An Efficient Deep Learning Approach to Detect Skin Cancer

by

Ashfaqul Islam
20341030
Daiyan Khan
19141024
Rakeen Ashraf Chowdhury
16141014

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
September 2021

© 2021. Brac University
All rights reserved.

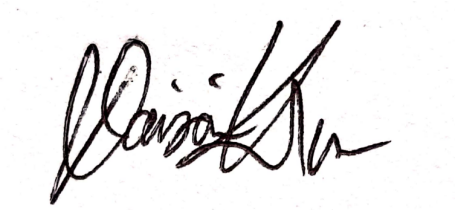
Declaration

It is hereby declared that

1. The thesis submitted is my/our original work while completing our degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material that has been accepted or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Students' Full Name & Signature:

Ashfaqul Islam
20341030



Daiyan Khan
19141024

Rakeen Ashraf Chowdhury
16141014

Approval

The thesis/project titled "An Efficient Deep Learning Approach to Detect Skin Cancer" submitted by

1. Ashfaqul Islam(20341030)
2. Daiyan Khan(19141024)
3. Rakeen Ashraf Chowdhury(16141014)

As of Summer 2021, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on September 26, 2021.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Ashraful Alam, Ph.D.
Assistant Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Md. Tanzim Reza
Lecturer
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi, Ph.D.
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Each year, millions of people around the world are affected by cancer. Research shows that the early and accurate diagnosis of cancerous growths can have a major effect on improving mortality rates from cancer. As human diagnosis is prone to error, a deep-learning based computerized diagnostic system should be considered. In our research, we tackled the issues caused by difficulties in diagnosing skin cancer and distinguishing between different types of skin growths, especially without the use of advanced medical equipment and a high level of medical expertise of the diagnosticians. To do so, we have implemented a system that will use a deep-learning approach to be able to detect skin cancer from digital images. This paper discusses the identification of cancer from 7 different types of skin lesions from images using CNN with Keras Sequential API. We have used the publicly available HAM10000 dataset, obtained from the Harvard Dataverse. This dataset contains 10,015 labeled images of skin growths. We applied multiple data pre-processing methods after reading the data and before training our model. For accuracy checks and as a means of comparison we have pre-trained data, using ResNet50, DenseNet121, and VGG11, some well-known transfer learning models. This helps identify better methods of machine-learning application in the field of skin growth classification for skin cancer detection. Our model achieved an accuracy of over 97% in the proper identification of the type of skin growth.

Keywords: cancer detection; convolutional neural networks; image classification; deep learning; machine learning algorithms

Acknowledgement

Firstly, all praise to Allah thanks to whom, this thesis has been completed without any setbacks.
Secondly, to our co-advisor Mr. Tanzim Reza sir for his kind support and advice with our work. He helped us whenever we required it.
And finally to our parents, without the support of whom this paper may not have been possible. With their kindness and prayers, we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

akiec Actinic Keratoses and Intraepithelial Carcinoma

ANN Artificial Neural Network

API Application Programming Interface

bcc Basal Cell Carcinoma

bkl Benign Keratosis

BNN Back-propagation Neural Network

CNN Convolutional Neural Network

df Dermatofibroma

DPI Dots Per Inch

HAM Humans Against Machine

KL Karhunen-Loève

mel Melanoma

nv Melanocytic Nevi

ReLU Rectified Linear Unit

RGB Red, Green, Blue

UV Ultraviolet

vasc Vascular Skin Lesions

VGG Visual Geometry Group

Chapter 1

Introduction

Each year, millions of people around the world are affected by cancer. Research shows that early and accurate diagnosis can have a major effect on improving mortality rates from cancer. Diagnosis sometimes requires a high-level of expertise and is often diagnosed completely wrong or as false-negative. Human diagnosis is prone to error, a deep-learning based computerized diagnostic system should be considered. We propose a system that will use a deep-learning approach to classify different types of skin lesions from digital scans and photographs. We will use image preprocessing to clip and crop the input data (digital images) to our required size and also to de-noise hazy images. Our model will transform and extract certain set features and create a connection between stimulus and the appropriate responses. The system will work more efficiently than the existing methods and will be working to do so with more accuracy.

Machine Learning has a subclass called Deep Learning. All of these algorithms take input data and extract information from them to classify the dataset and give predictions often along with the accuracy of the predictions. Deep learning algorithms excel at predicting results using layers and nodes especially when supervised learning is used. Supervised learning means teaching the deep learning algorithm by dividing the dataset into training and testing data.

There are about 2 to 3 million cases of skin cancer occurring globally each year with 1 out of every 3 growths being diagnosed as skin cancer. The report also states that the occurrence of cancer is increasing with each passing day [1].

Cancer treatment success depends on quick and proper diagnoses and early start of treatment. Non-melanoma cancer patients will often have a mortality rate of less than 5 percent. According to [2], the early diagnoses and treatment also apply to melanoma but since the nature of melanoma is more severe the mortality rates will be much higher.

The required number of oncologists will increase by about 40 percent whereas the number of actual oncologists will increase by only about 25 percent. The study also states that there needs to be an increase in the productivity of cancer specialists to cope with the growing demands, otherwise, the lack of doctors will start having

diverse effects on the growing number of patients [3].

Although dermatologists have hands-on experience, traditional skin cancer diagnosis methods require time and might produce erroneous results. A deep learning convolutional neural network-based system, that can detect skin cancer and the type from digital image inputs, can be developed and trained with a large and diverse dataset of images of non-cancerous and cancerous skin deformities. This will help oncologists work more efficiently and accurately and may help avoid several cases where complications might arise from false or inaccurate results.

Before images can be fed to an algorithm, the images need to be processed so that a dataset can be created where all the images are of the same size, resolution, etc. For use in skin cancer detection, the study mentions three stages of pre-processing images must be considered: image enhancement, image restoration and hair removal. Image enhancement helps improve the appearance of the images, focusing on the scaling and size of images, as well as the color and contrast of the images. The color space of all the images in the dataset must be made to be the same/similar to make it easier for the algorithm to have a standard to compare with. The contrast of images can be fixed using linear or non-linear contrast enhancement methods. (Image Scaling, Color Space Transformation, Contrast Enhancement). Image restoration helps to restore blurred or noisy images from corrupted sources. Spatial Filtering and Transform Domain Filtering are methods used by the study to aid with restoration, (Restoration from noise and blur) [4].

Images of skin cancer usually have fine hairs present in the picture as well as noise. This would lead to reduced accuracy due to foreign objects other than the skin cancer being present in the picture unless it is processed. Image pre-processing techniques were used by this study to mitigate this issue. The study mainly used Karhunen-Loève (KL) transform histogram equalization method, alongside contrast enhancement. The study also mentions the importance of maintaining a standard size for all images fed into the algorithm, by resizing the image to a fixed pixel width with a variable height. For de-noising they used a wavelet de-noise method (by two-dimensional bior3.3 wavelet). As this method leads to a "blur" effect while retaining image detail, a Median filter was used to remove the remaining noise in the images caused by the fine hairs [5].

For classifying the data, the U-net algorithm of CNN was used. Local Binary Pattern (LBP), Edge Histogram (EH), Histogram of Oriented Gradients (HOG) and Gabor method were utilized to attain features. These were passed into the Support Vector Machine (SVM), and also Random Forest (RF), K-Nearest Neighbor (KNN) and Naïve Bayes (NB) classifiers to predict if the image provided were of lesion that was benign or malignant [6].

The paper, "Max-min convolutional neural networks for image classification" modifies the standard convolutional block of CNN. Two well-known data repositories, "MNIST" and "CIFAR-10", are being used for the experiments. The results show that the convolutional net outperforms the standard CNN. Furthermore, a lot of deep networks could benefit from this Max-Min strategy with just slight modifications [7].

From the Blot et al., we can observe that a pooling layer is used to reduce the convoluted output's dimensions. Here, results are obtained using all three categories of pooling, i.e., attentive-pooling, max-pooling, and average-pooling. The results show that the better performer in terms of F1 score is max-pooling, which has a score of 64.56% whereas attentive pooling shows 59.92%. Lastly, the average-pooling is at 58.35%. Therefore, we know that max-pooling outperforms the other alternatives [8].

A deep CNN model with more than 50 layers was used to retrieve distinguishing features from the segments. With a residual learning implementation to avoid over-fitting, their implementation achieved the best and second-best results for classification and segmentation respectively [9].

In [10], they used "Google's Inception v3 CNN" architecture by removing the last layer and trained the system with their data. Back-propagation was used for the training, and for every 30 epochs, all layers were adjusted with the same global learning rate of 0.001 and a decay factor of 16..

In [11], Backpropagation Neural Network (BNN) for classifying their datasets under a training/testing method. After class labeling is used to categorize the images in their dataset, the BNN was used to assign weight and bias values to the images – starting with random values, the BNN assigned a set value based on the categorizing of the images (which was ascertained from the testing step), with training being continued until a minimum gradient was reached. Using this gradient, coefficients were calculated in the training step. The BNN was simulated during the testing step, categorizing the images into pictures of skin with melanoma, and without. Images with cancer present were assigned a value of 1, and 0 for those without. If simulation of the BNN yielded a value below 0.5, the image was considered cancer free. Accuracy was calculated after the training/testing paradigm had been completed.

Convolutional networks are the cornerstone of most state-of-the-art computer vision systems for a wide range of problems. Deep convolutional networks have become popularised since 2014, with significant advances in a variety of standards. For a number of use cases, including mobile vision and big-data scenarios, computational efficiency and low parameter count are still enabling aspects, even though for most operations, greater model size and computing expense may not always correlate to rapid quality improvements (as long as enough labeled data is provided for the training).

Chapter 2

Related Work

In [12], focused on diagnosing malignant melanoma from digital images, a type of skin disease that affects skin pigmentation by forming malignant tumors. This type of skin cancer is difficult to categorize unless done by a trained dermatologist, therefore the goal of the study was to develop an Artificial Neural Network (ANN) that could differentiate melanoma from other benign tumors.

Having separated the types of malignancies caused by melanoma, the study set about to use ANNs as Pattern Classifiers. As this method required them to classify the digital images into separate fixed categories, they chose to use a back-propagation Training Algorithm (Hyper-plane Classifiers) – i.e. "nodes in this network typically form a weighted sum of the inputs and pass this sum through a sigmoidal non-linearity". The study mitigated actual and desired outputs of the ANN with the help of the back-propagation under supervised training. The supervised training was conducted under the training/testing paradigm.

Implementation of the ANN was carried out on an average of 50 minutes at a time, providing results for the training/testing at 40/60 and 60/40 percentages (%).

Experimentation after implementation providing diagnostic results which yielded up to 86% accuracy in detecting malignant melanoma, while using reasonably sized data sets. Color characteristics and Tumor asymmetry were experimentally proven to be key factors in the diagnostic process.

In [13], we see that there are three key induction procedures. These are Heuristics, Formal Induction Methods, and Neural Networks. We will use a Convolutional Neural Network (CNN) to identify cancerous growths on the skin. The results from the training sets of this research paper show that the data in the input is ambiguous. In the second batch of training data, using more inputs and higher-level data, the outcomes are more unexpected, but better than without the higher-level data. This shows the importance of color information through computer vision. Variegated coloration was successfully identified using the "SCT/Center 2-D Split Color Image Segmentation" method. When the automatic induction method is used with the color segmentation technique, this feature can be identified at a rate of up to 92 percent. The feature vector techniques were about as successful as the expert

heuristics, with a success rate of 65%.

In recent years, digital dermoscopy has become widely used as a critical tool for assisting in the clinical identification of pigmented skin lesions. With the goal of comparing the validity of digital dermoscopy with an experienced dermatologist (with 5 years of experience) to doctors with the bare minimum of training and evaluating them using data from computer diagnostics. A total of three hundred and forty-one lesions were histopathologically examined and surgically excised, the sample consisted of both pigmented melanocytic and non-melanocytic skin lesions. A well-trained artificial neural network was used to frame and analyze digital dermoscopic images and calculated Cohen's k statistic to gauge the validity with true positive diagnoses of melanoma and non-melanoma. In-depth analysis and comparison led to the conclusion that experienced dermatologists placed higher in both sensitivity and specificity of the tests closely followed by inexperienced clinicians while computer analysis gave the highest number of false positives. Evidently, further development and usage of a well-trained artificial neural network can drastically improve the accuracy of melanoma diagnoses where an expert dermatologist might not be present [14].

Chapter 3

Different Types of Skin Cancer

Skin cancer is the abnormal proliferation of cells in the skin, that develops mostly on or in areas of the skin that have significant exposure to sunlight, leading to the damage of DNA, and as a result, mutations are triggered. Skin cells grow uncontrollably as a result of these defects, resulting in malignant tumors.
There are two main categories of skin cancer – melanoma and non-melanoma. These are the most common types of cancer out of all the possible types to occur in humans, with melanoma being the 19th most commonly occurring worldwide, and non-melanoma being the 5th most commonly occurring, according to World Cancer Research Fund. These values are likely an underestimation.

The 7 specific types of skin lesions and growths our paper will be covering include:

- Melanocytic nevi
- Melanoma
- Benign keratosis-like lesions
- Basal cell carcinoma
- Actinic keratoses and intra-epithelial carcinoma (aka Bowen's disease)
- Vascular lesions
- Dermatofibroma

We shall look into the types of chosen skin lesions and growths in further detail below.

Melanocytic nevi

A melanocytic nevus, otherwise known as a mole, is a benign lesion of the skin which comes from the excess growth of pigmented cells, or melanocytes. The collation of the melanocytes is what creates the dark moles that can be seen in men and women of all ages -which can either exist from birth or manifest at a later age in life.

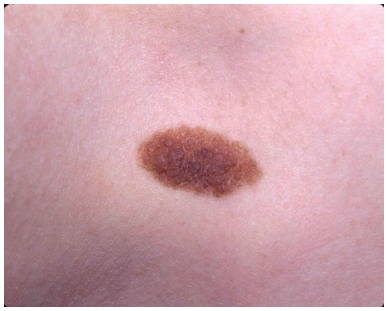


Figure 3.1: Melanocytic nevi, aka common mole

Commonly occurring amongst all individuals, it is usually found that fair-skinned people are more likely to have melanocytic nevi than people with darker skin. These may occur anywhere on the body (usually on the limbs), form with varied overlaying textures (rough, raised, flat, or thickened) and their color may vary from anywhere between flesh tones to black, at times making them difficult to identify through images. Some nevi accompany excess hair growth on them.

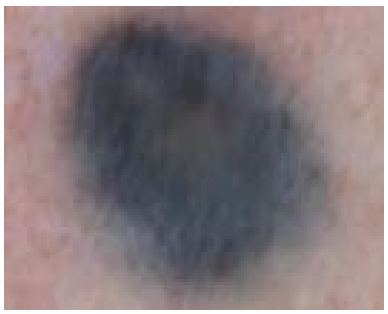


Figure 3.2: A dark-blue colored Melanocytic Nevi

The exact cause of melanocytic nevi manifestation is unknown, although research has shown that it is affected by genetic predisposition and level of exposure to the sun. With time, existing or new melanocytic nevi can change their shape and color, tending to enlarge and become unshapen. People with a higher number of nevi have been recorded to have a higher risk of developing melanoma than those with fewer nevi. 1 in 75 people is born with small melanocytic nevi, with larger nevi occurring once in every 20,000 births.

Melanocytic nevi are not to be confused with melanoma, a malignant form of melanocytes that commonly causes death from skin cancer. Nevi tend to have a symmetrical shape which can help distinguish them from melanoma.

Melanoma

Melanoma may not be the most commonly occurring form of skin cancer, but it is a typically dangerous form of skin cancer due to its tendency to grow and spread, from the mutation of melanocytes. Melanocytes are the cells in our skin and eyes that create and contain the pigment melanin, which attributes to the color of our physical selves. As with all types of skin cancer, melanoma is triggered when damage occurs to the DNA from UV radiation, resulting in the uncontrolled cellular growth of melanocytes.

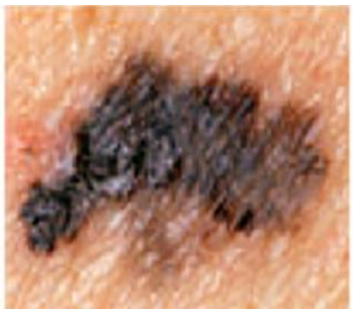


Figure 3.3: A Melanoma distinguishable by its asymmetry

The size, shape and color of melanomas vary, and can often be mistaken for melanocytic nevi thus making it difficult to ascertain a set of symptoms for melanoma at times. Some techniques that help identify the difference between melanomas and melanocytic nevi are symmetry, color, and evolution. Most melanomas tend to present themselves as asymmetrical, and sometimes pigmented with multiple shades of coloring. Melanomas also tend to evolve with time, changing their shape, texture, or even developing itching and in worse cases bleeding.

Melanoma is more than 20 times more likely to occur in fairer-skinned people than in darker-skinned people, with the overall risk of a fair-skinned person being at almost 3% across their lifetime. Though the percentage of risk increases with the increase of age, Melanoma still continues to be one of the most prevalent cancerous types of skin growth in young adults. There are many types of melanomas, some of which are outlined below:

Superficial Spreading Melanoma: Melanoma of this type is the most common type, which spreads for several years along with the outer layer of a person's skin. These are usually bumpy with irregular borders. It may grow along the skin for several years before invading layers of tissue below the epidermis.

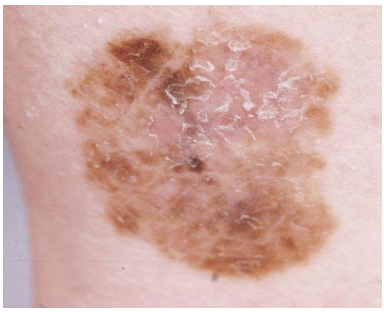


Figure 3.4: Superficial Spreading Melanoma

Nodular Melanoma: These count as 15% of all cases of melanoma, which presents as a bump or node that forms above the surface of the skin, and is firm to the touch. They typically look similar to melanocytic nevi, or moles.

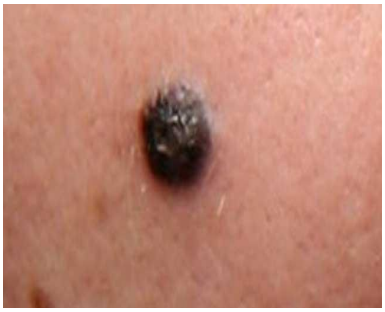


Figure 3.5: Nodular Melanoma

Lentigo Maligna Melanoma: This is a rare type of melanoma, which accounts for about 5% of all recorded melanoma cases. It is also known as "Hutchinson's melanotic freckle". This is one of the types of melanomas which is closely related to overexposure to the sun, more than other types of melanomas. This is because older people are more likely to develop lentigo maligna melanoma as they have been exposed to the sun's UV rays for a longer period over their lifetime. The most common location for diagnosis is the face.

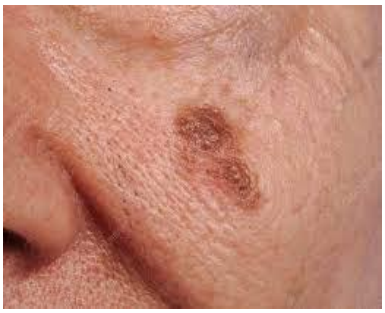


Figure 3.6: Lentigo Maligna Melanoma

Acral-lentiginous Melanoma: Acral-lentiginous melanoma is the most frequent type of melanoma that affects those with darker skin, accounting for 70% of cases in African Americans and 46% of cases in Asians. Melanoma on the soles of the feet, palms of the hands, and under the finger and toe nails are common sites for this type of growth.

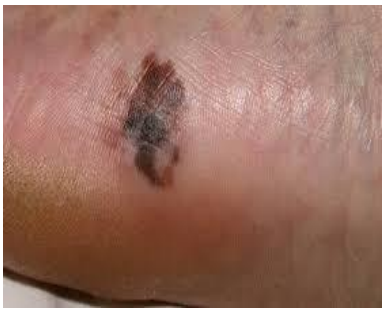


Figure 3.7: Acral-lentiginous Melanoma

Benign keratosis-like lesions

Benign keratosis-like lesions, also known as seborrheic keratoses, are benign (non-cancerous) growths that develop on the skin of some people as they age, which can be identified as a common sign of skin aging. These usually present themselves on the chest or the back, but can also occur anywhere else on the body. Most people develop at least one of these lesions across their body throughout their entire lifetime. The appearance of such lesions presents themselves as rough and bumpy, or the complete opposite – smooth and waxy.

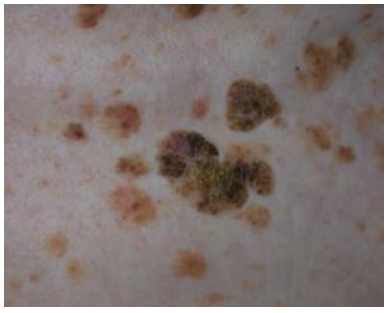


Figure 3.8: Benign keratosis-like lesions, aka Seborrheic Keratoses

Being extremely common across the world population, seborrheic keratoses have been recorded in an estimated 90% of all humans over the age of 30, but is highly uncommon in individuals under the age of 20. The lesions spread throughout the body over time.

The name may be misleading as the formation of these keratoses are not limited to seborrheic regions of the skin (i.e., scalp and chest regions).

Basal Cell Carcinoma

One of the most frequently recorded types of malignant skin growths, Basal Cell Carcinoma (BCC) arises from the uncontrollable and irregular growth of basal cells, one of the three primary types of cells present in the human skin's outermost surface, which usually sheds as new cells form. An estimated 3.6 million Americans are diagnosed with BCC each year, with more than 1 out of every 3 the BCC variety of skin cancer.

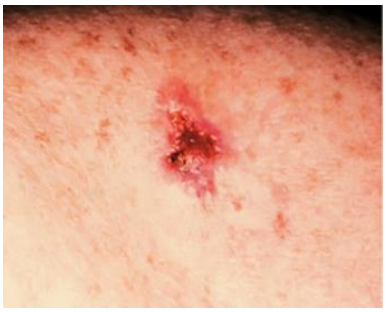


Figure 3.9: Basal Cell Carcinoma presenting as an open sore

BCCs are easy to spot as they usually appear as red or pink patches, sores or bumps, sometimes accompanied by oozing pus or bleeding. However, they may appear as pigmented on people with darker skin. This form of skin cancer is not as dangerous as others, and can be easily treated if caught early as they tend not to spread beyond their original point of inception. If left untreated, however, there is a chance that the carcinoma may grow invasively, penetrating tissue and bone in some cases.

Actinic keratoses and intraepithelial carcinoma

Actinic keratoses (AK), also known as Bowen's Disease or Squamous Cell Carcinoma, are all considered types of malignant keratinising tumors. It forms multiple pink papules (skin lesions) with a scaly surface, generally presenting itself on parts of the body with keratinoid cell structures, such as across the scalp, ears, arms, neck and back of the hands.

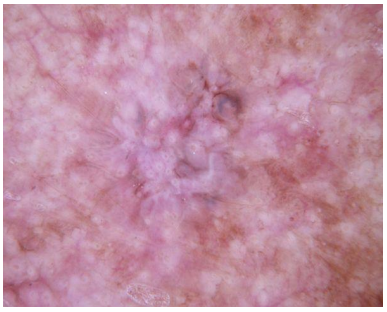


Figure 3.10: Actinic Keratosis

The most prevalent form of keratinocyte skin cancer is AK, which is caused by long-term sun exposure. There are two main forms of Actinic keratoses:

Pigmented Actinic Keratosis This form of AK mostly affects people with darker skin, with reported identifiers such as the irregular openings of follicles, and the presentation of grey dots arranged in symmetrical shapes between follicular structures. They form a prominent pseudo-network between all the affected follicular structures.

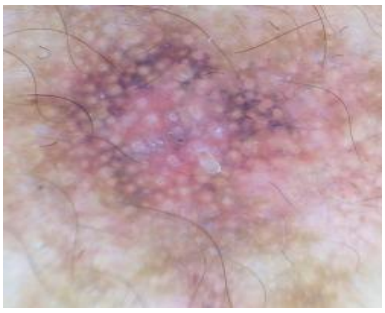


Figure 3.11: Pigmented Actinic Keratosis

Non-Pigmented Actinic Keratosis Non-pigmented forms of AK shows 4 specific features which allow for an accurate diagnosis with high specificity:

1. Fine, wavy vessels - these vessels form surrounding hair follicles.
2. Targetoid hair follicles - openings on the hair follicles that comprise of a yellowish solid that is surrounded by a white circular halo.
3. Erythema - shapeless areas of pale-red skin lacking any identifiable area of hypo-pigmentation.
4. Pink-to-Red pseudo-network - similar to Erythemas, but resembling network-like shape structure, corresponding with follicles of the skin.

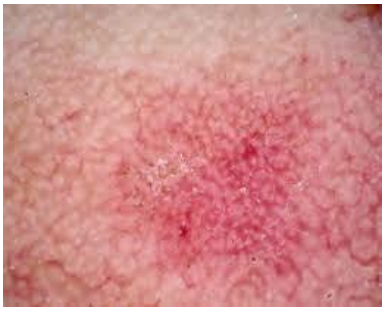


Figure 3.12: Non-Pigmented Actinic Keratosis

Vascular Lesions

These are skin and tissue abnormalities that are relatively common among people of all ages regardless of skin tone or gender, and are more commonly known as birthmarks. There are 2 categories of vascular lesions that may look similar, but

distinguish themselves in terms of their origin or required treatment. The categories are:

Hemangiomas: These are the most common forms of vascular lesions among children. They are non-cancerous tumors of the cells that line the blood vessels that appear at birth as reddish patches or overlying growths on the skin, and proceed to grow rapidly in size and thickness over a short period of time. As they grow, their pigmentation tends to darken and may sometimes lead to ulceration (a break of the upper skin layer) causing bleeding (eventually infection in worst cases). They grow for a period of 6-12 months, after which they stabilise and shrink in size over a few years, eventually fading out. Hemangiomas of the eyelid may lead to permanent loss of eyesight in some infants.

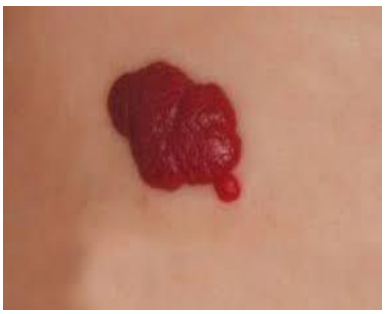


Figure 3.13: Non-Pigmented Actinic Keratosis

Pyogenic Granulomas: These are benign vascular lesions of the skin that usually form as the result of injury to local skin tissue. They have a rapid period of growth, forming red polyp-like masses that lead to ulcers that cause bleeding easily. This form of vascular lesions typically presents itself in children and pregnant women.

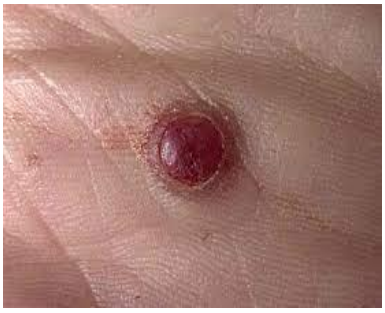


Figure 3.14: Non-Pigmented Actinic Keratosis

Dermatofibroma

These are small, benign (harmless) growths that present themselves inside the skin, typically pink in color in fair-skinned people and brownish on darker-skinned people. Irritation of the affected surface may lead to the darkening of the growth. Dermatofibromas are more likely to occur in women than in men, and are most commonly present on the legs.

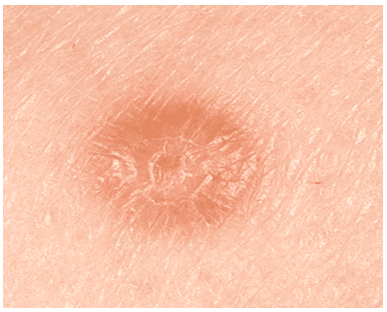


Figure 3.15: Dermatofibroma

These growths feel dense to the touch, feeling like a small stone or lump having formed under the skin. Typically, dermatofibromas will cause a dimpling of the skin around it, which is what helps distinguish this from other types of growths. They are typically painless, however, these growths may be accompanied by itching, tenderness or irritation.

The exact cause of dermatofibroma is unknown.

Chapter 4

Dataset Description

To train a diagnostic algorithm that is based on a neural-network, a large collection of sorted and labeled images is required. However, sources with reliable diagnoses of high-quality dermatoscopic images are either limited in data size or reliability.

Thanks to the work done by Tschandl et al [15], a large collection of annotated images has been collected into a dataset named "HAM10000 (Human Against Machine with 10000 training images)". This data repository combines data from many different population categories using multiple data collection methods over a period of about 20 years. Notable sources include the Department of Dermatology at the Medical University of Vienna, Austria, and the skin cancer practice of Cliff Rosendahl in Queensland, Australia. Due to this large variation in data collected, many processes of "cleaning" and preparing the data were undertaken to ensure that the images could be used to feed a specifically trained neural-network for testing and identification purposes. The HAM10000 dataset contains a total of 10,015 dermatoscopic simulacra that is accessible via the ISIC archive and in the Harvard Dataverse for public use in academic machine-learning research, acting as a benchmark for testing machine-learning techniques. Images in the dataset of seven (7) important diagnostic categories pertaining to pigmented lesions or growths. The images were either confirmed by pathology, follow-ups, in-vivo confocal microscopy or via expert consensus.

Design Type(s)	Database Creation Objective, Data Integration Objective, Image Format Conversion Objective
Measurement Type(s)	Skin Lesions
Technology Type(s)	Digital Curation
Factor Type(s)	Diagnosis, Diagnostic Procedure, Age, Biological Sex, Animal Body Part
Sample Characteristics(s)	Homo Sapiens, Skin of Body

Figure 4.1: Summary of meta-data in HAM10000 dataset

4.1 Dataset Background

The diagnostic technique that pertains to the examination of the skin using skin surface microscopy is known as Dermatoscopy, or dermoscopy, which is a method used to aid in the evaluation and identification of pigmented skin growths or lesions that may have otherwise been unable to have been identified by the naked eye. Dermatoscopic images allow for the training of artificial neural-networks so methods can be developed to automatically diagnose pigmented skin growths or lesions.

Binder et al.[16] trained an Artificial Neural Network (ANN) in 1994, using dermatoscopic images to differentiate melanomas from melanocytic nevi. However, the study was affected by the lack of available images due to the available datasets at the time being of a small sample size, as the training of ANNs require large sets of annotated and categorised images for reliable outcomes. The headway made in the field of graphics cards and machine learning techniques in recent years now allows for better and more complex neural-networks to be established, which allows better and more efficient identification of pigmented skin growths and lesions without any human supervision.

In 2000, a commercially available book named "Interactive Atlas of Dermoscopy" was released by Argenziano et al. containing a CD-ROM with over 2000 images of pigmented skin growths and lesions. Though this is a large dataset, the lack of public accessibility has made it a limited resource in the field of neural network training for skin cancer detection.

A dataset labeled PH^2 was publicly released in 2013 by Mendonca et al. [17] with 200 dermatoscopic figures, containing 160 images of melanocytic nevi and 40 images of melanoma confirmed via pathology. The public availability of the dataset allowed it to be a comprehensive benchmark for neural network training till recent years.

At the moment, the ISIC archive, a public repository of dermatoscopic images, contains about 23,665 images as of April 2020. The dataset is heavily focused towards melanocytic lesions and nevi, comprising about 90% of the archive. It is currently the world standard for the sourcing of dermatoscopic images for research purposes thanks to its large size, well-categorised structure and licensing permissions.

Class Label	Abbreviation	Class	Np. of Images
0	AKIEC	Bowen Disease	334
1	BCC	Basal Cell Carcinoma	583
2	BKL	Benign Keratosis-like Lesions	1674
3	DF	Dermatofibroma	122
4	MEL	Melanoma	2177
5	NV	Melanocytic Nevi	18,618
6	VASC	Vascular Lesions	157
		Total	**23,665**

Figure 4.2: Overview of ISIC archive and Distribution of Classes as of April 2020

Due to the unavailability of resources in the past, study and research were primarily concentrated on melanocytic lesions, and not on on-melanocytic pigmented lesions even though these are also a common occurrence. The discrepancy between the lack of diversity in available datasets and the variation of data due to the tendency of human behaviour, allowed for excellent results in closed conditions, but with modest results in clinical applications.

4.2 Dataset Classification Methods

As mentioned in the table above, The photographs in the HAM 10000 dataset were acquired over a period of 20 years from Cliff Rosendahl's skin cancer clinic in Queensland, Australia, and the Department of Dermatology at the Medical University of Vienna, Austria. The practice in Australia compiled their images in PowerPoint files and their metadata in Excel databases, whereas the Medical University of Vienna began compiling its data and images before the advent of digital cameras, resulting in photographs and data information being stored in various formats depending on the time of recording. This meant that the data had to be sorted and categorised into a singular form of formatting to allow for continuity and ease of access. The methods included:

Image and metadata extraction from PowerPoint files The Australian skin cancer practice compiled their monthly clinical reports and images into separate PowerPoint files, i.e. one file for every calendar month of diagnostics. Each slide of the monthly files contained a single image along with a note and a unique identifier. An automated approach using python was implemented by Tschandl et al to iterate through all the files and classify content according to their corresponding identifiers.

Digitisation of Diapositives The Department of Dermatology in Vienna used to store dermatoscopic images as diapositives before the invention of digital cameras. These diapositives were first digitised and saved as 8-bit JPEGs at 300 DPI. Images were then cropped with the lesions centered at an 800x600px ratio, with a manual histogram used to enhance the color and contrast of the scan as compared to the original.

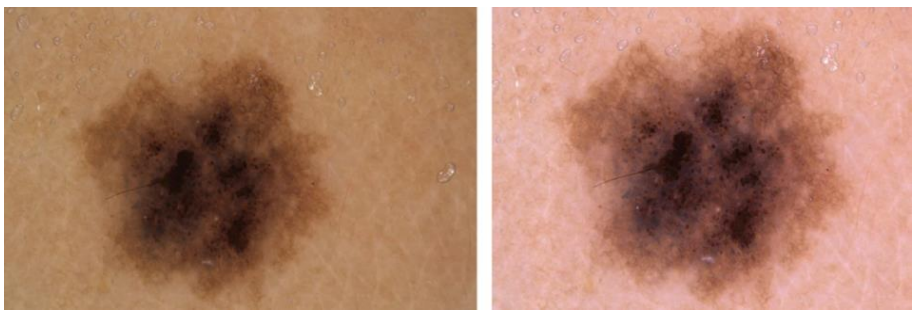


Figure 4.3: Original diapositive vs final image after manual quality review

Data Extraction from Digital Dermatoscopy System Years after the introduction of digital cameras, the Department of Dermatology shifted to the

usage of MoleMax HD, a digital dermatoscopy system. Data from the system was retrieved using proprietary software, and only three types of data were collected: benign non-melanocytic lesions, melanocytic nevi with more than 1.5 years of digital dermatoscopic follow-up, and excised lesions with a histopathologic report (which were later manually matched to their respective lesions). These images were also manually cropped to an 800x600px resolution, with the growths or lesions centered in the image.

Filtering of dermatoscopic images All the images collected contained both dermatoscopic images, as well as clinical close-ups, were taken by the medical staff. However, these additional images had no reliable source annotation, so the dermatoscopic images had to be manually separated from the rest of the images. 1501 images from the collection of the Australian data set were first labeled by hand into 3 categories – overviews, close-ups, and dermatoscopy. These hand-labeled images were then used to train an ANN to classify the remaining images automatically. A top-1 accuracy of 98.68% was achieved, and this helped speed up the process of future image categorisation for the HAM dataset, followed by a second manual review to remove any misclassifications.

Unifying pathological diagnoses Histopathological diagnoses, aka biopsies, recorded from both the Australian practice and the Department of Dermatoscopy showed multiple inconsistencies between their records consisting of typos, uncertain diagnoses, different terminologies and classifications used, and multiple different diagnoses per lesion.

Uncertain diagnoses were discarded, except for those pertaining to melanomas that were associated with nevi.

The rest of the diagnoses were classified into seven (7) generic classes, removing any ambiguity in generalisation. From both data retrieval sites, these groupings covered more than 95% of all recorded pigmented lesions and growths, which will be further discussed below.

All pictures in the data set were subjected to a final manual inspection and validation to exclude data that contained:

- Identifiable content (tattoos, garments, etc.)
- Blurry or out of focus images
- Images with artifacts (visible anomalies)
- Non-pigmented lesions
- Close-ups and overviews which were not automatically removed by the ANNs

This was followed by manual histogram corrections of color and contrast manually where required.

4.3 Dataset Sources and Comparisons

The HAM10000 database is available for public use at the Harvard Dataverse, falling under the Creative Commons Attribution-NonCommercial 4.0 International Public License, and is compiled with images and metadata the School of Medicine at the University of Queensland, and the ViDIR Group (Department of Dermatology, Medical University of Vienna, Austria).
The table below illustrates the publicly available sources of dermatoscopic image datasets, namely the PH² and the Interactive Atlas of Dermatoscopy, as well as the information about the HAM10000 dataset used in our project.

Dataset	Total Images	Pathological Verification	akiec	bcc	bkl	df	mel	nv	vasc
PH2	200	20.5%	-	-	-	-	40	160	-
Atlas	1024	unknown	5	42	70	20	275	582	30
ISIC 2017	13786	26.3%	2	33	575	7	1019	11861	15
Rosendahl	2259	100%	295	296	490	30	342	803	3
ViDIR Legacy	439	100%	0	5	10	4	67	350	3
ViDIR Current	3363	77.1%	32	211	475	51	680	1832	82
ViDIR MoleMax	3954	1.2%	0	2	124	30	24	3720	54
HAM10000	10015	53.3	327	514	1099	115	1113	6705	142

Table 4.1: Data repositories of dermatoscopic images that are freely available compared to HAM10000

The description of sourcing for the HAM10000 dataset as compiled by Tschandl et al. is as follows:

Rosendahl Images of growth or lesions for the Rosendahl part of the HAM10000 dataset were collected from the archives of the skin cancer practice of Cliff Rosendahl (CR), of the School of Medicine at the University of Queensland, after approval from the board of institutional ethics. Images prior compiled by CR were taken either using DermLite™ Fluid or DermLite™ DL3 with an immersion fluid (ultrasound gel). Each slide had a dermatoscopic image as well as a unique lesion ID number that linked the image to an Excel database that provided clinical metadata as well as histopathologic diagnoses. The image collection of this manner ranged from 2008 till May 2017, comprising 34.2GB of data and images, which allowed the extraction of 36,802 images with matching histological reports. The final Rosendahl dataset as outlined in the above table was achieved after removal methods applied were applied (refer to chapter 4.2).

ViDIR Group The ViDIR Group had image datasets compiled in different methods during different time periods, so information compilation of different methods were required, after approval from the ethics committee of the Medical University of Vienna.

1. ***Legacy Diapositives (ViDIR Legacy)*** The first set of ViDIR group dermatoscopic images and metadata predate the availability of digital cameras, from a time when dermatoscopic images had to be archived as diapositives, recorded by analog cameras. These photos were originally taken using the Heine Dermaphot system using an immersion fluid, and physical copies were prepared for research and archiving using the E-6 methodology.
2. ***Current Database (ViDIR Current)*** Dermatoscopic images collected at ViDIR post-2005 were documented using a DermLite™ FOTO system, and stored on a central database. 3 copies of each image were taken at different magnifications to allow for better visualisation.
3. ***MoleMax (ViDIR MoleMax)*** Patients diagnosed by the Department of Dermatology in Vienna diagnosed as high risk were offered digital dermatoscopic follow-ups for the surgical removal of early melanoma, to allow for a decrease in the removal of melanocytic nevi. The time period between follow-ups were between 6-12 months. Since 2015, the MoleMax HD System (Derma Medical Systems, Vienna) has been employed for the gathering and storage of digital dermatoscopic pictures. Multiple nevi were thus recorded for most patients.

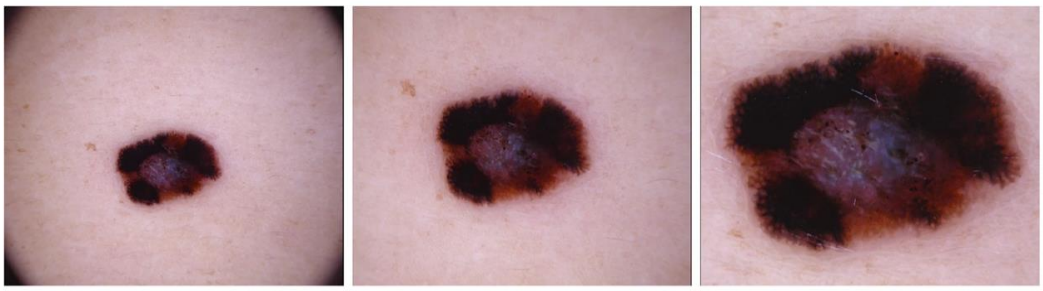


Figure 4.4: Original source of the images tended to have 3 images of different magnifications and angles instead of the cropping of a large source image

4.4 Technical Validation

Where required, expert dermatologists were consulted to help perform manual histogram corrections for the ease of visual access of human viewers. Underexposed photographs and images with a discernible tint offsetting the natural skin tone in the image were the only ones that needed to be corrected. Illuminant color estimations were tested post-correction, according to the grey-world assumption (an image white-balance method that assumes that most of the image in question is a neutral gray color). This allows for a graphing of the average color present in the image, to show the improvement in image clarity comparing the images pre- and post- histography corrections, as shown in the image below.

4.5 Technical Terms and Definitions

All images in the HAM10000 dataset have been categorised into a metadata csv file according to seven classification categories, each with multiple sub-categories for ease of access and rationalisation of data. These are:

- lesion_id
- image_id
- dx
- dx_type
- age
- localisation

Refer below for the explanation of technical terms and keywords used in the dataset.

Lesion Identification (*lesion_id, image_id)*

Every image in the dataset has its unique lesion identifying number, noted as the "lesion_id" in the metadata file that corresponds to each unique lesion recorded in the HAM10000 dataset. These IDs are followed by the prefix "HAM_". They also have a unique image identification number, or "image_id" that identifies the image with the corresponding image file from the ISIC database, followed by the prefix "ISIC_".

The dataset also includes a collection of data for flattened 28x28x3 (pixelated) versions of the available images [HMNIST_28_28_RGB], where each image is correlated with a label between zero (0) to six (6). Each label corresponds to the diagnosis (dx) according to the metadata file, in the order mentioned above (1 is akeic, 2 is bcc, etc.).

Lesion Diagnosis (*dx, dx_type)*

Each image in the dataset is classified using seven (7) unique identifying categories according to their respective medical diagnoses under the category type of "dx" (refer to chapter 3 for further details and explanations):

akiec Actinic Keratoses and Intraepithelial Carcinoma (aka Bowen's Disease)

bcc Basal Cell Carcinoma

bkl Benign keratosis (generic class including seborrheic keratoses, solar lentigo and lichen-planus like keratoses)

df Dermatofibroma

mel Melanoma

nv Melanocytic Nevi

vasc Vascular Skin Lesions

The dataset contains meta-data of the type of diagnosis (dx_type) for each image, which are labeled as follows:

Consensus **(consensus)** Tschandl et al. presented an expert consensus on typical benign cases that had no follow-up or histopathology, given that all authors independently provided the same absolute benign diagnoses. Lesions of this type were originally recorded for educational purposes and did not require a follow-up for diagnosis confirmation.

Histopathology **(histo)** These diagnoses were originally carried out by specialised dermatopathologists. All images under this category were manually reviewed with their corresponding histopathology to check for validity. If the diagnosis was invalid, the data was checked for sample mismatching and re-examined where necessary. Ambiguous diagnoses were discarded.

Follow-up examinations **(follow_up)** Initial diagnoses that showed no changes after 3 follow-up visits (or 1.5 years) were acknowledged as evidence of benignity. Nevi (nv) was the only diagnostic category that were labeled with this category of diagnostic type, as dermatofibromas (df), seborrheic keratoses (bkl) and vascular lesions (vasc) are not monitored by dermatologists after the first visit. Any changes in the lesions were validated by one of the authors who had 20 years of experience in digital dermatoscopic follow-up.

In-vivo confocal microscopy **(confocal)** An in-vivo imaging technique that allowed for a near-cellular level of resolution known as reflectance confocal microscopy was used to verify some benign keratoses in the facial region.

Other Categories

Three other categories were employed to label the data, including the age of the patient in question, labeled as “age”, the gender of the patient, labeled as “sex” (male, female and other), and the localisation of the lesion or growth on the body, ie where the lesion is located on the patient, labeled “localisation”.

Chapter 5

Dataset Pre-processing

5.1 Libraries Required

Multiple code-based libraries were required for the pre-processing of our dataset, the libraries we have used are mentioned in the table below.

Python Libraries	numpy, matplotlib, seaborn, PIL, pandas, datetime, os, cv2, itertools, tqdm, glob
Pytorch Libraries	torch, torch.autograd, torch.utils.data, torchvision
Sklearn (Scikit-Learn) Libraries	sklearn.metrics, sklearn.model_selection
Keras Libraries	tensorflow, keras.preprocessing.image, keras.callbacks, tensorflow.keras.models, tensorflow.keras.layers

Table 5.1: Libraries implemented in our code including their code source

5.2 Reading and Processing Data

The dataset contains a compiled csv file (hmnist_28_28_RGB.csv) of all the RGB data of the images, we have used this data for our model as we were aiming for our implementation to work even with very low resolution colored images. For VGG, ResNet and DenseNet, we have made a dictionary of images from both the HAM10000 images part1.zip and HAM10000 images part2.zip folders in JPG format. After this, we have labeled those images from 0 to 6 based on the type of cancer they represent by comparing the image file name with the data in the metadata.csv file

5.3 Data Cleaning

Often, especially large datasets, have missing or null values. These have to be filled, and we have searched for and filled in these values with their mean. Duplicate images might create a bias in the outcome and yield unexpected results, hence, we have checked for duplicates and used only those images that are unique. We found 4501 unique images from our database.

```
unduplicated    5514
duplicated      4501
Name: duplicates, dtype: int64
```

Figure 5.1: Check for image duplicates in our dataset

5.4 Exploratory data analysis (EDA)

This dataset also comes with an extensive amount of supplementary information about the localization and diagnosis of skin growth. The information also contains data about the sex and age of the affected people.

Distribution of type of cancer From the distribution of the type of cancer, we can see that akiec (Actinic keratoses and intraepithelial carcinoma) has a much higher occurrence rate than even all the other types combined. This might create huge bias in the learning of the implemented model and we will have to manipulate the data to feed a balanced set of data to our model.

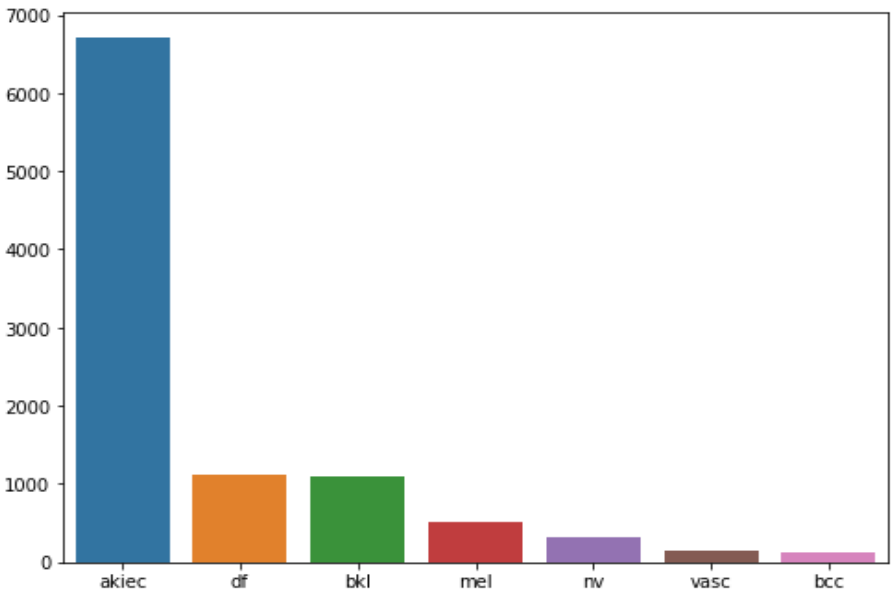


Figure 5.2: Distribution of type of cancer

Distribution with diagnosis type The distribution of different types of diagnostic techniques that were used to identify the skin growths.

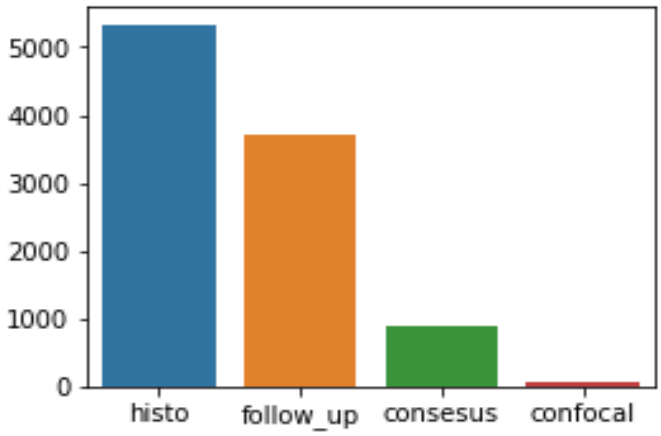


Figure 5.3: Distribution with diagnosis type

Distribution with localisation From the graph below we can see the region of the human body where skin growths occur and how prevalent they are.

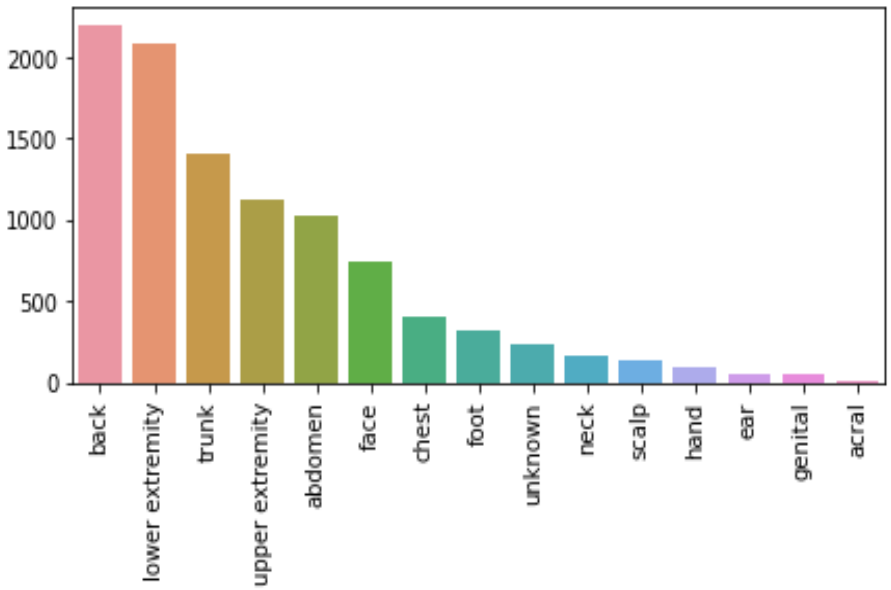


Figure 5.4: Distribution with localisation

Distribution with gender From the distribution of gender, we can see that more males are affected by cancer than females.

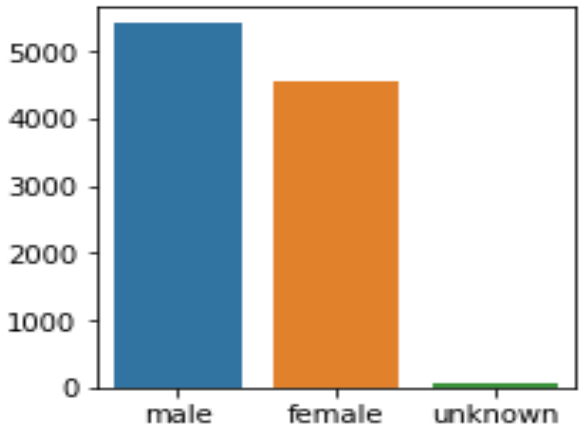


Figure 5.5: Distribution with gender

Distribution with age The data below shows that cancer is most common within the age range of 35 to 50. The affected number within this age group is almost as much as all the other ages combined.

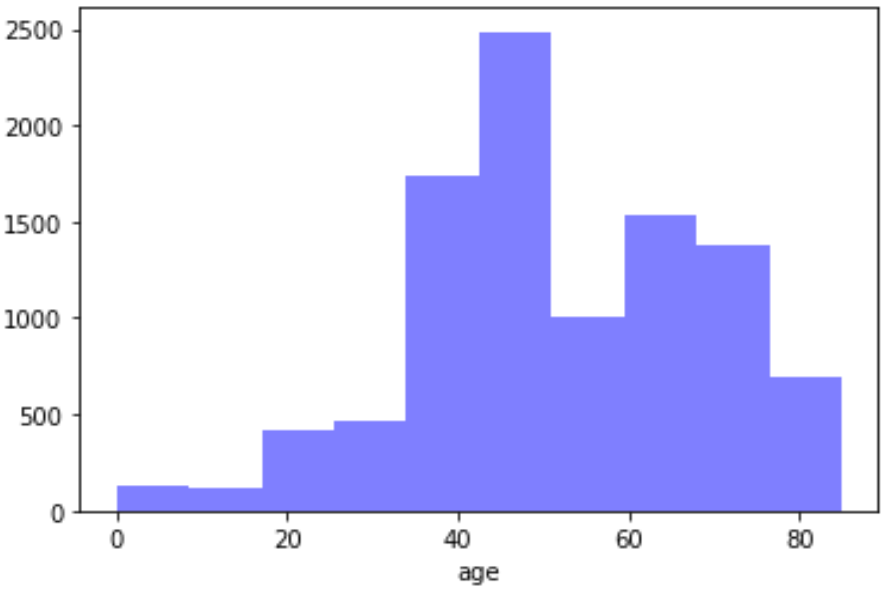


Figure 5.6: Distribution with age

Scatter-plot with distribution of age against distribution of cancer The scatterplot shows that nv (Melanocytic nevi) and vasc (Vascular lesions) is common in all age groups while the other 5 types mostly occur during adolescence, adulthood and old age. Also, bkl (Benign keratosis-like lesions) and mel (Melanoma) might occur in babies and young children.

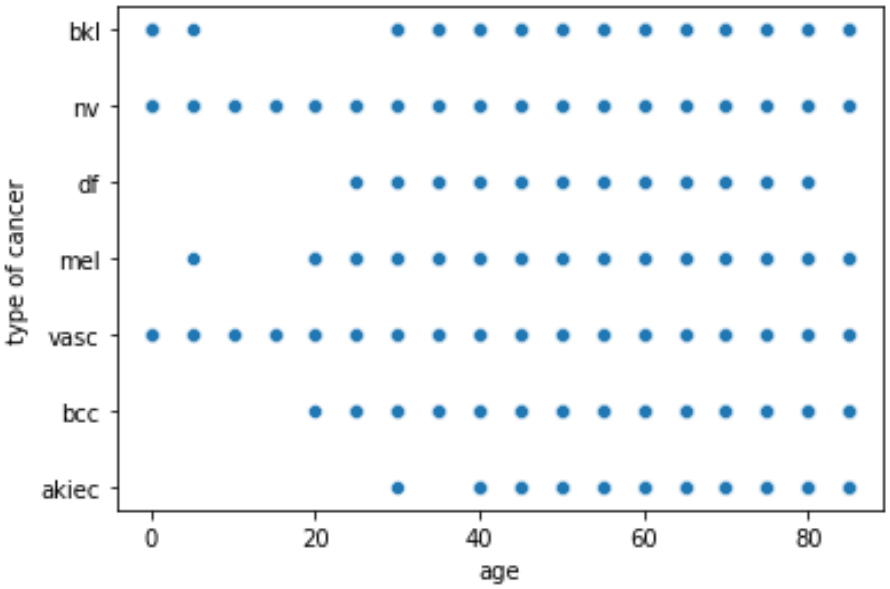


Figure 5.7: Distribution of age against type of cancer

Chapter 6

Model Training

6.1 Loading and Resizing of Images

TensorFlow cannot work with the images of the dataset directly as they have a much larger dimension (450x600x3) than what it can handle. Hence, for our model we have used the compiled RGB data of the provided CSV file within the dataset and for VGG, ResNet and DenseNet, we have used the provided images which were scaled down to 244x244x3.

6.2 Train-Test Split

We divided the dataset into an 80:20 training and testing set. This is a typical data science practice.The training dataset is the foundation for any computation that demonstrates how advanced technologies, such as neural networks, may be used to learn and generate complex outputs. It contains both information and the corresponding expected yield. The training dataset's goal is to provide "ground truth" data for our calculations. Regardless, the test dataset is used to assess how successfully our computation was produced using the preparation dataset. Because the calculation will almost certainly "know" the normal yield, we can't just repeat the training dataset in the testing step, which defeats the purpose of evaluating the computation.

6.3 Normalization

We normalized the train and test sets by subtracting from their respective mean values and then dividing by their value of standard deviation. All values from 0-255 are normalized to 0-1. Normalization is a data preparation procedure that is frequently described as a major element of AI or machine learning. The purpose

of normalization is to transform the numeric section evaluations in the dataset to a standard scale without misrepresenting quality scope discrepancies. It is only necessary when features have a wide range of values, like in our case.

6.4 Label-Encoding

There are seven different classes of skin cancer types on the labels, ranging from 0 to 6. The following is the indexing for data classification purposes:

Label name	nv	mel	bkl	bcc	akiec	vasc	df
Index value	0	1	2	3	4	5	6

6.5 Splitting Training and Validation Split

The training set was divided into two parts: a small fraction (20%) served as a validation set for the model, while the remainder (80%) was utilized to train the model. To avoid over-fitting, a validation dataset is utilized. We're not changing the network's weights using this data set; instead, we're confirming that any increase in correctness over the train set is indeed an improvement over data that the system doesn't already have, or at the very least haven't been exposed to. (i.e. validation data set). If the accuracy of the train set of data improves while the correctness of the validation set of data declines, your neural network is over-fitted, and you should cease training.

6.6 Data Augmentation

As Melanocytic nevi accounts for over half of the training data, we can observe that there is a severe class imbalance in the data. We believe we can enhance our data to tackle this challenge. To avoid the problems caused by over-fitting, we must artificially expand our HAM 10000 dataset. We can significantly increase the size of our current dataset. To reproduce the variations, the idea is to modify the preparation instructions with minor changes. Information increase strategies are procedures that vary the cluster depiction while keeping the mark the same by modifying the preparation information. Gray-scales, horizontal and vertical flips, random cropping, color jitters, translations, rotations, and other common augmentations are only a few examples

- We have re-scaled the images, which simply means to change the range of pixel values from [0, 255] to [0, 1].

- We rotated the photographs at random angles between 0 and 10 degrees, which resulted in some blank pixels that needed to be filled in. The default value for the fill mode argument is "nearest," which simply fills the empty area with the nearest pixel values. This is done so that our model may learn more when it looks at similar images from different perspectives.
- We have also shifted the images, both vertically and horizontally as realistically, images fed to the model will not always be centered
- We have also sheared the images randomly up-to 20%

Chapter 7

Model Building and Evaluation by CNN Model using Keras Sequential API

7.1 What is CNN?

Convolutional Neural Networks (CNNs) are developed to detect patterns in pixel images with minimal pre-processing. Convolutional Neural Networks (CNN) are a form of neural network with a grid-like structure that is used to evaluate data. As digital picture are essentially a binary representation of data made up of a sequence of pixels in a grid-like pattern, it is excellent for this type of neural network, with each cell storing a pixel value that depicts the color and brightness of each pixel.

When we humans view a picture, our brain uses neurons to analyze this huge amount of data. They are linked in a network and operate in such a way that their whole field of view is covered. In a CNN, each neuron responds only to inputs in its own receptive area. Layers are ordered in a manner such that the image is broken down into finer, simpler patterns including lines, curves, and so on, before moving on to more complex forms such as objects and facial features. CNN may be used to offer vision to computers in some ways.

CNNs are usually made up of three layers: convolution, pooling, and fully connected. The majority of the computations are done in the convolution layer. The kernel and the limited part of the receptive field are both matrices used in this calculation. The kernel matrix is smaller than a picture in terms of size, but it contains more information. So, if the picture has three RGB channels, the kernel will have a lower height and breadth in terms of spatial dimensions, but the depth will span all of them. A dot product of the two matrices are calculated in this layer.

According to the stride, the kernel moves over the image's height and breadth. The activation map is a 2D representation of the picture that results from this process. The activation map shows the kernel's reaction at each picture location. The formula

for Output Weight during Convolution is as follows,

$$W_{out} = \frac{W - F + 2P}{S} + 1 \tag{7.1}$$

This will result in an output volume, which has a size of,

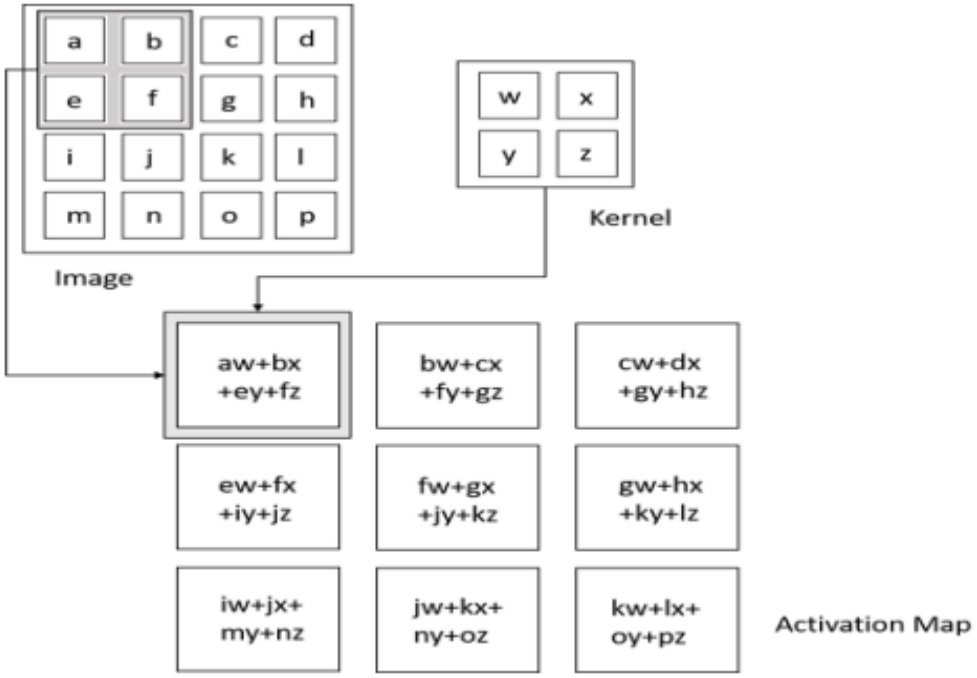


Figure 7.1: Convolution Operation (Source: Deep Learning by I. Goodfellow, Y. Bengio)

The pooling layer then sums adjacent outputs to determine the greatest output at specific places. As a result, the size of the representation is reduced, and computing time and weights are reduced. Max pooling is the most widely used of various pooling functions, and we utilized it in our thesis. Max pooling determines a neighborhood's maximum output.

If we have an activation map of size $W * W * D$, a pooling kernel of spatial size F, and stride S, then the size of output volume (output weight) during MaxPooling can be calculated using the following formula:

$$W_{out} = \frac{W - F + 2P}{S} + 1 \tag{7.2}$$

This will result in an output volume of size *(W * W * D)*. For the computer to detect an item that may appear anywhere in the picture, pooling is required.

The neurons of a completely linked layer are all connected to the preceding and subsequent layers. This cannot be calculated using basic dot multiplication. This

layer connects the input and output representations.

However, there is an issue. Images are not linear, although convolution is. Following the convolutional layer, non-linearity layers are immediately added for the determination of a non-linear activation map. Non-linear operations come in a variety of forms, the most common of which are listed below:

Sigmoid a sigmoid operation uses the mathematical formula,

$$\sigma(\kappa) = 1/(1 + e^{\kappa})$$

It transforms a real-valued number into the 0-1 range. The drawback of this nonlinearity is that the gradient is essentially nil when the activation occurs at either end. Backpropagation will almost definitely be 0 if the local gradient is extremely near to 0. In addition, if the input data is always positive, the output will be either entirely positive or entirely negative. This results in a zig-zag weight gradient pattern.

Tanh Real numbers are transformed in the range [-1,1]. The activation reaches a saturation point, yet the output remains zero-centered.

ReLU An acronym for "Rectified Linear Unit" which calculates the function,

$$f(\kappa) = max(0, \kappa)$$

The threshold for activation is set to zero.

ReLU is more dependable, and convergence occurs considerably more quickly (by six-fold), and this is what we have used in our work. If a huge gradient flows in such a manner that it updates in such a way that the neuron does not get updated any more, a problem may arise during the training stage. We were able to get around this by determining an appropriate learning rate.

Researchers established that it is typically better to go deeper when it comes to convolutional neural networks. However, it is observed that after traversing some depth, the performance actually degrades. There are better convolutional networks today with a huge database from ImageNet.

ImageNet is a collection of over 15 million high-resolution photos with over 22,000 categories that have been identified. As a result, we've taken advantage of this dataset to improve the accuracy of our results.

7.2 Our Model

For building our model, we have used the Keras Sequential API. Models with this API are built by adding layers upon layers. We need to specify the input shape because the models need to know what shape of input it will get.

The first layer is the Conv2D layer. Conv2D is a kernel that has dimensions which are smaller than the images that pass through it. This is so that the layer can go across the image. Conv2D gives us an integer value that helps determine the total number of output filters we will receive from convolution. 32 filters in Conv2D are the most common number but we have utilized Conv2D layers that go up to 128 filters. These layers basically alter and transform the image.

Another layer is the pooling layer (MaxPool2D) which usually comes after a convolutional layer. The layer basically allows us to downsample (pool) images. This layer compares pixels from the input images and chooses the maximum value from them to create the downsampled image. This method is used to aid in the extraction of smooth and sharp features. These features may include edges, points, etcetera.

The Flatten layer converts the images into a one-dimensional array so that each pixel data can be connected to the next layer. It brings together all the features that were extracted in the layers together. Next, we add and pass data through dense layers. Once the images pass through the previous convolutional layers, we do not know exactly how the outputs turned out.

The Dense layers function is to find the connection between all the features that were fed to it without any input parameters beyond convoluted layers.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 28, 28, 16)	448
conv2d_1 (Conv2D)	(None, 26, 26, 32)	4640
conv2d_2 (Conv2D)	(None, 24, 24, 32)	9248
max_pooling2d (MaxPooling2D)	(None, 12, 12, 32)	0
conv2d_3 (Conv2D)	(None, 12, 12, 32)	9248
conv2d_4 (Conv2D)	(None, 10, 10, 64)	18496
conv2d_5 (Conv2D)	(None, 8, 8, 64)	36928
conv2d_6 (Conv2D)	(None, 6, 6, 128)	73856
conv2d_7 (Conv2D)	(None, 6, 6, 128)	147584
max_pooling2d_1 (MaxPooling2	(None, 3, 3, 128)	0
conv2d_8 (Conv2D)	(None, 1, 1, 64)	73792
conv2d_9 (Conv2D)	(None, 1, 1, 64)	36928
max_pooling2d_2 (MaxPooling2	(None, 1, 1, 64)	0
flatten (Flatten)	(None, 64)	0
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 64)	4160
dense_2 (Dense)	(None, 32)	2080
dense_3 (Dense)	(None, 32)	1056
dense_4 (Dense)	(None, 16)	528
dense_5 (Dense)	(None, 7)	119

Total params: 423,271
Trainable params: 423,271
Non-trainable params: 0

Figure 7.2: Architecture of Our Model

7.3 Annealing

ReduceLROnPlateau

Reduce the learning rate when a metric no longer improves.

Models often benefit from lowering the learning rate by a factor of 2-10 when learning becomes static. If no progress is noticed after a 'patience' number of epochs, this callback watches the quantity and slows down the learning rate.

We observed that after about the 19th or 20th epoch, the learning rate reduction came into play and this yielded better results.

7.4 Setting Optimizer

For our model, we have used the Adam optimizer.

Adam is a deep neural network training algorithm that uses an adaptive learning-rate optimization mechanism.

The approach is "computationally efficient, has little memory requirement, invariant to diagonal re-scaling of gradients, and is well suited for problems that are large in terms of data/parameters" according to Kingma et al.[18].

7.5 Fitting the Model

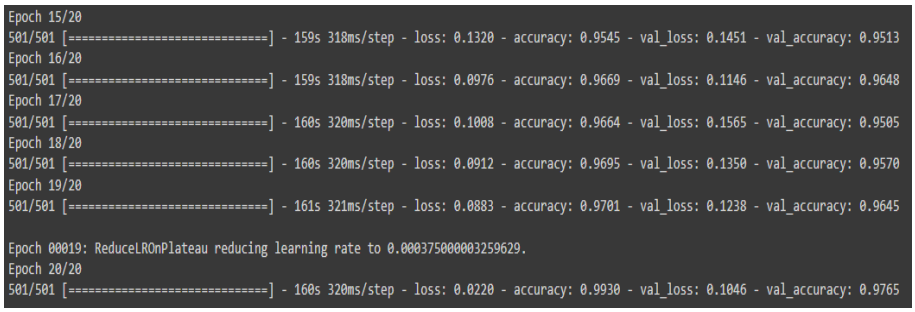


Figure 7.3: Validation accuracy (last 5 epochs of our model)

7.6 Model Evaluation Using CNN

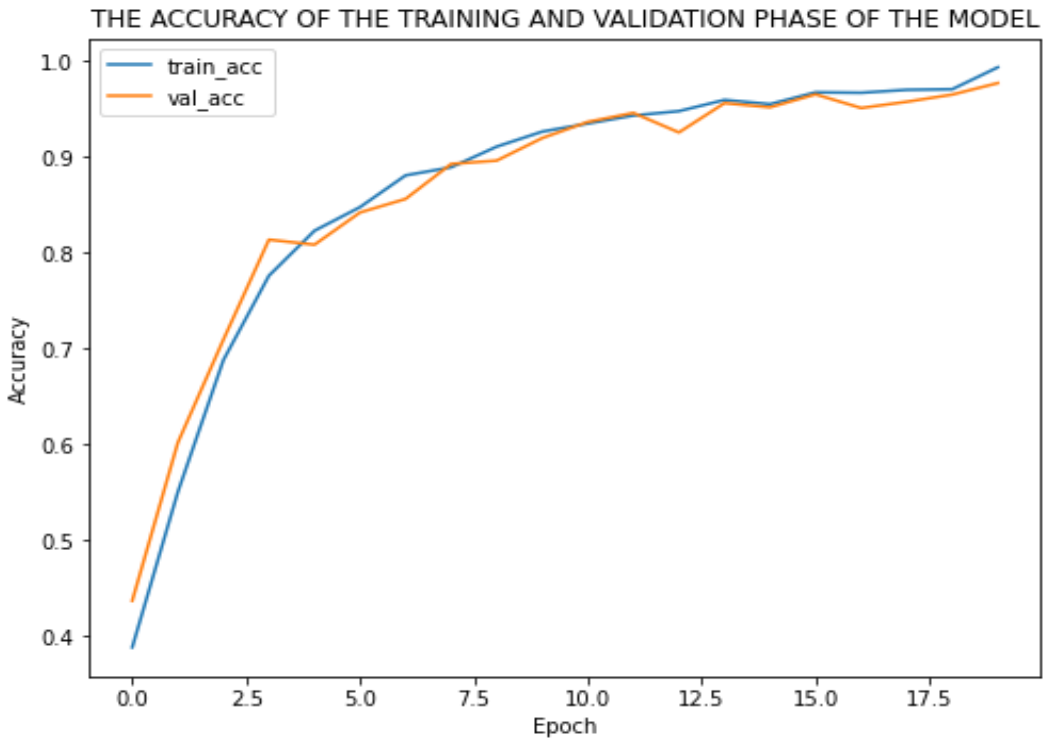


Figure 7.4: Accuracy of the Training and Validation Phase of Our Model

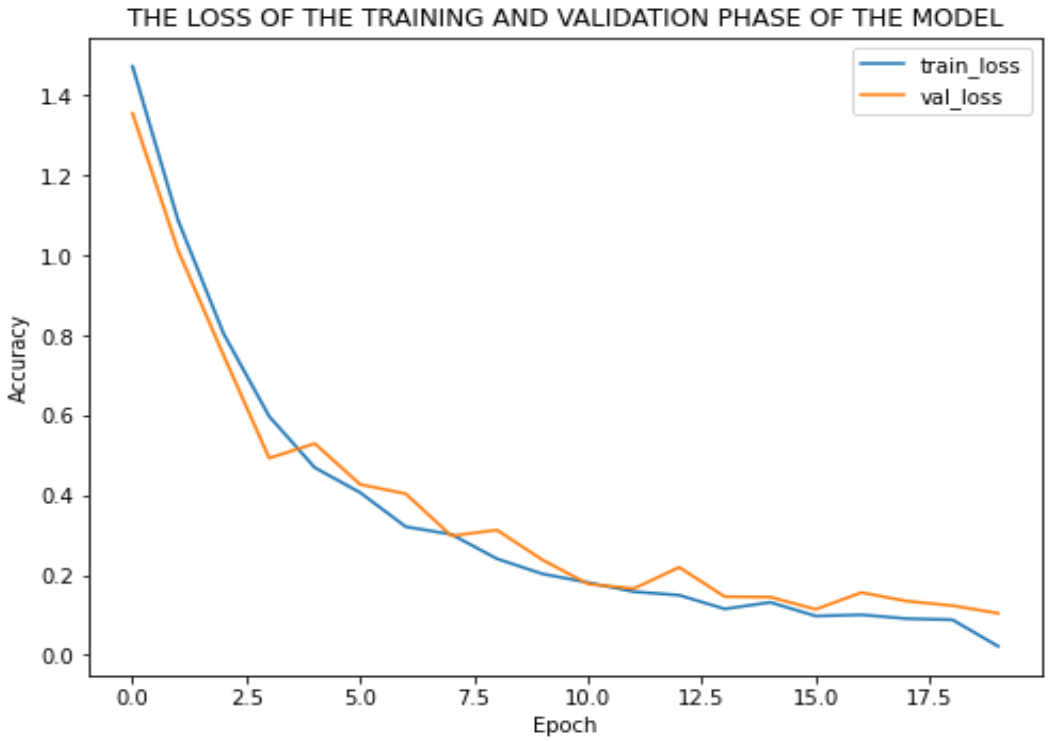


Figure 7.5: Loss of the Training and Validation Phase of Our Model

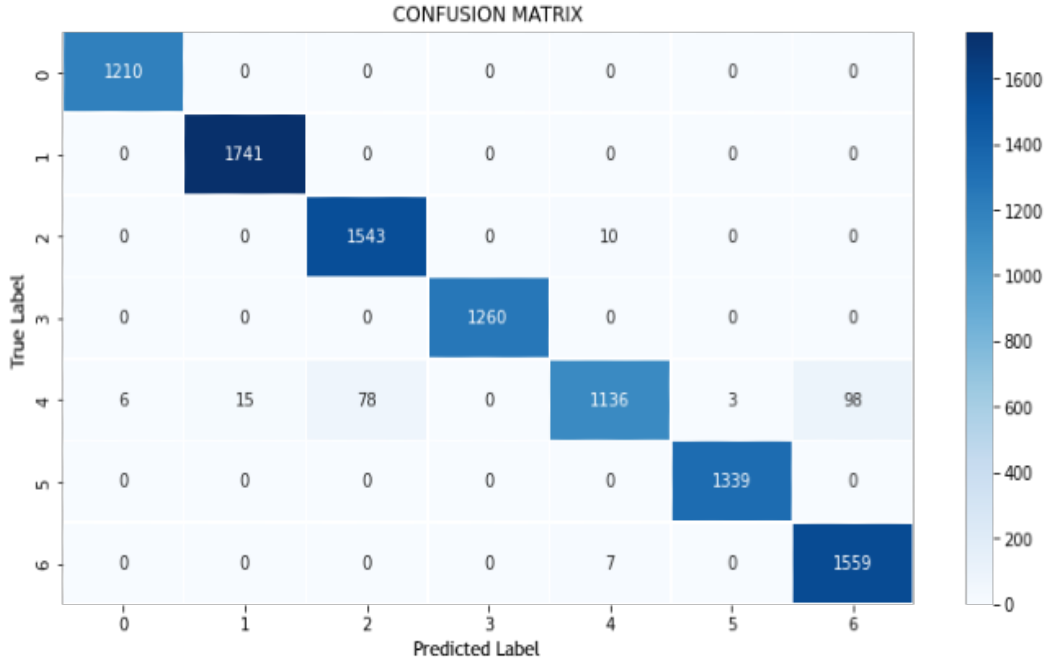


Figure 7.6: Confusion Matrix of Our Model using Testing Data

	precision	recall	f1-score	support
nv	1.00	1.00	1.00	1210
mel	0.99	1.00	1.00	1741
bkl	0.95	0.99	0.97	1553
bcc	1.00	1.00	1.00	1260
akiec	0.99	0.85	0.91	1336
vasc	1.00	1.00	1.00	1339
df	0.94	1.00	0.97	1566
accuracy			0.98	10005
macro avg	0.98	0.98	0.98	10005
weighted avg	0.98	0.98	0.98	10005

Figure 7.7: Precision, Recall and F1 score of our model

Chapter 8

Model Building and Evaluation using RESNET50

8.1 ResNet & ResNet50

A residual neural network (ResNet) is a type of artificial neural network (ANN) created in 2015 by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, and it is based on structures known from pyramidal cells in the cerebral cortex, according to the article[19]. Residual neural networks employ skip connections, also known as shortcuts, to skip over some layers. The bulk of ResNet models employ batch normalization in between double- or triple-layer skips with non-linearities (ReLU).

To solve a complicated problem, we generally stack some extra layers in Deep Neural Networks, which increases accuracy and speed. The idea behind adding more layers is that as time goes on, these layers would learn increasingly complex characteristics. In the instance of image identification, the first layer may learn to detect edges, the second layer may learn to recognize textures, and the third layer may learn to do the same. The traditional Convolutional neural network model, on the other hand, has been discovered to have a maximum depth threshold. By enabling the gradient to travel through an extra shortcut channel, ResNet's skip connections solve the problem of vanishing gradients in deep neural networks. These connections also help the model learn the identity functions, guaranteeing that the higher layer performs as well as, if not better than, the bottom layer.

The ResNet network uses a 34-layer simple network design based on VGG-19, after which the shortcut connection is added. These shortcut connections, as seen in the picture below, transform the design into a residual network.

The ResNet network employs a VGG-19-inspired 34-layer plain network architecture, after which the shortcut connection is added. As seen in the diagram below, these shortcut connections change the design into the residual network.

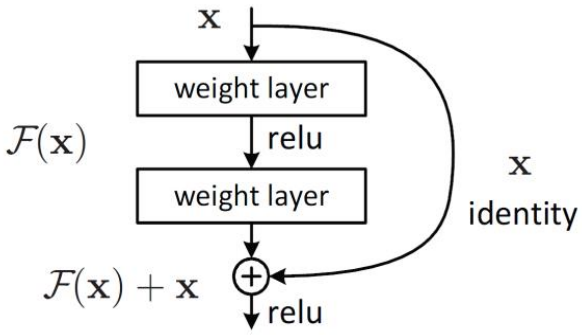


Figure 8.1: ResNet50 Algorithm (Source:neurohive.io)

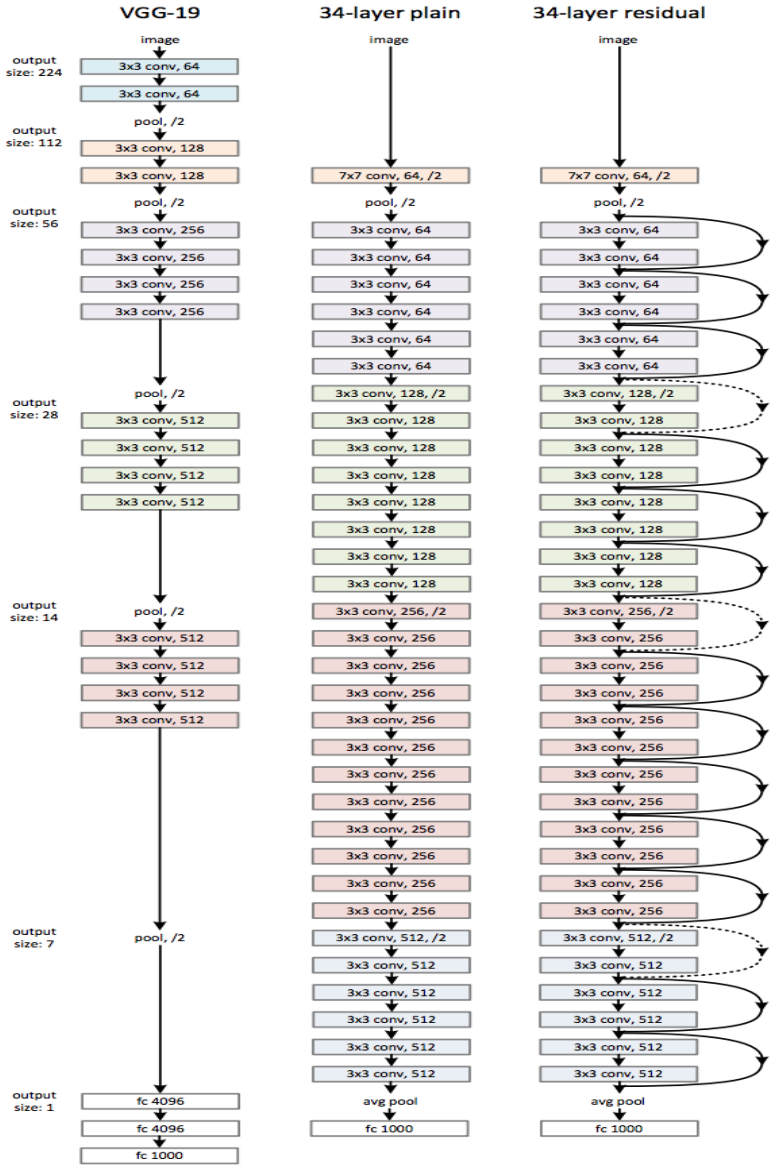


Figure 8.2: Difference between Plain and Residual Networks (Source: medium.com)

8.2 Model Evaluation using ResNet

```
[epoch 10], [iter 100 / 1124], [train loss 0.23978], [train acc 0.90531]
[epoch 10], [iter 200 / 1124], [train loss 0.23824], [train acc 0.90406]
[epoch 10], [iter 300 / 1124], [train loss 0.25137], [train acc 0.90115]
[epoch 10], [iter 400 / 1124], [train loss 0.24240], [train acc 0.90500]
[epoch 10], [iter 500 / 1124], [train loss 0.24044], [train acc 0.90669]
[epoch 10], [iter 600 / 1124], [train loss 0.24163], [train acc 0.90594]
[epoch 10], [iter 700 / 1124], [train loss 0.24518], [train acc 0.90496]
[epoch 10], [iter 800 / 1124], [train loss 0.24393], [train acc 0.90617]
[epoch 10], [iter 900 / 1124], [train loss 0.24207], [train acc 0.90677]
[epoch 10], [iter 1000 / 1124], [train loss 0.24099], [train acc 0.90663]
[epoch 10], [iter 1100 / 1124], [train loss 0.24057], [train acc 0.90687]
------------------------------------------------------------
[epoch 10], [val loss 0.39622], [val acc 0.87399]
------------------------------------------------------------
```

Figure 8.3: Validation Accuracy of ResNet50 (last 5 epochs)

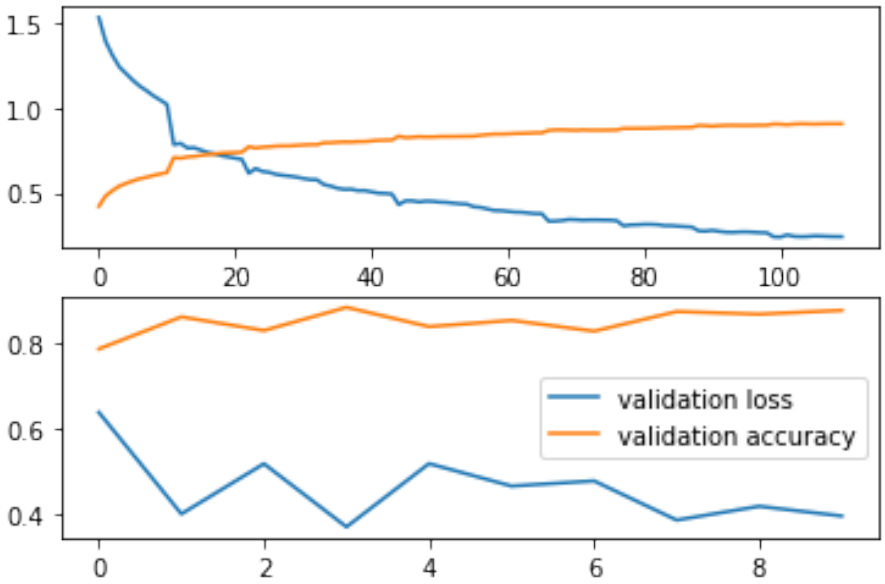


Figure 8.4: Loss and Accuracy of the Training and Validation Phase of the ResNet50 Model

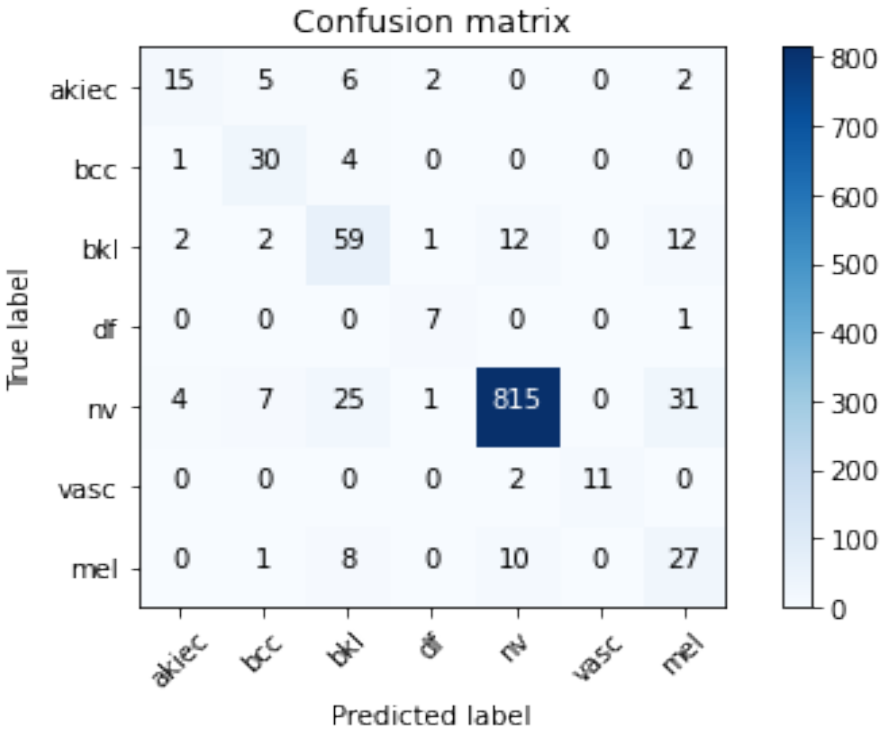


Figure 8.5: Confusion Matrix of ResNet50 using Testing Data

	precision	recall	f1-score	support
akiec	0.68	0.50	0.58	30
bcc	0.67	0.86	0.75	35
bkl	0.58	0.67	0.62	88
df	0.64	0.88	0.74	8
nv	0.97	0.92	0.95	883
vasc	1.00	0.85	0.92	13
mel	0.37	0.59	0.45	46
accuracy			0.87	1103
macro avg	0.70	0.75	0.71	1103
weighted avg	0.90	0.87	0.88	1103

Figure 8.6: Precision, Recall, F1 score and Support values for ResNet50

Chapter 9

Model Building and Evaluation using DENSENET121

9.1 DenseNet121

Dense Convolutional Network (DenseNet) is a convolutional neural network in which each layer is connected to all levels below it. It's a type of convolutional neural network that uses Dense Blocks to connect all layers (with matching feature-map sizes) directly, resulting in dense connections between them. Each layer gets fresh inputs from all preceding levels and passes on its own feature-maps to all following layers to maintain the feed-forward nature. DenseNets not only improve parameter efficiency, but they also make it easier to train networks. They can offer improved gradients and information flow throughout the network.

The DenseNet-121 model is a member of the DenseNet family designed to perform image classification. The models were first trained on Torch and subsequently converted to Caffe format by the authors. The ImageNet picture database was used to pre-train all DenseNet models [20].

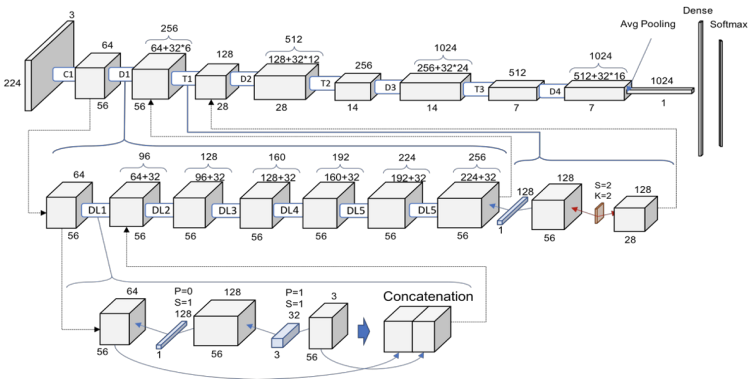


Figure 9.1: Full schematic representation of DenseNet121

In the above figure, we can see that these 32 new feature maps are added to the preceding volume by each layer. This is why we go from 64 to 256 after six levels. Transition Block also has 128 filters and may be used as a 1x1 convolution. After that, a 2x2 pooling with a stride of 2 was conducted, resulting in a half-division of the volume and number of feature maps. At the new deeper level representing the first Dense Layer within the first Dense Block, we can see how this behavior of adding 32 times the number of layers is achieved. To decrease the size of the feature maps, we do a 1x1 convolution with 128 filters, as advised by the authors, and then a more expensive 3x3 convolution (remember to include the padding to ensure the dimensions remain constant) using the 32 feature maps of growth rate. The input volume and the outcomes of the two operations (which are identical for every Dense Layer inside every Dense Block) are concatenated in order to add new information to the network's common knowledge.

9.2 Model Evaluation using DenseNet121

```
[epoch 10], [iter 100 / 1124], [train loss 0.19296], [train acc 0.92594]
[epoch 10], [iter 200 / 1124], [train loss 0.20307], [train acc 0.92297]
[epoch 10], [iter 300 / 1124], [train loss 0.19224], [train acc 0.92573]
[epoch 10], [iter 400 / 1124], [train loss 0.19027], [train acc 0.92672]
[epoch 10], [iter 500 / 1124], [train loss 0.18394], [train acc 0.92875]
[epoch 10], [iter 600 / 1124], [train loss 0.18693], [train acc 0.92745]
[epoch 10], [iter 700 / 1124], [train loss 0.18645], [train acc 0.92804]
[epoch 10], [iter 800 / 1124], [train loss 0.18328], [train acc 0.92953]
[epoch 10], [iter 900 / 1124], [train loss 0.18265], [train acc 0.92944]
[epoch 10], [iter 1000 / 1124], [train loss 0.18251], [train acc 0.92966]
[epoch 10], [iter 1100 / 1124], [train loss 0.18206], [train acc 0.92997]
------------------------------------------------------------
[epoch 10], [val loss 0.37241], [val acc 0.89262]
------------------------------------------------------------
```

Figure 9.2: Validation Accuracy of DenseNet121 (last 5 epochs)

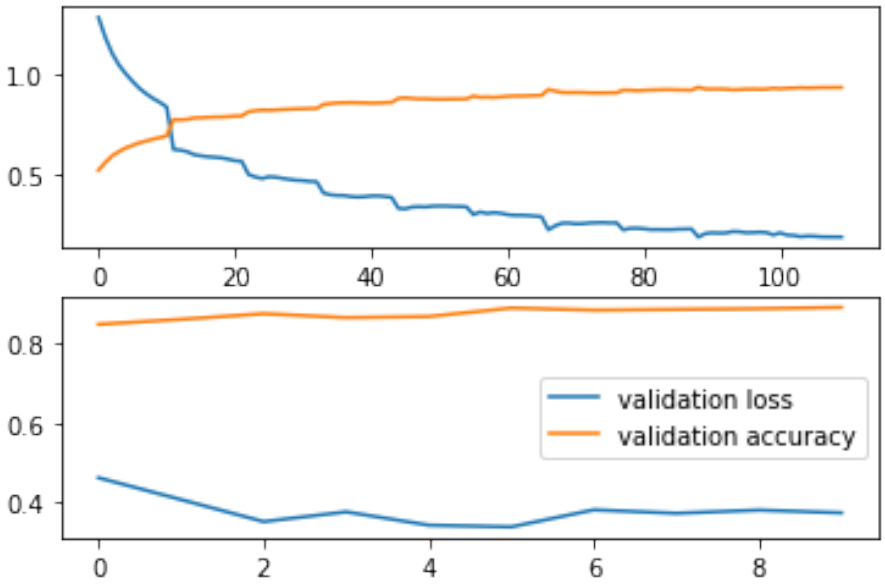


Figure 9.3: Loss and Accuracy of the Training and Validation Phase of the DenseNet121 Model

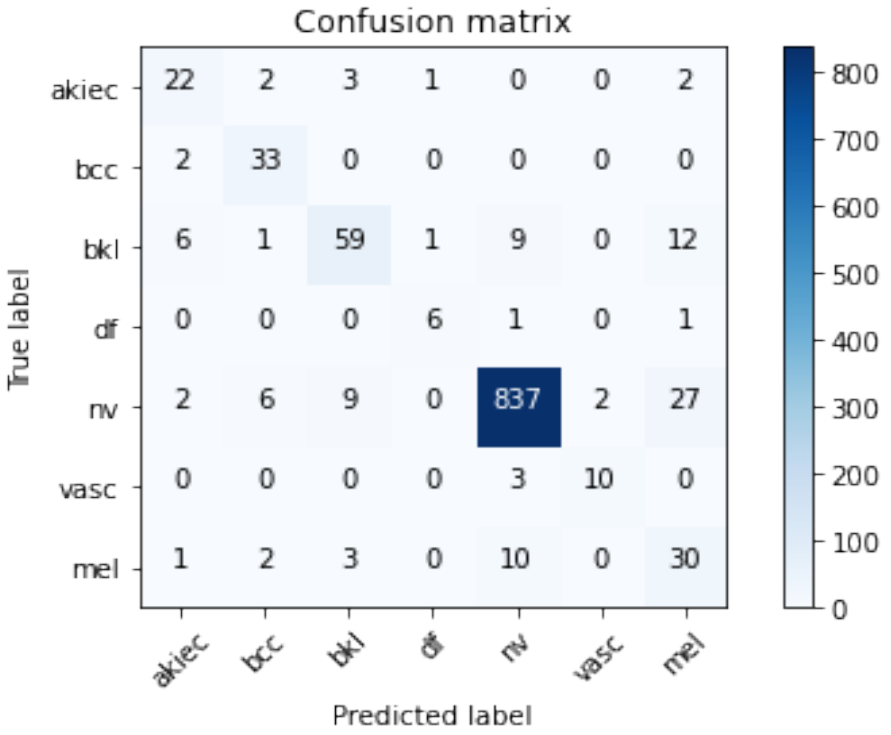


Figure 9.4: Confusion Matrix of DenseNet50 using Testing Data

	precision	recall	f1-score	support
akiec	0.67	0.73	0.70	30
bcc	0.75	0.94	0.84	35
bkl	0.80	0.67	0.73	88
df	0.75	0.75	0.75	8
nv	0.97	0.95	0.96	883
vasc	0.83	0.77	0.80	13
mel	0.42	0.65	0.51	46
accuracy			0.90	1103
macro avg	0.74	0.78	0.75	1103
weighted avg	0.92	0.90	0.91	1103

Figure 9.5: Precision, Recall, F1 score and Support values for DenseNet121

Chapter 10

Model Building and Evaluation using VGG11

10.1 VGG11

AlexNet was a breakthrough improvement when it was released in 2012; it improved on classic Convolutional Neural Networks (CNNs) and became one of the top image categorization models until VGG came along. While prior AlexNet derivatives focused on smaller window sizes and strides in the first convolutional layer, VGG tackles one of CNNs' most important features: depth.

VGGNet was created by the Visual Geometry Group (by Oxford University). At ILSVRC 2014, this architecture came in second place in the classification challenge, whereas GoogLeNet came in first. Because many modern image classification models are built on top of VGGNet, it's critical to understand it.

This is a model with weights that represent the dataset's characteristics that has been pre-trained on a dataset.

VGG models accept a 224×224 pixel picture as input, which must be in RGB format. If the reader is curious as to why just 224 out of the 0 to 255 pixel range of RGB was used to cope with a fixed image size.

Moving on to convolutional layers, VGG-11 contains three nodes that are fully linked. The upper two completely linked channels each have 4096 channels, as seen in the picture below. Because each channel corresponds to one class, the third completely linked layer contains 1000 channels.

There are 11 layers of VGG11:

1. Convolution using 64 filters + Max Pooling
2. Convolution using 128 filters + Max Pooling
3. Convolution using 256 filters
4. Convolution using 256 filters + Max Pooling
5. Convolution using 512 filters
6. Convolution using 512 filters + Max Pooling
7. Convolution using 512 filters
8. Convolution using 512 filters + Max Pooling
9. Fully connected with 4096 nodes
10. Fully connected with 4096 nodes
11. Output layer with Softmax activation with 1000 nodes

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 10.1: Layers of VGG11

The study's main result is that VGG-11 performs well with a large number of layers due to the modest size of the convolution filter. This explains the model's high accuracy and strong performance results [21].

10.2 Model Evaluation using VGG11

```
[epoch 10], [iter 100 / 1124], [train loss 0.34314], [train acc 0.87281]
[epoch 10], [iter 200 / 1124], [train loss 0.34449], [train acc 0.87406]
[epoch 10], [iter 300 / 1124], [train loss 0.35620], [train acc 0.86740]
[epoch 10], [iter 400 / 1124], [train loss 0.34587], [train acc 0.87109]
[epoch 10], [iter 500 / 1124], [train loss 0.34656], [train acc 0.87050]
[epoch 10], [iter 600 / 1124], [train loss 0.35261], [train acc 0.86875]
[epoch 10], [iter 700 / 1124], [train loss 0.35776], [train acc 0.86665]
[epoch 10], [iter 800 / 1124], [train loss 0.35538], [train acc 0.86770]
[epoch 10], [iter 900 / 1124], [train loss 0.35688], [train acc 0.86660]
[epoch 10], [iter 1000 / 1124], [train loss 0.35660], [train acc 0.86638]
[epoch 10], [iter 1100 / 1124], [train loss 0.35693], [train acc 0.86668]
------------------------------------------------------------
[epoch 10], [val loss 0.43513], [val acc 0.83345]
------------------------------------------------------------
```

Figure 10.2: Validation Accuracy of VGG11 (last 5 epochs)

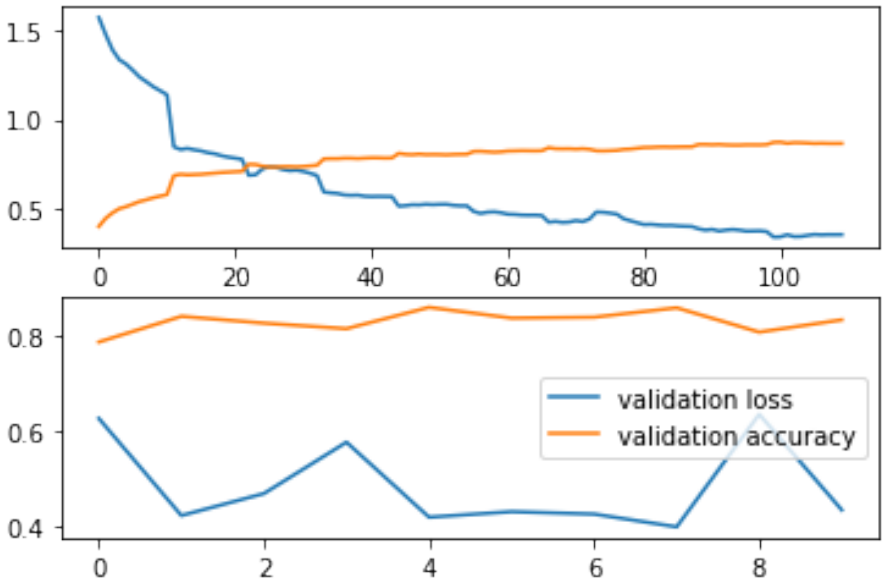


Figure 10.3: Loss and Accuracy of the Training and Validation Phase of the VGG11 Model

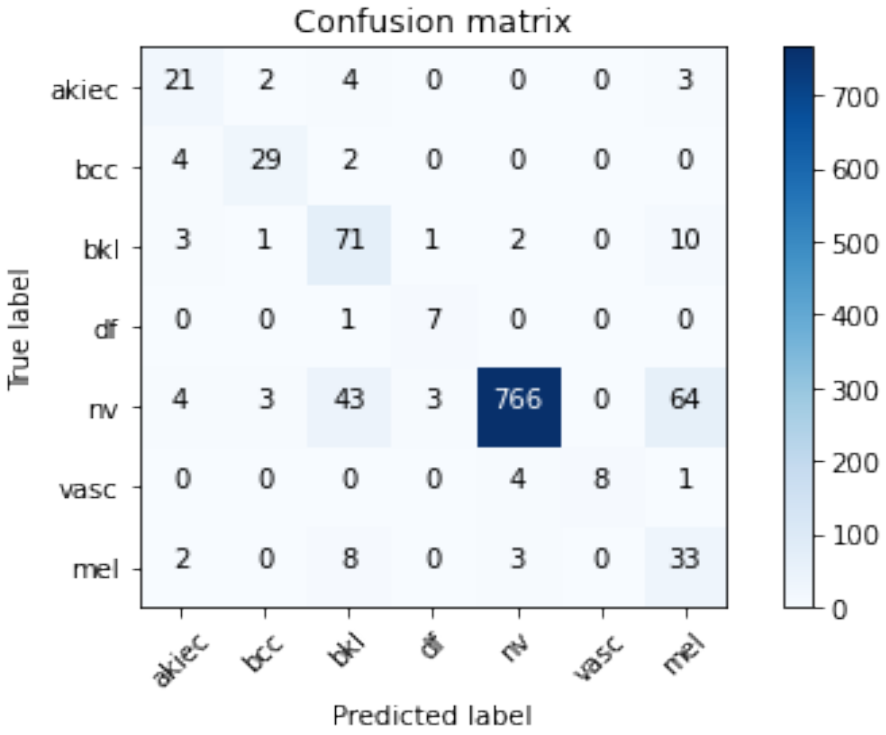


Figure 10.4: Confusion Matrix of VGG11 using Testing Data

	precision	recall	f1-score	support
akiec	0.62	0.70	0.66	30
bcc	0.83	0.83	0.83	35
bkl	0.55	0.81	0.65	88
df	0.64	0.88	0.74	8
nv	0.99	0.87	0.92	883
vasc	1.00	0.62	0.76	13
mel	0.30	0.72	0.42	46
accuracy			0.85	1103
macro avg	0.70	0.77	0.71	1103
weighted avg	0.91	0.85	0.87	1103

Figure 10.5: Precision, Recall, F1 score and Support values for VGG11

Chapter 11

Conclusion

A model of an Artificial Neural Network which is able to detect skin cancer from images with at least 90% accuracy, can have many use cases where the model can be used as an efficient method of skin cancer detection compared to the human eye. The purpose of our paper was to discuss the concept of detecting different forms of skin cancer from images by implementing a CNN model that uses Keras Sequential API. We began with a set kernel size for our convolution and a fixed filter size for our max-pooling function. The results were flattened into 1D single vectors and used as input variables for the ANN from Keras. The database we used contained 10,015 images, from which duplicate images were removed, and the remaining images split by an 80:20 ratio for the purposes of a test-train split.The training set was then subdivided into a 20% validation set.

Our model allowed us to achieve an accuracy of around 97%, where the model was run for 20 epochs. To compare our results with that of previously set models, we looked to other known algorithms, mainly ResNet50, DenseNet121 and VGG11. We ran the algorithms, with pre-trained data from ImageNet, through the three aforementioned algorithms (using 10 epochs) and recorded the accuracy of these compared to that achieved by our model. For comparison and evaluation, graphs of loss and accuracy, confusion matrices as well as precision, support and F1 scores of all the algorithms were recorded. We ascertained that the accuracy achieved by our model will allow it to be implemented in real-world applications in the future, with intended efficacy to aid in applicable fields to work alongside human efforts.

Bibliography

[1] *Skin cancer - symptoms and causes*, Dec. 2020. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/skin-cancer/symptoms-causes/syc-20377605.

[2] *Radiation: Ultraviolet (uv) radiation and skin cancer*, Oct. 2017. [Online]. Available: https://www.who.int/news-room/q-a-detail/radiation-ultraviolet-(uv)-radiation-and-skin-cancer.

[3] *Skin cancer survival rate*, Jan. 2018. [Online]. Available: http://moffittcloud.aviddesign.com/cancers/skin-cancer-nonmelanoma/survival-rate/.

[4] P. Grint, *A shortage of oncologists*, Nov. 2018. [Online]. Available: https://www.clearityfoundation.org/a-shortage-of-oncologists/.

[5] A. N. Hoshyar, A. Al-Jumaily, and A. N. Hoshyar, "The beneficial techniques in preprocessing step of skin cancer detection system comparing," *Procedia Computer Science*, vol. 42, pp. 25–31, 2014.

[6] H. T. Lau and A. Al-Jumaily, "Automatically early detection of skin cancer: Study based on nueral netwok classification," in *2009 International Conference of Soft Computing and Pattern Recognition*, IEEE, 2009, pp. 375–380.

[7] R. Seeja and A. Suresh, "Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (svm)," *Asian Pacific journal of cancer prevention: APJCP*, vol. 20, no. 5, p. 1555, 2019.

[8] M. Blot, M. Cord, and N. Thome, "Max-min convolutional neural networks for image classification," in *2016 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2016, pp. 3678–3682.

[9] V. Suárez-Paniagua and I. Segura-Bedmar, "Evaluation of pooling operations in convolutional architectures for drug-drug interaction extraction," *BMC bioinformatics*, vol. 19, no. 8, pp. 39–47, 2018.

[10] A. Estava, B. Kuprel, R. Novoa, J. Ko, S. Swetter, H. Blau, and S. Thrun, "Dermatologist level classification of skin cancer with deep neural networks [j]," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[11] H. Mhaske and D. Phalke, "Melanoma skin cancer detection and classification based on supervised and unsupervised learning," in *2013 International conference on Circuits, Controls and Communications (CCUBE)*, IEEE, 2013, pp. 1–5.

[12] F. Ercal, A. Chawla, W. V. Stoecker, H.-C. Lee, and R. H. Moss, "Neural network diagnosis of malignant melanoma from color images," *IEEE Transactions on biomedical engineering*, vol. 41, no. 9, pp. 837–845, 1994.

[13] S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "Applying artificial intelligence to the identification of variegated coloring in skin tumors," *IEEE engineering in medicine and biology magazine*, vol. 10, no. 4, pp. 57–62, 1991.

[14] D. Piccolo, A. Ferrari, K. Peris, R. Daidone, B. Ruggeri, and S. Chimenti, "Dermoscopic diagnosis by a trained clinician vs. a clinician with minimal dermoscopy training vs. computer-aided diagnosis of 341 pigmented skin lesions: A comparative study," *British Journal of Dermatology*, vol. 147, no. 3, pp. 481–486, 2002.

[15] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data*, vol. 5, no. 1, pp. 1–9, 2018.

[16] M. Binder, A. Steiner, M. Schwarz, S. Knollmayer, K. Wolff, and H. Pehamberger, "Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: A pilot study," *British Journal of Dermatology*, vol. 130, no. 4, pp. 460–465, 1994.

[17] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "Ph 2-a dermoscopic image database for research and benchmarking," in *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, IEEE, 2013, pp. 5437–5440.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[19] H. Mujtaba, *What is resnet or residual network: How resnet helps?* Oct. 2020. [Online]. Available: https://www.mygreatlearning.com/blog/resnet/.

[20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[21] S. Gupta, *Vgg-11 architecture*, May 2020. [Online]. Available: https://iq.opengenus.org/vgg-11/.