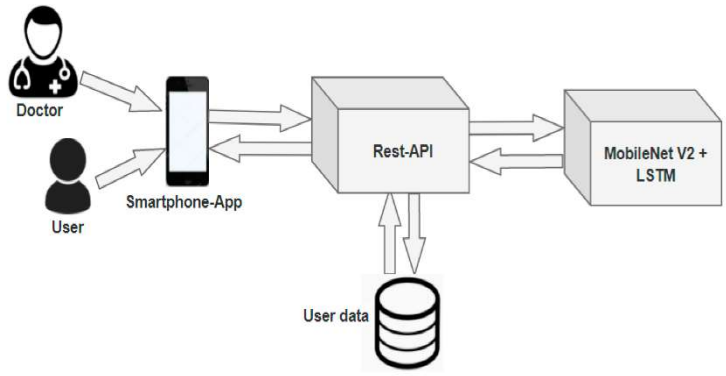


Figure 4.4a. The framework of the web app along with the MobileNet V2.

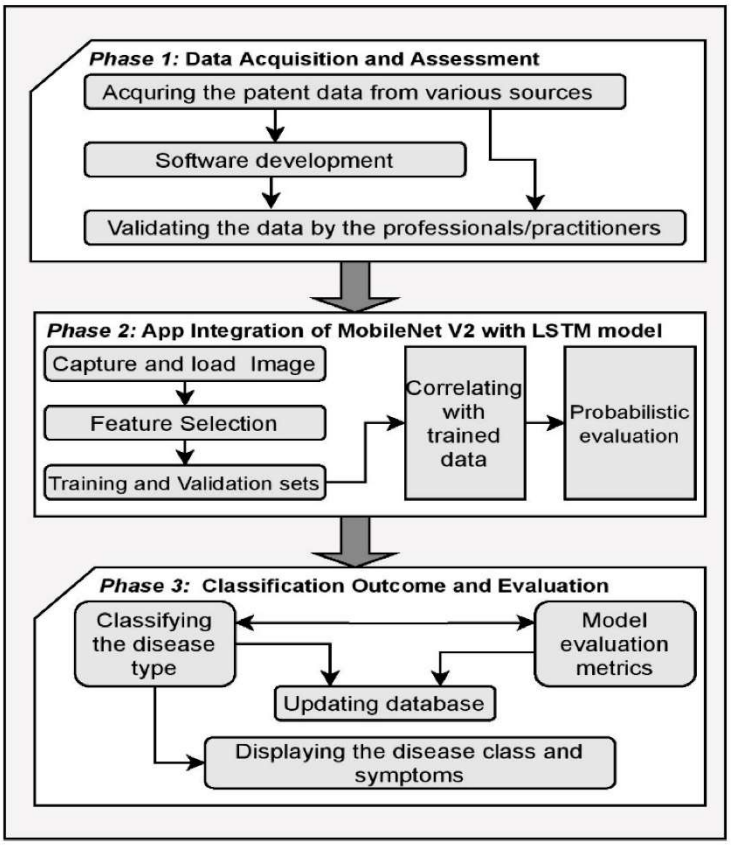


Figure 4.4b. Overall process used in the project: dataset to build model development followed by the model integration for website deployment.

MobileNet has been used for model development and a Keras model is used along with Tensorflow.js which powers the app backend.

The web application is developed using TensorFlow.js, which operates in the backend to download the model from the project repository upon initialization. Patients are instructed to take a clear photograph of the skin lesion and upload the image using the "Choose image" option on the web application. Upon detecting a new image upload, TensorFlow.js triggers an event handler that processes the image by invoking the `predict_model()` function. The top three diagnostic categories resulting from the prediction are subsequently displayed to the patient.

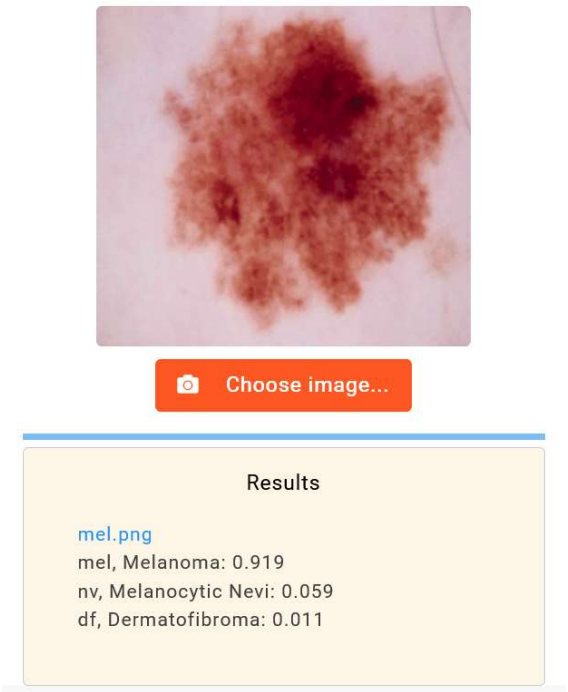


Figure 4.4c. The interface of the web application to upload image of skin lesion. Upon uploading, the results are presented in the form of the top three predicted diagnostic classes of skin lesions, each accompanied by the probability for its respective class.

The website is open-sourced and publicly available and offers a user-friendly interface where doctors / patients can input relevant lesion images. The website prioritizes user privacy by ensuring that the model is downloaded directly onto the user's device. Consequently, lesion images are not uploaded to any server; instead, predictions are performed locally using the downloaded model. Once the model is downloaded, it is stored in the device's cache, eliminating the need for subsequent downloads for future predictions. This approach not only conserves user data bandwidth but also enhances the speed of predictions, as the model is readily available on the device.

Chapter 5

Conclusion & Future Work

The skin cancer incidences are intensifying over the past decades; the need of an hour is to move towards an efficient and robust automated skin cancer classification system, which can provide highly accurate and speedy predictions.

In this study, I demonstrated the effectiveness of deep learning in automated dermoscopic multi-class skin cancer classification with the MobileNet model trained on a total of 38,569 dermoscopy images from HAM10000 dataset. I matched the performance of expert dermatologists across seven diagnostic tasks with an overall accuracy of 75.47% for seven classes in the dataset, whereas top2 and top3 accuracy of 88.27% and 94.24%, respectively. Also, the weighted average of precision, the weighted average of recall, and the weighted average of f1-score were found to be 86%, 81%, and 81%, respectively. I conclude that MobileNet model can be used to develop an efficient real-time computer-aided system for automated medical diagnosis systems. As compared to previously proposed models the MobileNet model has shown accurate and robust performance in addition to its faster and lightweight architecture.

The high accuracy of the model developed in this approach is of great significance in the context of early skin cancer diagnosis prediction. Identifying individuals at risk at an early stage allows for timely interventions and preventive measures, which can significantly improve patient outcomes and reduce the burden of skin cancer, including economical burden. The accuracy and robustness of the model make it a valuable tool for healthcare professionals in identifying high-risk individuals and implementing targeted interventions.

Future work in skin cancer diagnosis could involve incorporating personalized patient data such as age, skin colour and genetic information alongside current diagnostic methods. Integrating these additional features could enhance the development of personalized computer-aided systems for skin cancer diagnosis. The website can be upgraded to accommodate these enhancements. The dataset, being very unbalanced, could also be updated to include a balanced

representation of all skin cancer types, which can be used to further improve the accuracy of the model.

Moreover, the website could be further enhanced to automatically detect skin cancer from skin lesions and provide valuable educational resources and lifestyle recommendations to reduce the risk of skin cancer. Additionally, implementing a database to store and synchronize patient image data could be valuable. This database could store images of skin lesions for each patient, enabling the tracking of disease prognosis, which could provide valuable insights for medical professionals.

References

1. Kawahara J, Hamarneh G. Multi-resolution-Tract CNN with Hybrid Pretrained and Skin- Lesion Trained Layers. In: Wang L, Adeli E, Wang Q, Shi Y, Suk HI (eds) Machine Learning in Medical Imaging. MLMI 2016. Lecture Notes in Computer Science. 2016;10019.
2. Esteva A et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017; 542(7639):115–118.
3. M. A. A. Milton, "Automated Skin Lesion Classification Using Ensemble of Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge," Jan. 2019.
4. Philipp Tschandl et al, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions".
5. Parvathaneni Naga Srinivasu et al, "Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM"
6. Stewart BW, Wild C. International Agency for Research on Cancer, and World Health Organization. World cancer report 2014.
7. Cakir BO, Adamson P, Cingi C. Epidemiology and Economic Burden of Nonmelanoma Skin Cancer. Facial Plast. Surg. Clin. North Am. 2012;20(4):419–422.
8. Rogers HW, Weinstock MA, Feldman SR, Coldiron BM. Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the U.S. Population, 2012. JAMA Dermatol. 2015;151(10):1081-1086.
9. Stern RS. Prevalence of a history of skin cancer in 2007: results of an incidence based model. Arch Dermatol. 2010 Mar;146(3):279-82.
10. WHO. Skin cancers WHO; 2017
11. Almeida, M.A.M.; Santos, I.A.X. Classification Models for Skin Tumor Detection Using Texture Analysis in Medical Images. J. Imaging **2020**, 6, 51
12. Ki, V.; Rotstein, C. Bacterial skin and soft tissue infections in adults: A review of their epidemiology, pathogenesis, diagnosis, treatment and site of care. The Canadian journal of infectious diseases & medical microbiology. Can. J. Infect. Dis. Med. Microbiol. **2008**, 19, 173–184.
13. Cahan, A.; Cimino, J. A Learning Health Care System Using Computer-Aided Diagnosis. J. Med. Internet Res. **2017**, 19, PMCID:PMC5362695.
14. Sae-lim, W.; Wettayaprasit, W.; Aiyarak, P. Convolutional neural networks using MobileNet for skin lesion classification. In Proceedings of the 16th

International Joint Conference on Computer Science and Software Engineering, Chonburi, Thailand, 10–12 July 2019; pp. 242–247
15. Castillo, D.; Lakshminarayanan, V.; Rodríguez-Álvarez, M.J. MR Images, Brain Lesions, and Deep Learning. Appl. Sci. **2021**, 11, 1675
16. SivaSai, J.G.; Srinivasu, P.N.; Sindhuri, M.N.; Rohitha, K.; Deepika, S. An Automated Segmentation of Brain MR Image through Fuzzy Recurrent Neural Network. In Bio-Inspired Neurocomputing; Bhoi, A., Mallick, P., Liu, C.M., Balas, V., Eds.; Studies in Computational Intelligence; Springer: Singapore, 2021; Volume 903
17. Hafiz, A.M.; Bhat, G.M. A Survey of Deep Learning Techniques for Medical Diagnosis. In Information and Communication Technology for Sustainable Development; Advances in Intelligent Systems and Computing; Tuba, M., Akashe, S., Joshi, A., Eds.; Springer: Singapore, 2020; Volume 933
18. Civit-Masot, J.; Luna-Perejón, F.; Domínguez Morales, M.; Civit, A. Deep Learning System for COVID-19 Diagnosis Aid Using X-ray Pulmonary Images. Appl. Sci. **2020**, 10, 4640.
19. Yamanakkanavar, N.; Choi, J.Y.; Lee, B. MRI Segmentation and Classification of Human Brain Using Deep Learning for Diagnosis of Alzheimer's Disease: A Survey. Sensors **2020**, 20, 3243.