# Skin Lesion Identification Using Deep Learning

Hammad Mir - 200709330

1 Abstract

Skin cancer is defined as the uncontrolled growth of abnormal cells in the epidermis, the skin's outermost layer, caused by unrepaired DNA damage that results in mutations. Skin cancer is one of the most frequent types of cancer in the globe. The three most frequent types of skin cancer are squamous cell carcinoma, basal cell carcinoma, and melanoma. The clinical therapy of a skin lesion is mostly determined by its prompt discovery. The prevalence of skin cancer is on the rise, particularly melanoma, which is aggressive due to its high metastatic rate. As a result, early detection is crucial for therapy before the beginning of malignancy. To overcome this issue, medical imaging is employed for dermoscopic image processing and classification/identification. This gives an opportunity to create an automated model for identifying skin lesions. Using Convolutional Neural Networks, we investigated an automated technique for lesion diagnosis (CNN). We trained our model using Kaggle's "Skin Cancer MNIST: HAM10000" dataset, which contains a huge collection of multi-source dermoscopic pictures of pigmented lesions. We did data augmentation on the training data to boost generalization and classification performance. We employed the VGG16 and ResNet50 architectures for skin lesion picture classification and got the greatest accuracy and f1 score of 0.7651 and 0.7569 for ResNet50 with augmented data, respectively.

2 Introduction

A lesion is a section of skin that has an abnormal growth or appearance as compared to the surrounding skin. Skin lesions are classified into two types: primary and secondary. Primary lesions are abnormal skin conditions that can be present from birth or developed throughout the course of a person's life. Skin lesions that have been inflamed or modified cause secondary skin lesions. For example, if a mole is scratched until it bleeds, the subsequent lesion, a crust, is now a secondary skin lesion. Skin cancer is defined as the uncontrolled proliferation of aberrant cells in the epidermis, the skin's outermost layer, produced by un-repaired DNA damage that causes mutations. Skin cancers are one of the most common cases of cancer worldwide. Squamous cell carcinoma, basal cell carcinoma, and melanoma are the three most common kinds of skin cancer. Melanoma is far less frequent than the others, but it is far more likely to infiltrate adjacent tissue and spread to other regions of the body and is the leading cause of death from skin cancer. These mutations cause skin cells to grow quickly, resulting in the formation of malignant tumours. A skin lesion may or may not be cancerous and thus early detection is critical, as the estimated 5-year survival rate for skin cancer such as melanoma drops from over 99% if detected in its earliest stages to about 14% if detected in its later stages.

Type of skin cancer
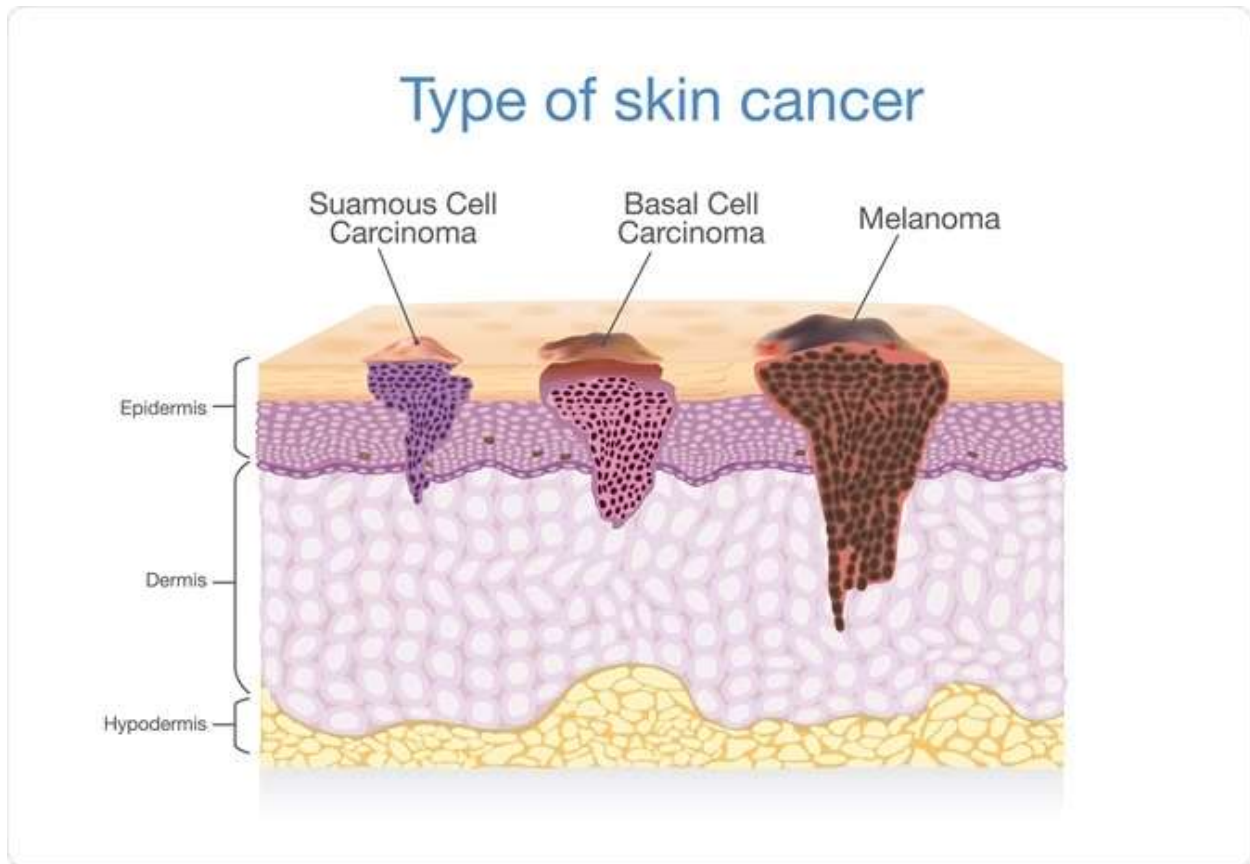
3 Diagnosis

Skin cancers are primarily diagnosed visually, beginning with an initial clinical screening and followed potentially by dermoscopic analysis, a biopsy and histopathological examination. Dermoscopy is the technique used to observe inner layers of skin by using microscope and special illumination. It allows detailed examination of skin lesions and provides inner view of many dermoscopic structures. It also helps dermatologist in diagnosis of malignant lesions. Histopathology refers to the microscopic examination of tissue in order to study the manifestations of disease. Specifically, in clinical medicine, histopathology refers to the examination of a biopsy or surgical specimen by a pathologist, after the specimen has been processed. Traditionally a specialist examines these images (Dermoscopic/Histopathological). However, these judgements depend on their personal experience and expertise and often lead to considerable variability. Thus, it becomes important to develop computational tools and learning techniques for automated diagnosis that operate on quantitative measures, thus helping the dermatologists diagnose skin cancers more efficiently.

4 Motivation

The research community has paid close attention to computer-aided technology for the diagnostic interpretation of medical images. Object Recognition has been a key area in computer vision and Deep learning techniques like Convolutional Neural Networks (CNN's) offer high performance for object recognition. Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in visual object recognition, object

detection and many other domains. Deep convolutional networks have brought about breakthroughs in processing images, video, speech and audio. With early detection of cancers being paramount in the treatment and survival of patients, it is imperative to develop systems that can identify potential cancers with relatively high accuracy. As such we will be exploring CNN on the HAM 1000 Dataset.

5   Previous Work

A lot number of research in the field of melanoma detection have been made over the last decade, which covers a wide range of computer vision and pattern recognition techniques. Segmentation and classification of images are the most used techniques, found in the literature. Segmentation techniques include assigning of threshold binary values to effected region and non-effected region. The proposed technique provides good results away from the lesion boundary, but accuracy falls near boundary where skin colour overlaps. Nikhil Cheerla et al. proposed an algorithm using texture and local binary pattern for segmentation of melanoma effected images. After application of segmentation technique, images were classified which returns a sensitivity of 97%. The paper addressed the classification scheme for both melanoma and non-melanoma images. For this a step-by-step approach was applied including K-mean clustering for segmentation, feature extraction from different domains and finally classifying Melanoma, Basal cell carcinoma (BCC), Seborrhoeic keratosis (SK) and Nevus using support vector machine (SVM) classifier. An accuracy of 96.8% was recorded in their work. Almaraz et al. proposed a method of hybrid feature extraction using ABCD rule (A for Asymmetry, B for the Border, C for colour and D of the Diameter of effected skin area) and textural features. After getting hybrid feature set a classification scheme using support vector machine was applied and a performance rate of 75.1% was achieved.

Another neural network technique presented by Alencar et al. used color feature of ABCD rule and edge features for development of feature set that is further classified using MLP trained network. The proposed system achieved a specificity of 93%. ABCD rule features played a vital role in a number of classification schemes proposed by many authors. Majorly used properties include shape, colour, edge and texture feature extraction. A multi-layer feed forward neural network technique for Dermoscopic images' segmentation was proposed by Mehta et al. these segmented images were then presented for feature extraction using Fuzzy C-Mean Clustering and targets less error for segmentation as compared to ELM segmentation procedure. Barata et al. in their research investigated different algorithms to deal with robustness in Dermoscopic images. Their research showed that colour constancy helped improving classification results and gained a sensitivity of 79.7%. Celebi et al. presents an approach to deal with colour schemes in Dermoscopic images by reducing the range of colours in images to a specific small number. For this they used K-means clustering algorithm. These reduced colour images are then classified to achieve a sensitivity of 62%. It is observed that most of the work found in the literature is based on computer vision and segmentation techniques. Methods using pattern recognition algorithms used different kind of features for the detection of skins lesions due to melanoma. Unlike already proposed methods, the following method focuses on extracting novel set of simple statistical color and texture features along with the use of SVM classifier to classify melanomas images from all dermoscopic images. The above work was performed on the PH2 dataset containing 200 dermoscopic images including 80 common nevi, 80 atypical nevi and 40 melanomas, making up 160 non-melanomas and 40 melanomas.

Hartanto et al. developed an Android app that detects skin cancer using MobileNet v2 and Faster R-CNN algorithms. Both proposed designs were taught to recognise images of actinic keratosis and melanoma skin cancer targets. The dataset used consisted of 600 images divided into two categories: actinic keratosis images and melanoma images, with no regard for gender, age, or other variables. They obtained classification accuracy of 87 percent for Faster R-CNN and 86 percent for MobileNet V2, respectively. Zafar et al. introduced Res-Unet, an automated approach for segmenting lesion borders that incorporates two designs, the U-Net and the ResNet. They also employed picture in-painting to remove hair, which dramatically enhanced the segmentation findings. The ISIC 2017 dataset was used for training and testing, as well as the PH2 dataset. On the ISIC 2017 test set, the proposed model achieved a Jaccard Index of 0.772, and on the PH2 dataset, it achieved a Jaccard Index of 0.854, which are comparable results to currently available state-of-the-art techniques.

6 Dataset

We focused on the HAM10000 dataset ("Human Against Machine with 10000 training Images"), which is publicly available on Kaggle and was originally released as part of the ISIC-2018 competition. The collection comprises of 10015 dermatoscopic images from various populations that were captured and archived using various modalities. Cases include a representative collection of all important diagnostic categories in the realm of pigmented lesions: actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc). Histopathology (histo) was employed to confirm more than 50% of lesions; the rest of the cases were confirmed relying on follow-up investigation (follow-up), expert consensus (consensus), or confirmation by in-vivo confocal microscopy (confocal). The dataset contains lesions with many pictures, which can be tracked using the HAM10000 metadata file's lesion-id column.
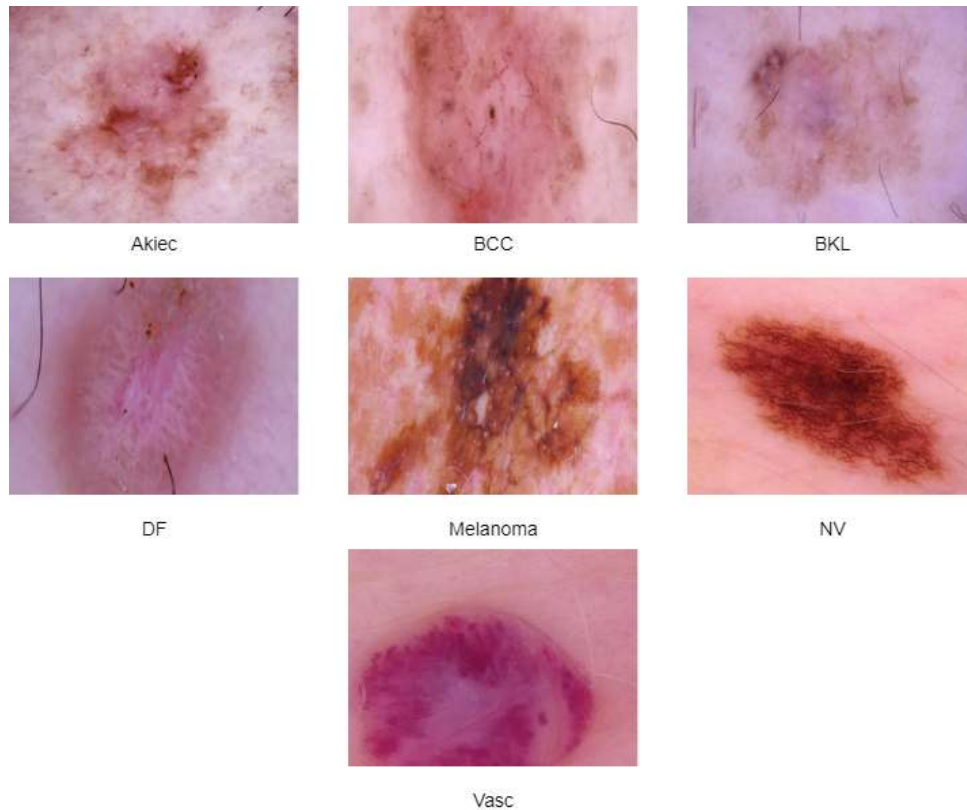
7 Work Done

7.1 Exploratory Data Analysis

The metadata file ("HAM10000_metadata.csv") contains details of all the images, such as lesion_id, image_id, lesion type (dx), labelling method (dx_type), age and gender of the patient and the location of the lesion. The structure of the metadata is as shown below.

```
##      lesion_id      image_id dx dx_type age  sex localization
## 1 HAM_0000118 ISIC_0027419 bkl   histo  80 male        scalp
## 2 HAM_0000118 ISIC_0025030 bkl   histo  80 male        scalp
## 3 HAM_0002730 ISIC_0026769 bkl   histo  80 male        scalp
## 4 HAM_0002730 ISIC_0025661 bkl   histo  80 male        scalp
## 5 HAM_0001466 ISIC_0031633 bkl   histo  75 male          ear
## 6 HAM_0001466 ISIC_0027850 bkl   histo  75 male          ear
```

We performed exploratory data analysis on the metadata to get a better understanding of the dataset. From the dataset we have the following lesion distribution of the overall dataset. Of all the lesion akiec, bcc and mel are the malignant types and the rest are benign.

- Lesion types and frequency

```
##
## akiec   bcc   bkl    df   mel    nv  vasc
##   327   514  1099   115  1113  6705   142
```

Following this, we retrieved summary of the age of patients for all lesions, benign lesions and malignant lesions. These are as given below:

- All lesions age summary

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   40.00   50.00   51.86   65.00   85.00      57
```

- Malignant lesions age summary

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    5.00   55.00   65.00   63.28   75.00   85.00       2
```

- Benign lesions age summary

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   40.00   50.00   49.08   60.00   85.00      55
```

We also created several graphics to help us better comprehend the dataset. According to the dataset, while most of the samples obtained belong to a male population (54%), samples collected from persons under the age of 50 years have more females and samples gathered from people over the age of 50 years have more men. Secondly, we have a higher number of images belonging to benign classes, with 67% of images classifying as NV showing that the dataset is heavily imbalanced. Finally, we have also plotted the Gender-wise Lesion localization, gender-wise lesion distribution and lesion-wise (Malignant and benign) age distribution plots to bet a better visual understanding of the data. From the plots we observe that most of the malignant lesions occur in later years, confirming to the malignant lesion summary above, showing the first quartile at 55 years of age.
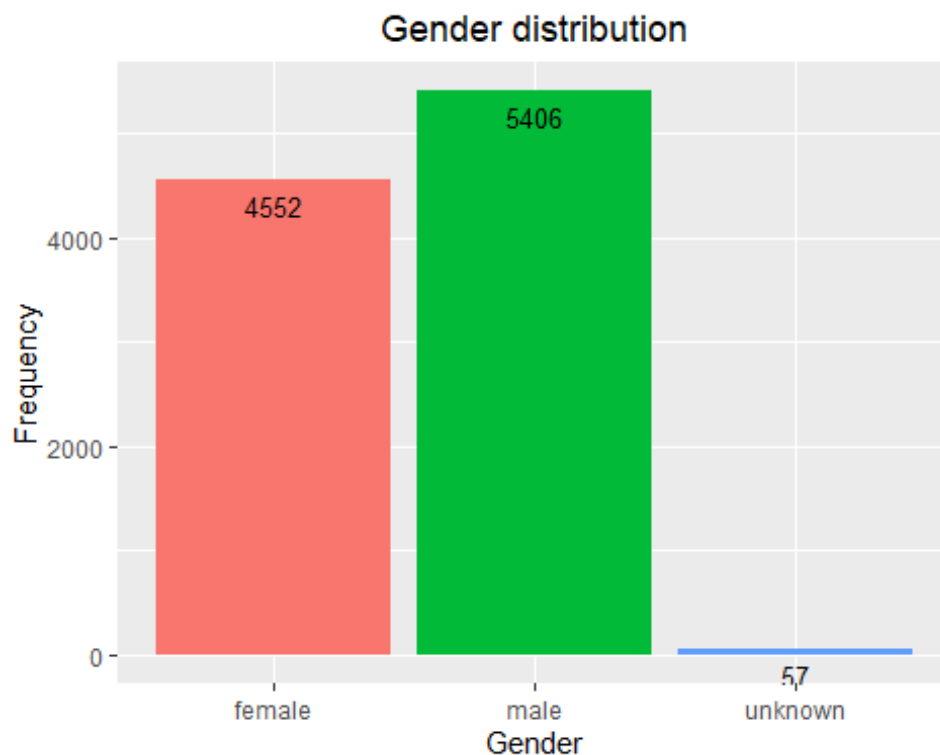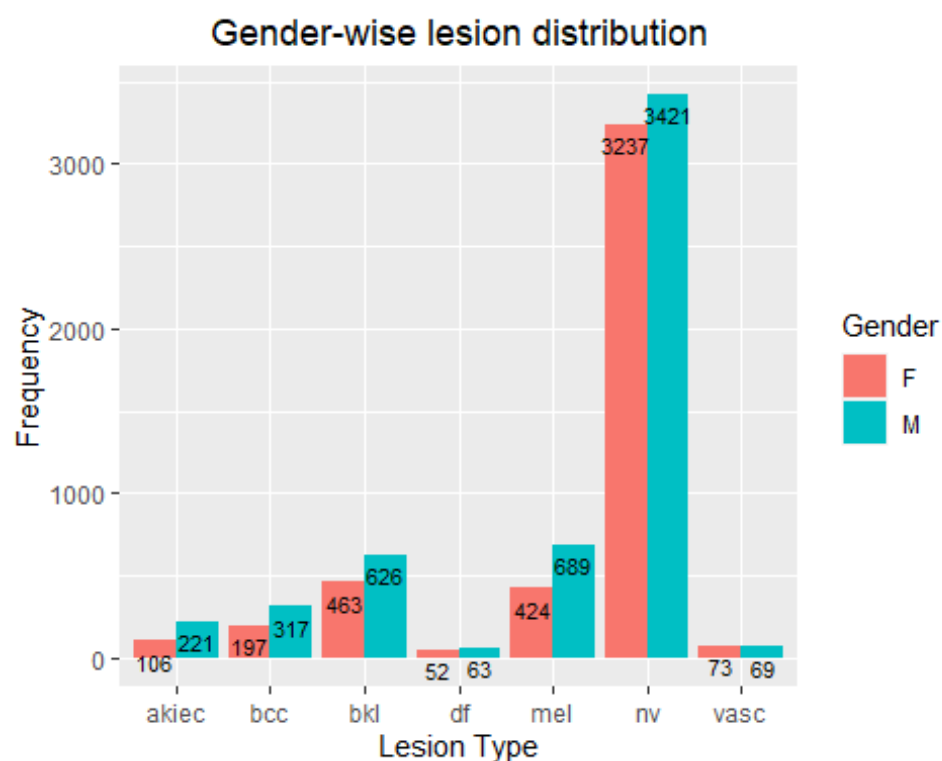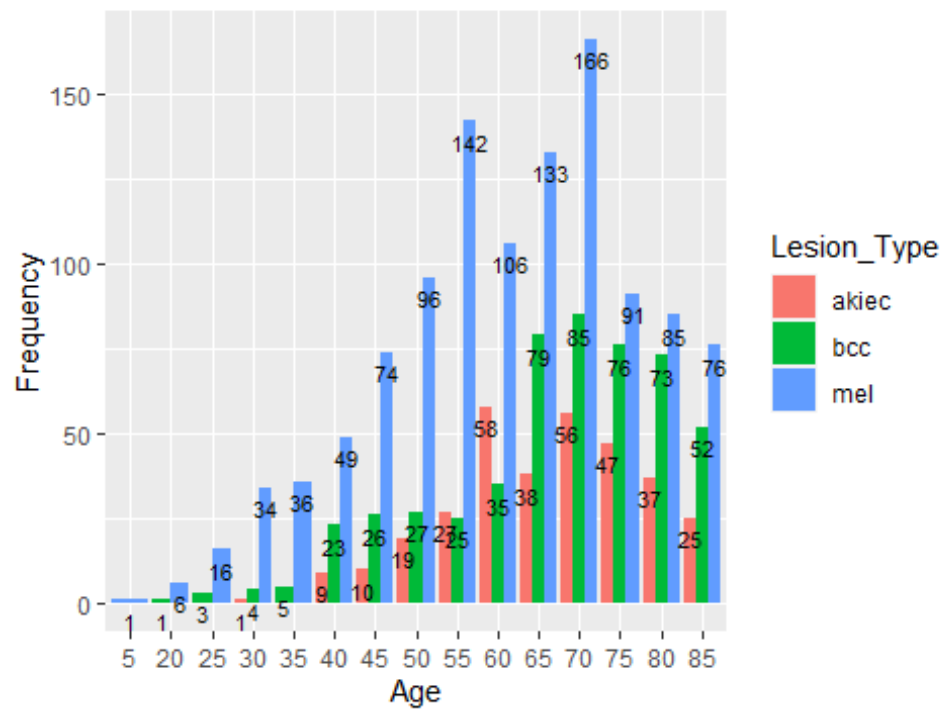
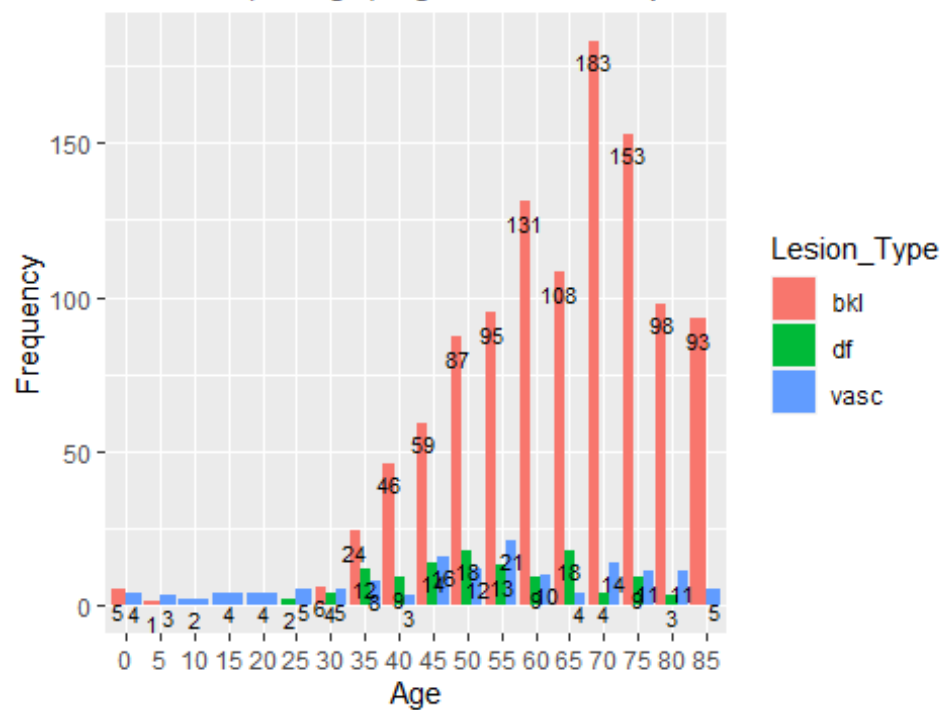# Gender-wise age distribution
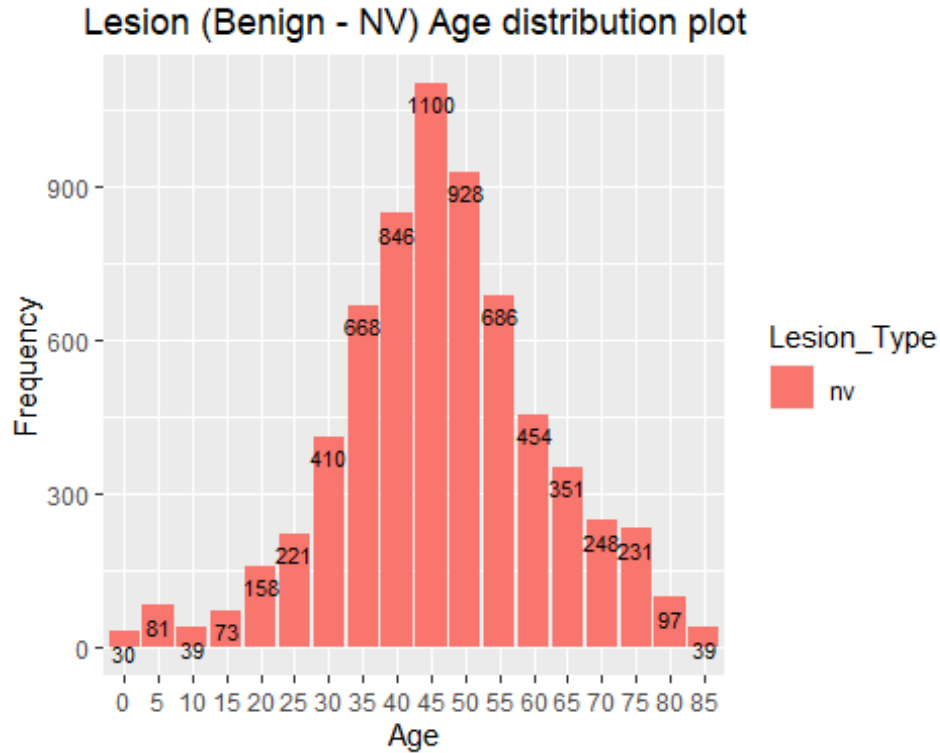

Leision type distribution

Gender-wise Lesion localization
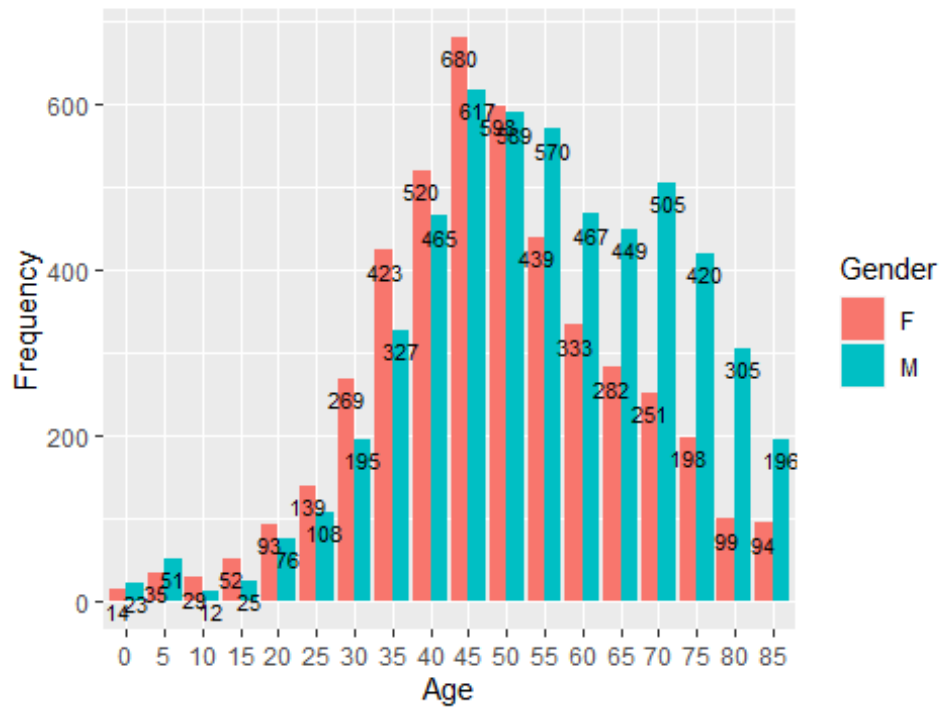
Gender-wise lesion distribution
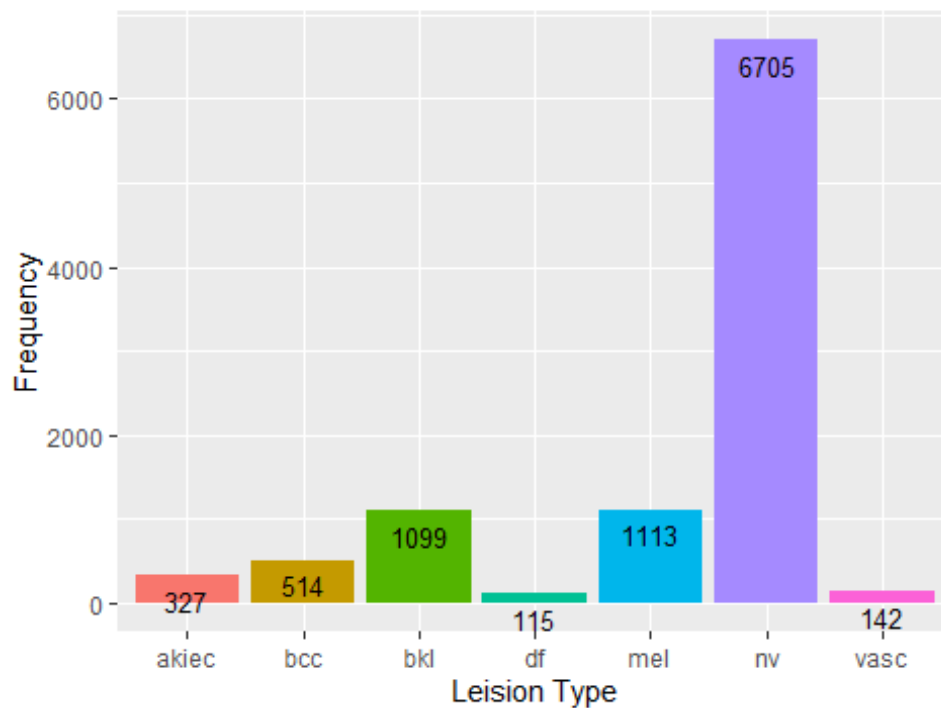
Lesion (Malignant) Age distribution plot

Lesion (Benign) Age distribution plot

Lesion (Benign - NV) Age distribution plot

7.2 Data Pre-Processing

For our project, we initially separated the dataset into 7 different folders corresponding to their class using the metadata file. The dataset was then split into training and testing data to a ratio of 8:2. To get a good split, it is important to pick images randomly without any bias. Instead of doing this manually we built a custom python script which takes as input the dataset path, the output path, split ratio and optionally a random seed value. The script builds the train and test image datasets for all image categories given a split ratio randomly by first shuffling all the images randomly then splitting them into two parts. To increase the amount of training data for better classification and to avoid overfitting, we performed augmentation on our training dataset by applying random rotations, flips and zoom operation to the images This was done by using the augmentor python library using which we increased the number of training images from 8,010 to 20,000.

The pre-processed images are available on google drive (here).
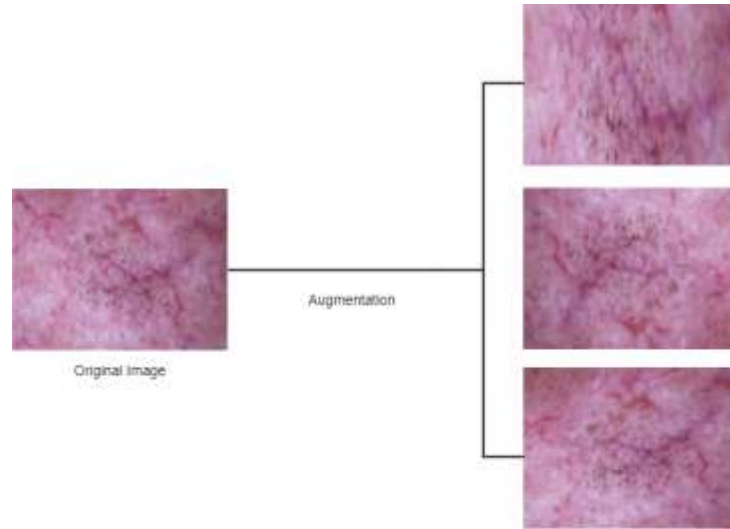
Data Augmentation

7.3   Classification Model

Keeping in line with the previous work on skin cancers and owing to the prowess of CNNs for image classification tasks, we decided to explore the application of available popular CNN models/architectures for this project. We explored various state of the art architectures and keeping in mind the time and resource constraints we decided to explore the performance of VGG16 and ResNet50 architectures for the skin lesion classification task. The models were used with their default configuration, an input size of 224x224x3 and a 'softmax' output with 7 nodes, giving probability of a image belonging to each of the given seven classes.

7.3.1   VGG16
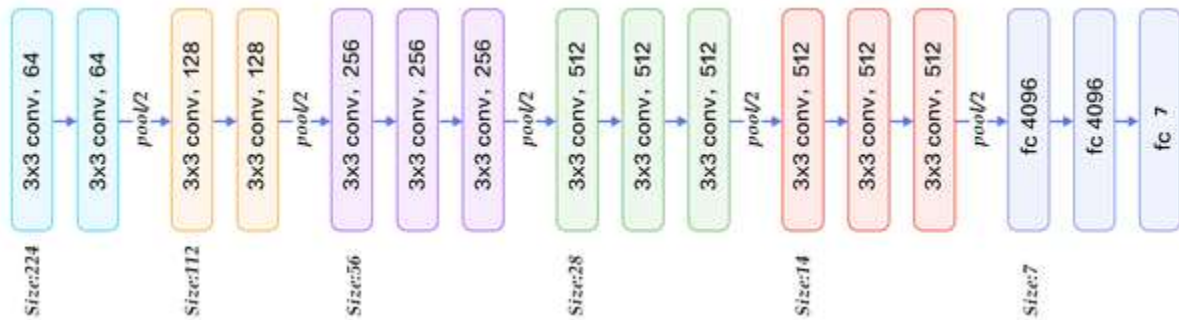
7.3.2 ResNet50

7.4 Network Training

The models were trained on their default configuration as described above on Google Colab using TensorFlow Keras python library. During training, we found the best results by using the Nadam optimizer with a learning rate of 0.0002 and categorical cross-entropy loss. The models were trained for 50 to 100 epochs and the model with best validation accuracy were saved using Keras call-backs. To account for time and resource constraints, the data was loaded while training using Keras sequence class and the images were resized from their original size of 450x600x3 to 224x224x3. The links to the saved models are given below:

- VGG16 with non-augmented data
- VGG16 with augmented data
- ResNet50 with non-augmented data
- ResNet50 with augmented data

7.5 Evaluation

Since the dataset is highly imbalanced, during training the model with the best accuracy was saved and the final evaluation was based on the weighted f1-score. The metrics are calculated as follows:

$$Accuracy = \frac{No. \, of \, correct \, prediction}{Total \, predictions}$$

$$Precision = \frac{True \, Positives}{True \, Positives + False \, Positives}$$

$$Recall = \frac{True \, Positives}{True \, Positives + false \, Negatives}$$

$$F1 \, Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

8 Results

During testing, we found that ResNet50 had the best performance on augmented data with an accuracy score of 0.7651 and an f1 score of 0.7569.

Network	Accuracy	F1 Score
VGG16 (non-augmented)	0.7566	0.7456
VGG16 (augmented)	0.7625	0.7494
ResNet50 (non-augmented)	0.7516	0.7337
ResNet50 (augmented)	**0.7651**	**0.7569**

9 Future Work

In the future, we intend to work towards developing an improved classification system for skin lesion detection and explore the possibility of deploying the system using normal images rather dermoscopic images for lesion identification.