



Media Engineering and Technology Faculty
German University in Cairo

Skin Lesion Segmentation and Melanoma Detection using Deep Learning

Bachelor Thesis

Author: Youssef Shaarawy
Supervisors: Assoc. Prof. Mohammed Abdel-Megeed Salem
Submission Date: 1 August, 2021



Media Engineering and Technology Faculty
German University in Cairo

Skin Lesion Segmentation and Melanoma Detection using Deep Learning

Bachelor Thesis

Author: Youssef Shaarawy

Supervisors: Assoc. Prof. Mohammed Abdel-Megeed Salem

Submission Date: 1 August, 2021

This is to certify that:

- (i) the thesis comprises only my original work toward the Bachelor Degree
- (ii) due acknowledgment has been made in the text to all other material used

Youssef Shaarawy
1 August, 2021

Acknowledgments

I thank Assoc. Prof. Mohamed for his help and his valuable insight during this journey. His knowledge and advice were critical to the conduction of this research. I would also like to thank my friends and family for their support. Finally, I would like to thank my late mother, a dermatologist herself, who has always been and will always be my role model and hero.

Abstract

The rates of Melanoma have been rising rapidly over the past few decades. Detecting Melanoma early is critical to the reduction of the mortality rates associated with it. Despite several methods being available for segmentation of lesion region in images of skin diseases, there is still scope for exploring new models, which are efficient and provide better segmentation. Recently, researchers have been leaning more towards the use of Deep Learning algorithms for the diagnosis of Melanoma. The method proposed in this thesis consists of a primary stage where the image is cropped around the Region of Interest, eliminating objects that may interfere with the diagnosis using the Mask and Region based Mask_R CNN. The following stage involves the use of transfer learning to state-of-the-art networks which have been used in the classification of other objects, and using them to classify Melanoma. The three models explored are ResNet152V2, Xception and InceptionResNetV2. The first stage was trained and validated using the ISIC 2017 challenge dataset. The second stage was trained, validated and tested using the ISIC 2018 challenge dataset. The model achieves an accuracy of 89.2% on the test dataset, while achieving a specificity of 91.0% and sensitivity of 80.8%. Achieving a sensitivity of 80% and higher, while having a specificity of above 90%, shows that the model is highly efficient and differentiates accurately between the two classes without being biased towards any of them.

Contents

1	Introduction	1
1.1	Motivation	2
1.2	Problem Statement	2
2	Background	4
2.1	Concepts Overview	4
2.1.1	Deep Learning	4
2.1.2	Neural Networks	5
2.1.3	Convolutional Neural Networks	6
2.1.4	Mask-R CNN	8
2.1.5	Transfer Learning	8
2.2	Literature review	9
2.2.1	Non-Deep Learning Algorithms	9
2.2.2	Deep Learning Algorithms	12
2.3	Datasets per paper	14
3	Methodology	15
3.1	Segmentation of skin lesions	16
3.2	Detection of Melanoma	18
4	Results	23
4.1	Evaluation Metrics	23
4.2	Results	24
4.2.1	Initial Results	24
4.2.2	Further Experimentation	25
4.3	Experiment setting and Dataset	29
4.3.1	Tools used	29
4.3.2	Dataset	30
5	Conclusion and Future Work	31
5.1	Conclusion	31
5.2	Future Work	32
	References	35

List of Figures

1.1	Skin Cancer Classes	2
2.1	Hierarchy of Artificial Intelligence, Machine Learning and Deep Learning[1]	4
2.2	Machine Learning vs Deep Learning[2]	5
2.3	Neural Network[3]	6
2.4	Convolution[4]	7
2.5	Edge Detection using Convolution	7
2.6	R-CNN workflow[5]	8
2.7	Transfer Learning	9
3.1	Block diagram	15
3.2	Ground Truth mask of a skin lesion	16
3.3	17
3.4	Ink marks filtered out	17
3.5	ResNet Architecture[6]	19
3.6	Xception Architecture[7]	19
3.7	InceptionResNetV2 Architecture[8]	20
3.8	Flowchart of the classification network	22
4.1	Confusion Matrices for all 3 models	25
4.2	Confusion Matrices for the five new models	27
4.3	Final Confusion Matrices	28

Chapter 1

Introduction

Skin diseases are ranked as the fourth most common cause of human illness [9]. According to the Skin Cancer Foundation, 1 in 5 Americans will develop skin cancer by the age of 70. In addition, more than 2 people die of skin cancer in the U.S. every hour. Skin cancer, the most common human malignancy[10], is primarily diagnosed visually, beginning with an initial clinical screening and followed potentially by dermoscopic analysis, a biopsy and histopathological examination. Skin cancers can be classified into two classes; Melanoma and non-Melanoma cancers. Melanoma, the most serious form of skin cancer, is characterized by the uncontrolled growth of pigment-producing cells that control skin colour. What differentiates Melanoma from non-Melanoma skin cancers is that Melanoma is likely to spread to other parts of the body and is known to be an aggressive cancer.

The first five letters of the alphabet are a guide to help recognize the warning signs of melanoma.[11]

- **A** is for Asymmetry. Most melanomas are asymmetrical. If you draw a line through the middle of the lesion, the two halves don't match, so it looks different from a round to oval and symmetrical common mole.
- **B** is for Border. Melanoma borders tend to be uneven and may have scalloped or notched edges, while common moles tend to have smoother, more even borders.
- **C** is for Color. Multiple colors are a warning sign. While benign moles are usually a single shade of brown, a melanoma may have different shades of brown, tan or black. As it grows, the colors red, white or blue may also appear.
- **D** is for Diameter or Dark. While it's ideal to detect a melanoma when it is small, it's a warning sign if a lesion is the size of a pencil eraser (about 6 mm, or $\frac{1}{4}$ inch in diameter) or larger. Some experts say it is also important to look for any lesion, no matter what size, that is darker than others. Rare, amelanotic melanomas are colorless.

- **E** is for Evolving. Any change in size, shape, color or elevation of a spot on your skin, or any new symptom in it, such as bleeding, itching or crusting, may be a warning sign of melanoma.

In this paper, we propose a method for classification of Melanoma using CNNs (Convolutional Neural Networks), which have been used previously in the segmentation, classification and detection of several diseases accurately.

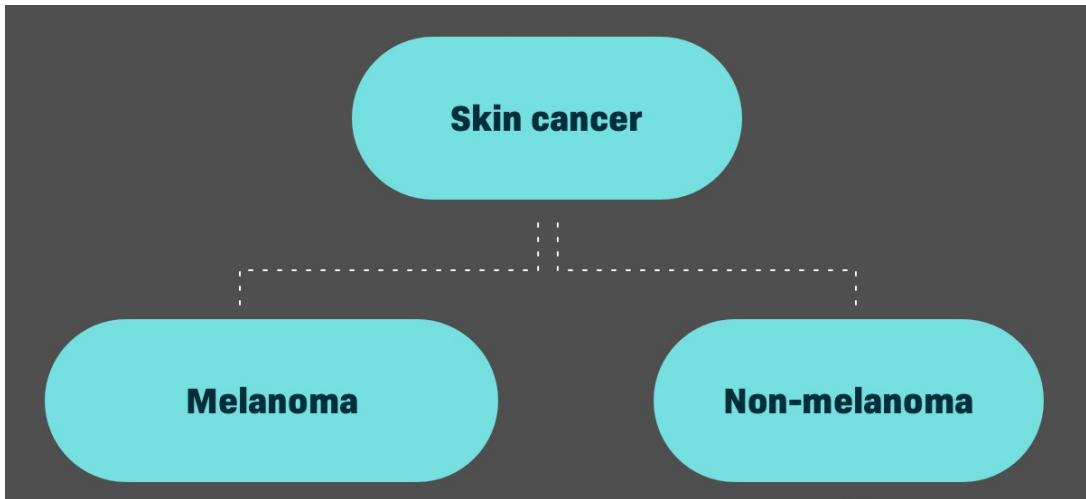


Figure 1.1: Skin Cancer Classes

1.1 Motivation

As stated before, Melanoma is highly aggressive. It can spread to other parts of the body, if not treated early, and cause death. The good news is that 99 percent of all cases are curable if they are diagnosed and treated early enough[12]. Our goal is to introduce an accurate and easily-accessible system which can detect Melanoma as early as possible.

1.2 Problem Statement

The algorithm has to provide results accurate enough for it to be safe to use. Classification of dermatoscopic images' difficulties are: the presence of hairs, inks, ruler markings, colored patches, glimmers of light, drops, oil bubbles, blood vessels and/or inflammation around the lesion.

- **Concept Overview:** A brief explanation of all technologies used throughout the paper.

- **Literature Review:** A compilation of all previous work done in the past regarding the implemented project or any of the technologies used.
- **Methodology:** Workflow overview on how the project was implemented over two parts and how the environment developed could be replicated.
- **Results:** The section is split into parts. The first part shows the initial results and the final results after further experimentation. The second part lists the Tools and Datasets used.
- **Conclusion:** A brief discussion of what was achieved in this paper, and the future work that could be done to excel the results and further development.

Chapter 2

Background

2.1 Concepts Overview

2.1.1 Deep Learning

Deep Learning, a subset of machine learning, is a function based on Artificial Intelligence. It is based on how human brains process information, in order to create patterns used for decision making. It has networks that can learn from data that is not structured or labeled without being supervised. Object recognition, speech recognition, and language translation are some of the common problems tackled by Deep Learning. The difference between Machine Learning and Deep learning is that Deep Learning networks extract the features without supervision as highlighted in figure 2.2, which is why it usually requires a large dataset.

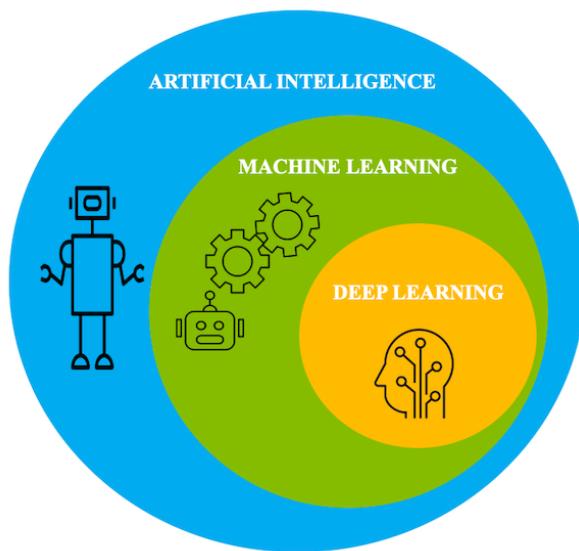


Figure 2.1: Hierarchy of Artificial Intelligence, Machine Learning and Deep Learning[1]

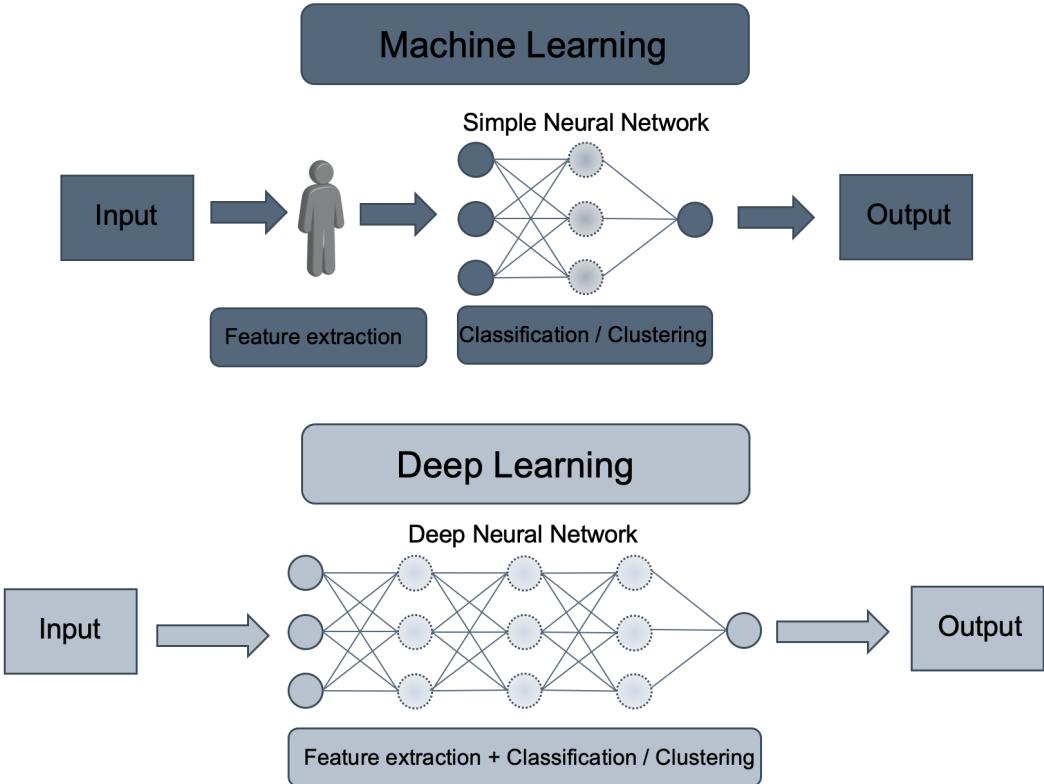


Figure 2.2: Machine Learning vs Deep Learning[2]

2.1.2 Neural Networks

Artificial Neural Networks (ANNs), or just simply Neural Networks (NNs), are a collection of nodes in a graph called neurons, based on the neurons of a human brain. Each connection, called edges, can send signals containing data, specifically a real number to other neurons just like the synapses of a human brain. The output of a neuron is calculated by a non-linear function of the sum of its inputs. Neurons and edges have weights, that are adjusted as the network is trained. This weight increases the strength of the signal at a connection, and impacts how much this signal is taken into consideration at the next neuron.

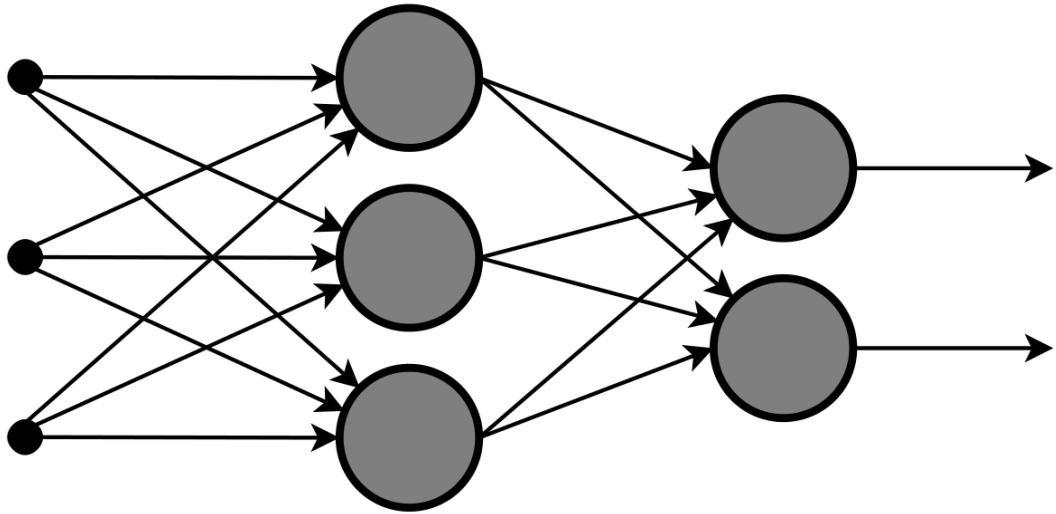


Figure 2.3: Neural Network[3]

2.1.3 Convolutional Neural Networks

CNNs (Convolutional Neural Networks) are used for Image Analysis. Each layer in the network can be used to detect specific patterns (edges, circles, corners, etc). There are many types of layers like convolutional layers, pooling layers and fully connected layers. Each convolutional layer has a matrix that acts as a filter. The dot product of the $f \times f$ matrix and each group of $f \times f$ pixels in the input is calculated and stored in the output as shown in figure 2.4, which is then passed on to the next layer. This operation, known as Convolution, hence where the name of these types of networks comes from, results in the output being of smaller size than the input, which can restrict the number of layers we can use. Therefore, padding the input before "convolving" it can be useful. This also increases the weight of the values at the edges of the input image in the output. Pooling layers group $f \times f$ pixels together and apply a specific function on them (Max, Min, Average, etc).

2.1. CONCEPTS OVERVIEW

7

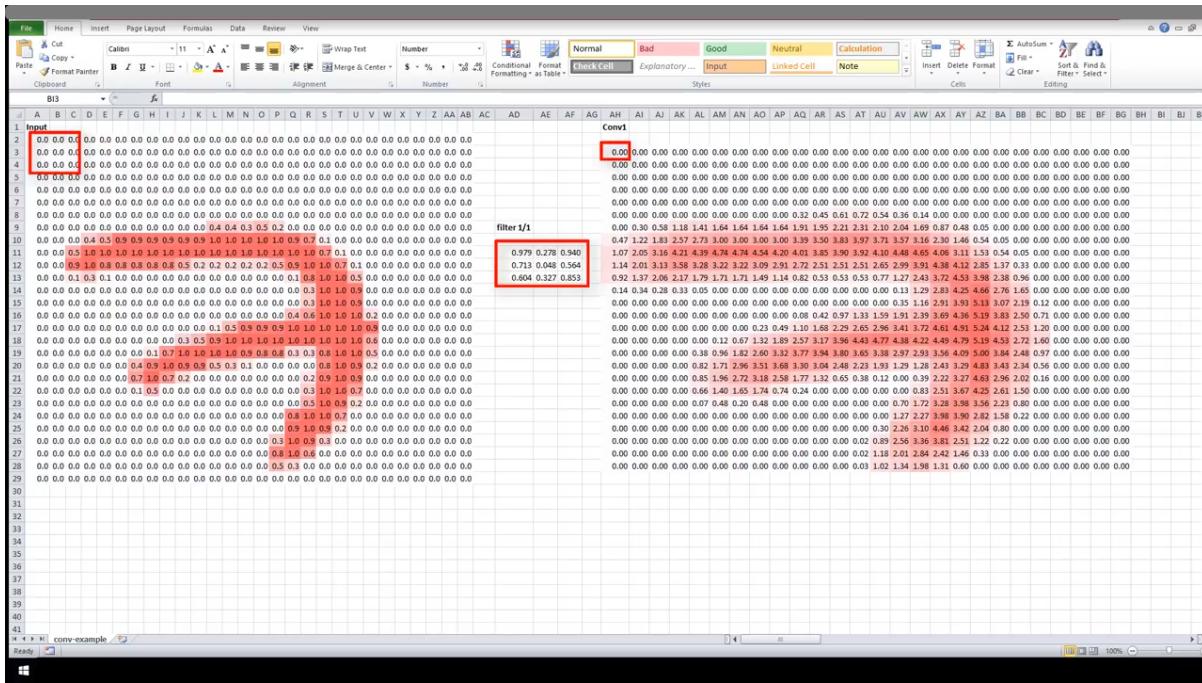


Figure 2.4: Convolution[4]

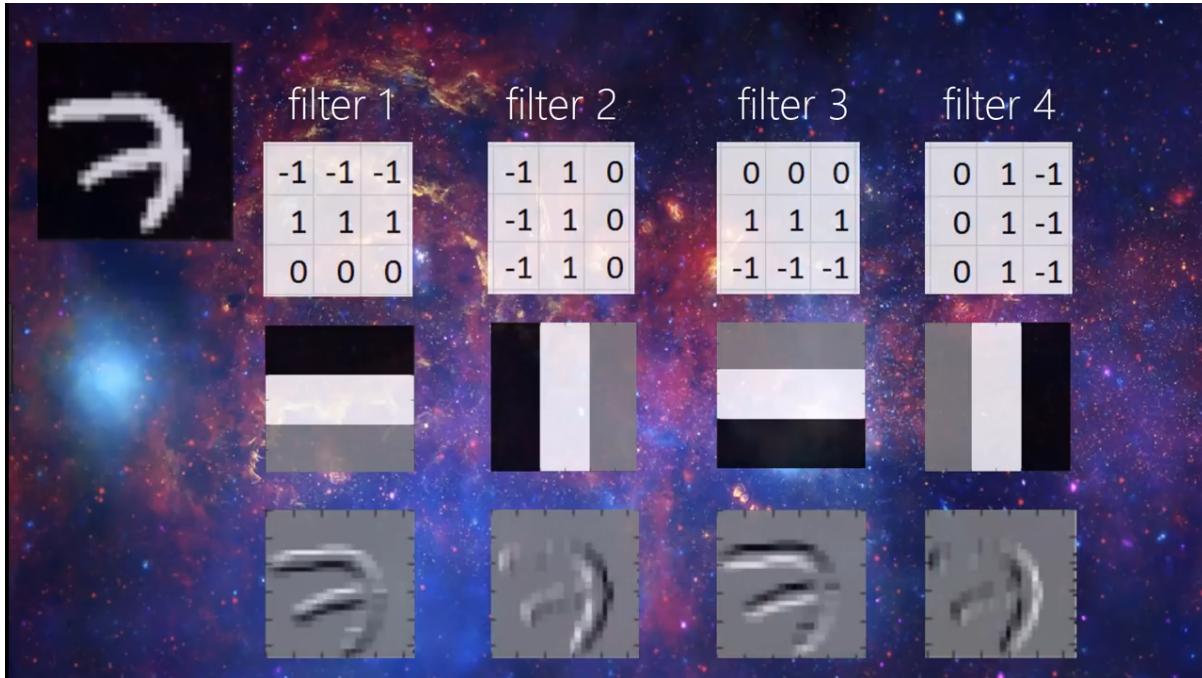


Figure 2.5: Edge Detection using Convolution

2.1.4 Mask-R CNN

R-CNN, which stands for Region based Convolutional Neural Network, is a family of networks used for computer vision tasks, such as object detection. Figure 2.6 shows how R-CNN works. Mask R-CNN is a Convolutional Neural Network that was developed on top of Faster R-CNN, which outputs a class, bounding box and mask for each detected object.

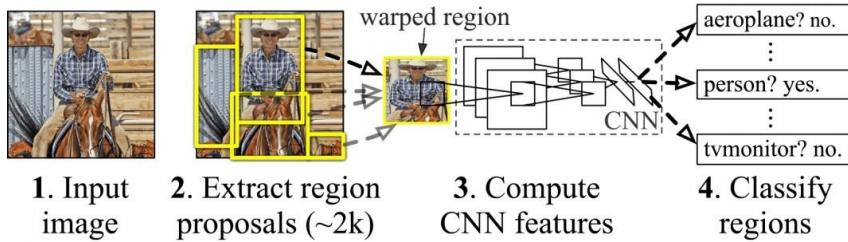


Figure 2.6: R-CNN workflow[5]

2.1.5 Transfer Learning

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. The concept is used when the dataset is not big enough to train a network from scratch. The model is imported with its pre-trained weights, but without the top layer in the network, which is the classification layer. We replace the removed layer by our own classifier. The network is trained with our dataset at first while the base model layers are frozen (their states and values can not change) and then these layers are unfrozen and the whole network is trained again using our dataset for fine-tuning.

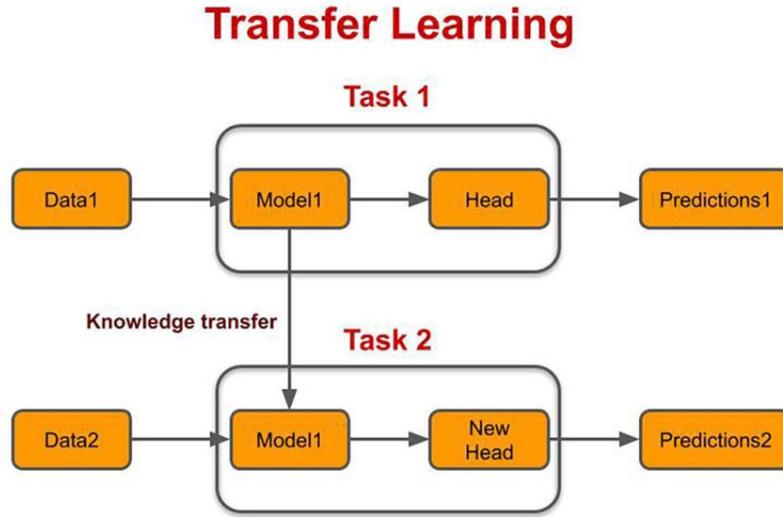


Figure 2.7: Transfer Learning

2.2 Literature review

In the recent past various methods have been proposed for segmentation and feature extraction from lesion regions. Some papers proposed methods based on features derived from pixel color to drive the segmentation, simple thresholding [13], active contours[14], or region merging[15]. However, these algorithms struggle during segmentation of lesions that lack contrast with the surrounding skin or lesions which have fuzzy edges. Recently, most methods focus on deep learning algorithms, especially CNNs due to the increasing availability of datasets required to train the networks enough for them to produce accurate results and for the increasing number of already highly trained networks. Also, CNNs have produced outstanding results in the classification of images and are now being increasingly used in the medical field.

2.2.1 Non-Deep Learning Algorithms

Green et al.[16] proposed a system, consisting of a hand-held device incorporating a color video camera and color frame grabber mounted in a microcomputer, in a pigmented lesion clinic. Analysis software extracted features relevant to the size, color, shape, and boundary of each lesion, and these features were correlated with the characteristics of a skin tumor, which were mentioned in chapter 1. Sixteen of 18 melanomas, and 89% of pigmented lesions overall, were correctly classified by the image analysis system over a 20 month period. All were captured by the camera in the system.

C. Sagar and L. M. Saini[17] proposed a simple and efficient method for the automatic segmentation of clinical images using colorspace analysis and improved binary thresholding algorithm. The method first does some pre-processing of the images to fix the illumination. Then, extra steps are taken to remove hairs or other objects that are irrelevant. The process of spatial clustering by thresholding using Otsu's method is then used, which differentiates the grayscale image into a binary image with each pixel of the image assigned a value either 0 or 1, depending on the grayscale pixel values. Two clusters of pixels are formed corresponding to 0 as lesion and 1 as background. The accuracy of the algorithm used is approximately 93.71% in detecting the correct lesion from an input digital image.

Celeb et al.[15] presented a quick and unsupervised approach to border detection of skin lesions based on the SRM algorithm. It consists of three main stages: Pre-processing, segmentation and Post-processing. Pre-processing the image included the removal of black frames by calculating the lightness of each pixel, and pixels with low lightness were discarded. The segmentation phase includes pixel couple ordering and region merging. Finally, the post-processing phase removes regions which belonged to background skin, removing the isolated regions, combining the remaining regions, and expanding the initial border to obtain the final result. The results were compared with three sets of borders that were determined by dermatologists acting as the ground truth. The approach achieved quick and accurate results.

Erkol et al.[14] investigated gradient vector flow (GVF)snakes to find the border of skin lesions in dermoscopy images. Snakes can be defined as curves within an image domain that can move under the influence of internal forces defined within the curve, from features such as smooth-ness. GVF fields are computed through a diffusion process. Preprocessing of the dermoscopy images is performed in two steps, including automatic snake point initialization and skin lesion image preprocessing, to facilitate application of the GVF snake algorithm. The experimental results show that 76 of 100 images have percentage border errors less than 20% and that 96 of 100 images have errors less than 30%.

Aljanabi et al.[18] proposed a new approach based on the artificial bee colony (ABC) algorithm. The advantages of this method is that it is much simpler than other methods and requires fewer parameters. Also, it requires no pre-training. The ABC algorithm was used to find the optimum threshold value for the melanoma segmentation, which was used for the melanoma segmentation like the thresholding method used in the Otsu method. For the melanoma detection, the method achieved an average accuracy and Jaccard's coefficient in the range of 95.24–97.61%, and 83.56–85.25%

Ganster et al.[19] introduced a system that segments skin lesions and is responsible for the detection of melanoma from images obtained from Epiluminescence microscopy (ELM). For the segmentation of skin lesions, the fusion of several simple segmentation algorithms (thresholding, color clustering) was used. The features used in the system are based on the famous ABCD rule used by dermatologists to segment lesions. The algorithm was successful in the segmentation of more than 96% lesion images. A non-parametric 24-NN classifier was used to detect Melanoma. A classification percentage of

73% was achieved which is comparable to rates observed in clinical routine and depicts a good performance for an automated system, which was trained on a dataset obtained from a daily clinical routine and was not optimized specially for this experiment.

Møllersen et al.[13] presented a segmentation algorithm using density analysis. Only the background skin is detected by density analysis and then the location of the lesion is estimated. The algorithm starts off by correcting the illumination of the image by creating a binary image from the luminance component in the IYQ colour space. Then, it removes hairs present in the image by detecting them using a morphological operation on the binary image. The test sample contained 91 images which were evaluated by dermatologists. They evaluated the segmentation of the algorithm as 'good' in 85% of the images. However, the algorithm struggled significantly to segment lesions which lacked contrast with the surrounding skin.

Peruch et al.[20] presented a new thresholding technique for segmentation of lesions called Mimicking Expert Dermatologists' Segmentations (MEDS) that is based on the thinking process of dermatologists when detecting lesions. It is a fast algorithm that doesn't require great computational resources to execute. The technique operates on the color histogram of the image, that associates to each color c the number of pixels $h(c)$ of that color. The colors (and thus pixels) are then separated into two clusters corresponding, respectively, to lesional and non-lesional skin. The images are then down-sampled to increase the efficiency of the algorithm. The results of the algorithm are compared with segmentations produced by four different expert dermatologists. MEDS obtained, on average, 12.35% disagreement with the expert dermatologists.

Melli et al.[21] compared the most famous color clustering techniques at the time, which were median cut, k-means, fuzzy-c means, and mean shift by applying them to a segmentation technique, and then comparing the upper bound in accuracy that can be reached with each approach. The main objective was to separate pixels into two classes, lesion and non-lesion. The approach did not only rely on the colors of the pixels, but also on their positions. The algorithm assumes that lesions occupy the center part of the image so they use the color of the skin in the corners of the image as basis for non-lesion skin color. The algorithm achieved accuracies higher than 93% on the dataset they used.

Cavalcanti et al.[22] explore how the concentration of the two main classes of melanin present in human skin, eumelanin and pheomelanin, can potentially be used to identify melanocytic skin lesions as benign or malignant, when combined with current diagnostic methods. Their method classifies lesions based on the estimation of eumelanin and pheomelanin contents in melanocytic skin lesions, uses these features in a two-stage classification technique. The first stage is performed, using the 52 features based on the ABCD rule and assigns a benign/malignant label to the lesion. In the second stage, all images that have been initially classified as benign are re-classified now using the melanin variation features extracted from inside the lesion and from the healthy skin area surrounding the lesion. The algorithm achieved a sensitivity of 100%, specificity of 97.78% and accuracy of 99.34% on the dataset they used.

2.2.2 Deep Learning Algorithms

Acosta et al.[23] discussed the use of Deep Learning based on Convolutional Neural Networks, which have been widely used for segmentation, classification and detection of several diseases. The method used in this paper first automatically crops the region of interest containing lesions using the Mask and Region based Convolutional Neural Network technique, removing objects such as hairs or ink marks that can interfere with the classification accuracy of the model, and a second stage based on a ResNet152 structure, which classifies lesions as either “benign” or “malignant”. On the test data set, the proposed model achieves an increase in accuracy and balanced accuracy of 3.66% and 9.96%, respectively, with respect to the best accuracy and the best sensitivity/specificity ratio reported in the ISIC 2017 challenge for melanoma detection, which shows that the model is good for accurate discrimination between benign and malignant lesions.

Venkatesh et al.[24] proposed an automatic approach to skin lesion region segmentation based on a deep learning architecture with multi-scale residual connections. The architecture of the proposed model is based on UNet [25] with residual connections to maximise the learning capability and performance of the network. From the results, the boundaries of lesion regions and the background are well separated and well-differentiated. Furthermore, the proposed model uses around 16M parameters only which is a small number when compared to other well-known conventional deep architectures used in complex problems.

Esteva et al.[26] demonstrated the classification of skin lesions using a single CNN, trained end-to-end from images directly, using only pixels and disease labels as inputs. A CNN is trained using a data-set of 129,450 clinical images, consisting of 2,032 different diseases. The performance is tested against 21 board-certified dermatologists on biopsy-proven clinical images. The CNN achieves performance on par with all tested experts across both tasks, showing that the model is capable of performing similarly to expert dermatologists when it comes to classification of skin cancer.

A. A. Adegun and S. Viriri[27] proposed a novel CNN based approach with an enhanced deep supervised encoder-decoder network to extract strong and robust features of skin lesions images. This network is able to deduce sophisticated features from the lesion images Due to its multi-stage approach in which the encoder stage of the network learns the general appearance of a lesion while the decoder stage learns the characteristics of a boundary of a lesion. After extracting the features, they use a new method called ”Lesion-classifier” to perform the classify skin lesions into melanoma and non-melanoma in a pixel-wise manner. This was tested on 200 skin lesion images. It achieved accuracy and dice coefficient of 95% and 92% respectively

Nasr-Esfahani et al.[28] introduced a system which takes clinical images as inputs, which could contain illumination and noise effects, and pre-processes them to reduce the effect of such artifacts. Afterward, the enhanced images are fed to a pre-trained convolutional neural network and applies the concept of transfer learning. The CNN then distinguishes between melanoma and non-melanoma cases. For training, only a

small dataset was available. By cropping, scaling, and rotating of images the number of images was increased to train the network using a bigger dataset in order to produce more accurate results. The system achieved an accuracy of 81%.

Yu et al.[29] used a CNN to detect acral melanoma (AM), which is a rare disease in Asians, from the dermoscopy images of pigmentation caused by the lesions on the hands and feet of the patients. As a result of their dataset being insufficient, they adopted a transfer learning technique to use learned features from a CNN model pre-trained on a large-scale natural image dataset. Then, the accuracy of diagnosis was calculated comparing it with a dermatologist's and a non-expert's evaluation. The accuracies achieved by the convolutional neural network were 83.51% and 80.23%, which was higher than the non-expert's evaluation (67.84%, 62.71%) and close to that of the expert (81.08%, 81.64%).

X. Zhang[30] used an improved neural network framework to achieve effective feature learning, and segment melanoma images efficiently. The architecture of the network includes multiple convolution layers, dropout layers, softmax layers, multiple filters, and activation functions. To overcome the insufficiency of the dataset, the number of images was increased by applying various transformations to the dataset. A non-linear activation function (such as ReLU and ELU) is used to overcome gradient disappearance and improve accuracy. The proposed model achieved an accuracy of 91%, which was higher than the other models it was compared to in the paper.

Attia et al.[31] proposed using a hybrid method that utilizes two popular deep learning methods: convolutional and recurrent neural networks. They used convolutional neural networks with auto encoder-decoder architectures such as FCN and SegNet, which outputted coarse segmentation masks. To overcome this problem, a recurrent neural network (RNN) is trained to form contextual connections between pixels. RNNs are used to preserve local and global contextual dependencies even over large distances between pixels. They achieved a segmentation average accuracy of 0.98 and Jaccard index of 0.93.

L. T. Thao and N. H. Quang[32] presented deep learning based approaches to solve two problems in skin lesion analysis using a dermoscopic image containing skin tumor, segmentation of a lesion and classification of it. The first step is using a convolutional deconvolutional simple network for the segmentation of skin lesions. The second step is done using two different approaches: 1.Using a simple convolutional network trained by the dataset from scratch 2.Using a convolutional network based on the VGG-16 architecture adjusted using transfer learning. Both had the same goal: Differentiating Melanoma lesions from non-Melanoma lesions. The first step achieved an 87% accuracy. For the second step, the first approach achieved an accuracy of 49% while the second approach achieved an accuracy of 80%.

Al-Masni et al.[33] introduced a new full resolution convolutional network (FrCN) method for skin lesion segmentation. As opposed to other deep learning based segmentation architectures, the proposed model was able to produce full spatial resolution features for each pixel of the input dermoscopy images. As a result, each pixel was considered as a training sample for the model, which removed the need for pre- or post-processing techniques for artifact removal or low contrast adjustment and improved the performance

of pixel-wise segmentation. The algorithm produced adequate results with an average Jaccard index of 77.11% and an overall segmentation accuracy of 94.03% on the ISIC challenge dataset used.

2.3 Datasets per paper

The following table presents the dataset used for each of the papers reviewed in section 2.2.

Paper	Dataset
Yu et al	Yonsei University Health System and Keimyung University Health System
Nasr-Esfahani et al	UMCG
Green et al	Live patients
Peruch et al	Live patients
Adegun and S. Viriri	ISIC 2017 and PH2
Al-Masni et al	ISIC 2017 and PH2
Zhang	ISIC 2017
Venkatesh et al	ISIC 2017
Acosta et al	ISIC 2017
L. T. Thao and N. H. Quang	ISIC 2017
Aljanabi et al	ISIC 2016, 2017, PH2 and Dermis
Attia et al	ISIC 2016
Celebi et al	EDRA Interactive and dermatologists
Erkol et al	EDRA interactive
C. Sagar and L. M. Saini	Dermnet
Cavalcanti et al	Dermnet
Møllersen et al	Dermatologist
Ganster et al	Department of Dermatology at the Vienna General Hospital.
Esteva et al	18 clinicians, Stanford University Medical Center, ISIC, the Edinburgh Dermofit Library
Melli et al	-

Table 2.1: Dataset per Paper

Chapter 3

Methodology

We propose a new approach for the segmentation of skin lesions and the detection of Melanoma. The approach is inspired by the work of Acosta et al[23] and L. T. Thao and N. H. Quang[32], but with improvements on the algorithm they used. It is built on splitting the algorithm into two parts. The first part is the detection of the skin lesion in the image, and then drawing a bounding box around it using the Mask and Region based CNN called Mask R-CNN [34]. The image is then cropped at the bounding box coordinates in order to produce a clearer zoomed-in image of the lesion. The second part is the classification of the skin lesion, whether it is Malignant or Benign using several state-of-the-art CNNs adjusted to our dataset using transfer learning.

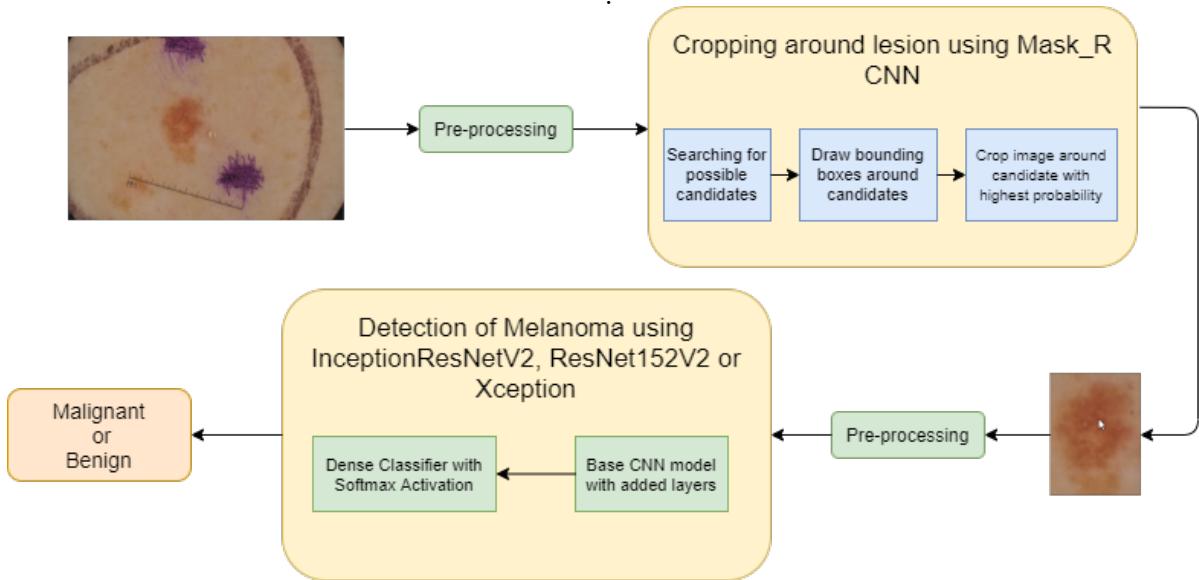


Figure 3.1: Block diagram

3.1 Segmentation of skin lesions

For this part, the ISIC 2018 challenge dataset was used, as it was the most recent and largest ISIC dataset containing ground truth masks of each lesion at the time of the study. It consists of 2594 training images, 1000 test images and 100 validation images of lesions, with their corresponding ground truth masks. The ground truth masks are binary images, with the segmented lesions having white pixels and non-lesion parts of the skin having black pixels, as shown in figure 3.2.

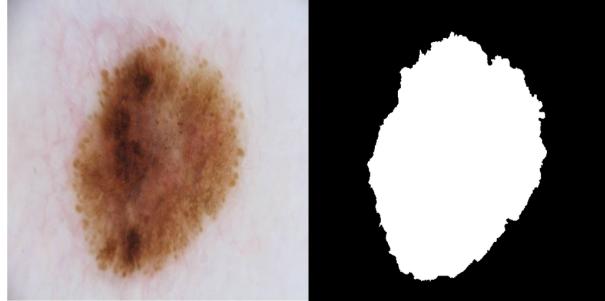


Figure 3.2: Ground Truth mask of a skin lesion

Mask R-CNN is trained using the 2594 images and their corresponding ground truths, to be able to understand what a lesion looks like. After training the model, it is used to predict on the test data set where the lesion is. When making a prediction, it searches the image for candidates that may correspond to a skin lesion, draws a bounding box around each one and, then, generates the probability of each candidate being a lesion. As shown in figure 3.3a, the network detected two candidates, the outer box having a probability of 0.904 and the inner box having a probability of 0.99. The candidate having the highest probability in the image is selected and the picture is cropped using the bounding box generated around it, as in figure 3.3b. This is done in order to crop a bounding box around the Region of Interest (ROI) and filter out objects which may interfere with the classification of the lesion such as ink marks and hairs, such as in figure 3.4.

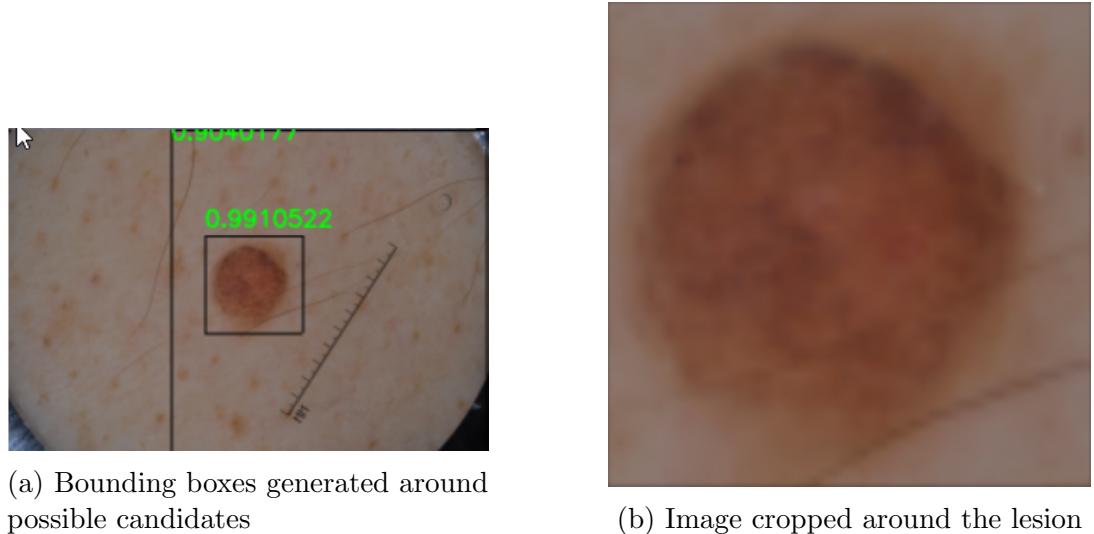


Figure 3.3

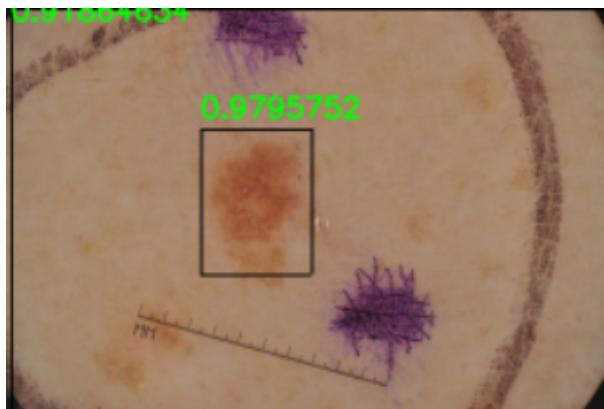


Figure 3.4: Ink marks filtered out

For training, all images were resized to 512 x 512 and data augmentation techniques, such as flipping horizontally and vertically and rotation, were applied in order to make the dataset bigger. The model was pre-trained on The MS COCO (Microsoft Common Objects in Context) dataset, which is a large-scale object detection, segmentation, key-point detection, and captioning dataset. The dataset consists of 328,000 images. All layers in the model were trained, which is different from what was done in section 3.2. The model is trained for 50 epochs primarily at a learning rate of 0.001, and then for further improvement and fine-tuning, the model was re-trained for another 50 epochs at a much lower learning rate of 0.0001.

After confirming that the model produced satisfactory results on the test images by correctly identifying the lesions and producing clear images cropped around them, the model was used to crop the images in the dataset that was used for part two of the algorithm as described in the next section, and thus, resulting in a cascaded dataset.

3.2 Detection of Melanoma

Not only can Mask R-CNN be used for the segmentation of the skin lesions, but it can also be used for the detection of Melanoma. However, several state-of-the-art architectures have been presented since the introduction of Mask R-CNN, that have proven more efficient and produced better results in the classification of Melanoma. In this section, the dataset of the 2019 ISIC challenge was used, as it was the most recent and largest dataset in the ISIC archive containing lesion images and their diagnosis at the time this study was conducted. It consists of 25,331 training images, composed of 4,522 Melanoma lesions and 20,809 non-Melanoma lesions, with one ground truth file stating the diagnosis of each of these images (whether it was a Melanoma lesion or not). At the time this study was conducted, the ISIC hadn't released the ground truth file for the test images. Therefore, only the training images were used in order to be able to evaluate the performance of the used networks. The images were split randomly into training, validation and test datasets, with each one having a percentage of 80%, 10% and 10% of the images respectively, while keeping the ratio between Melanoma and non-Melanoma images constant in all 3 datasets. This produced 20,264 training images, 2,533 validation images and 2,534 test images.

Three different state-of-the-art networks were used for the second part, which have produced significant results in several other classification problems. The three networks are: ResNet152V2, Xception and InceptionResNetV2. Their number of parameters, top-1 and top-5 accuracies, which refer to their performance on the ImageNet validation dataset are shown in table 3.1. The architectures of the three models are shown below in figures 3.5, 3.6 and 3.7.

Model	Top-1 Accuracy	Top-5 Accuracy	Parameters
ResNet152V2	78.0%	94.2%	60,380,648
Xception	79.0%	94.5%	22,910,480
InceptionResNetV2	80.3%	95.3%	55,873,736

Table 3.1

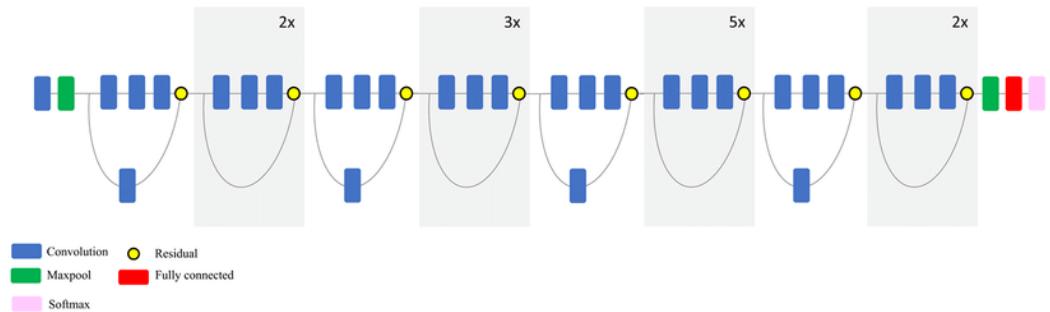


Figure 3.5: ResNet Architecture[6]

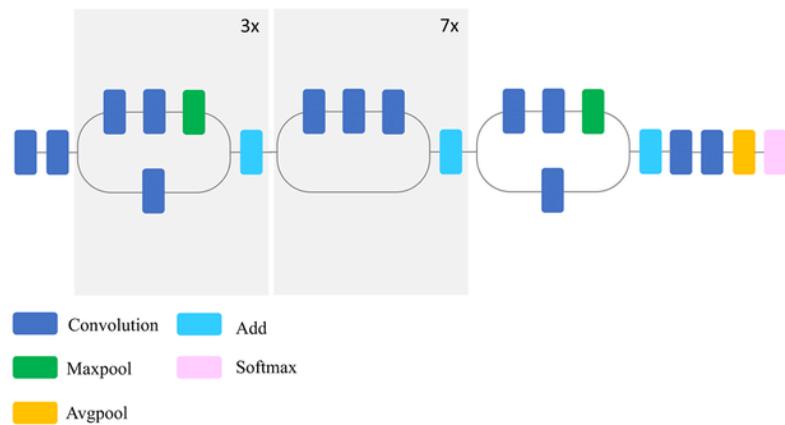


Figure 3.6: Xception Architecture[7]

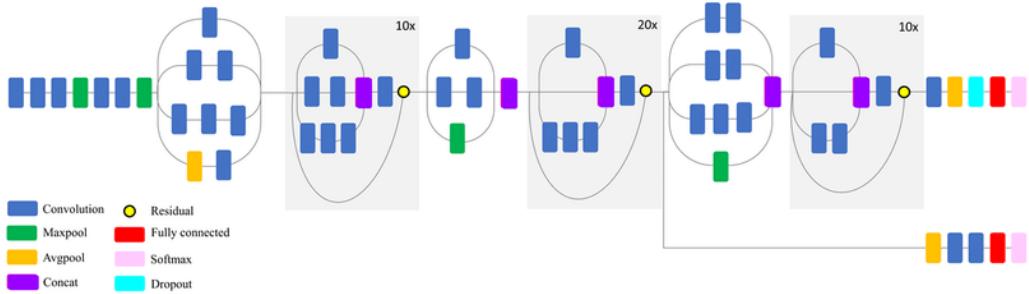


Figure 3.7: InceptionResNetV2 Architecture[8]

The dataset used was not enough to train the networks from scratch efficiently. As a result, two techniques have been used to overcome this problem. The first one was to randomly augment all training and validation images at each epoch, to increase the size of the dataset, using various image transformation techniques, such as Horizontal flipping, rotation between 20 degrees clockwise and anti-clockwise and shearing in both the x and y directions. Therefore, technically the number of images used for training was 20,264 multiplied by the number of epochs the model was trained for, which produced up to 202,640 images for training and 25,330 images for validation. The second technique was to import ImageNet weights, which hold 1,281,167 images for training and 50,000 images for validation, organised in 1,000 categories, taking up to 150 GBs of space, before training. The concept of transfer learning was then applied to each of these networks, as explained in detail in the next paragraphs.

The same process was applied to each one of the networks and in the end their results were compared. First, the networks were imported having been already pre-trained with ImageNet weights. The classifier at the end of these networks was removed and replaced by an average pooling layer and a dropout layer, to reduce over-fitting. Over-fitting refers to the scenario where a network can't generalize or fit well on an unseen dataset. A clear sign of machine learning over-fitting is if its error on the testing or validation dataset is much greater than the error on training dataset. Also, as another measure to prevent over-fitting, each of the models were configured to auto-stop their training if the validation data loss, which was calculated for the model after every epoch, didn't improve after 10 epochs from the best validation data loss obtained yet, which also massively saved resources and time. Finally, a dense classifier layer with softmax activation was added at the end of the network which outputs an array for each image, containing the probabilities of the image belonging to each of the 2 classes (non-Melanoma and Melanoma). The first entry in the array corresponded to the probability of the lesion belonging to the first class (non-Melanoma), and the second entry corresponded to the probability of

belonging to the second class (Melanoma). The sum of both probabilities was 1. The class with the higher probability was chosen as the model's prediction for each image. The base network's layers, which is the imported network without the three added classification layers, were frozen at first, which means that all their layers were not trainable and would not lose their values and states, in order not to destroy the pre-trained features of the original model. The resultant network was then trained using the training and validation datasets for up to 50 epochs (due to auto-stopping) with a learning rate of 0.001. After the training was done, the model was fine-tuned by un-freezing the layers of the base imported model and making them trainable. The model was then trained again for up to 50 epochs at a learning rate of 0.00001. An Adam optimization function was used. To obtain the best model, the validation loss was calculated after every epoch, and the model which produced the lowest value was selected and saved at the end. Each one of the models had a different pre-processing function, even though they did the same thing, which was to scale input pixels between -1 and 1. The functions were used on all the datasets before using them for training, validation or testing. In addition, each one of the networks requires a specific input image size. ResNet152V2 requires an input image size of 224 x 224, while both Xception and InceptionResNetV2 require an input image size of 299 x 299. Therefore, all images were resized to the size each network required before use. The whole process of training the network is shown in figure 3.8.

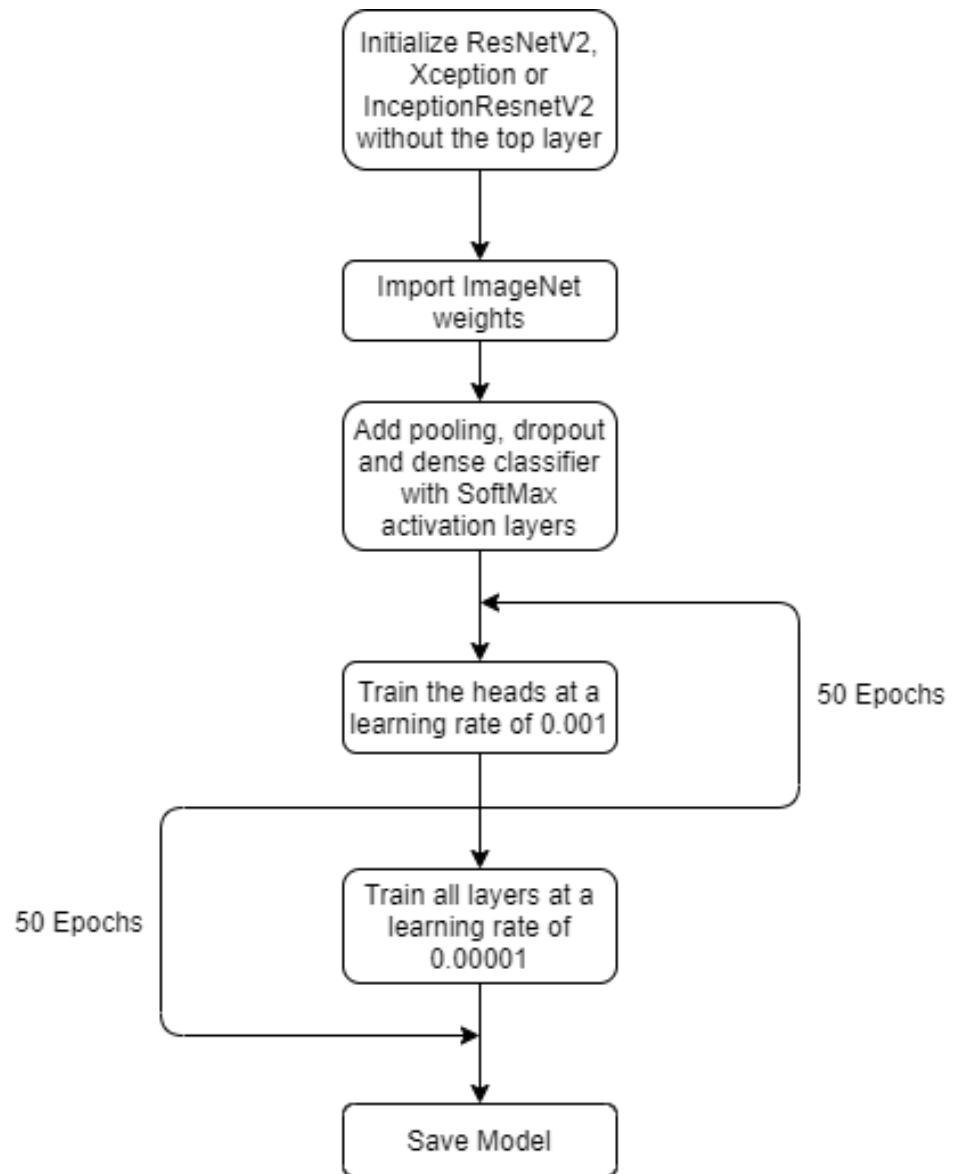


Figure 3.8: Flowchart of the classification network

Chapter 4

Results

4.1 Evaluation Metrics

The following metrics were used to evaluate the three models used:

Accuracy: The ability of a test to correctly diagnose a case.

Sensitivity: The ability of a test to correctly identify patients with a disease.

Specificity: The ability of a test to correctly identify people without the disease.

Balanced Accuracy: The mean between the sensitivity and the specificity. This metric is used for binary classifiers, and is usually used for cases where the number of pictures of both classes is unbalanced, which in our case is true.

$$Accuracy = \frac{(TruePositive + TrueNegative) * 100}{TruePositive + TrueNegative + FalsePositive + FalseNegative} \quad (4.1)$$

$$Sensitivity = \frac{(TruePositive) * 100}{TruePositive + FalseNegative} \quad (4.2)$$

$$Specificity = \frac{(TrueNegative) * 100}{TrueNegative + FalsePositive} \quad (4.3)$$

$$BalancedAccuracy = \frac{Sensitivity + Specificity}{2} \quad (4.4)$$

4.2 Results

4.2.1 Initial Results

As seen in table 4.1, although InceptionResNetV2 had the highest balanced accuracy, it performed very similarly to the ResNet152V2 model, with InceptionResNetV2 being the best one at classifying Melanoma (highest sensitivity), and Xception and ResNet152V2 being the best ones at classifying the lesion as benign (highest specificity). On the other hand, the Xception model's performance was subpar compared to the other two models in terms of sensitivity, which decreased its balanced accuracy.

Base Model	Accuracy	Specificity	Sensitivity	Balanced Accuracy
ResNet152V2	90.3%	96.3%	62.7%	79.5%
Xception	88.0%	96.3%	49.9%	73.1%
InceptionResNetV2	91.0%	97.0%	63.6%	80.3%

Table 4.1: Initial results of all 3 models compared

Confusion matrices were generated for the 3 models in figures 4.1a, 4.1b and 4.1c. The models were obviously biased towards the benign class, which was also confirmed by the high specificity percentages of the 3 models. This is due to the large number of benign images in the dataset compared to Melanoma images. Therefore, as an effort to improve the results, further experiments were conducted on the InceptionResNetV2 model. The reason this model was chosen was that it produced the highest sensitivity out of all three models, and sensitivity is the most important evaluation metric since that classifying whether a lesion belongs to the Melanoma class is our main objective. Also, the consequences of classifying a Melanoma lesion as benign are much more detrimental than classifying a benign lesion as Melanoma.

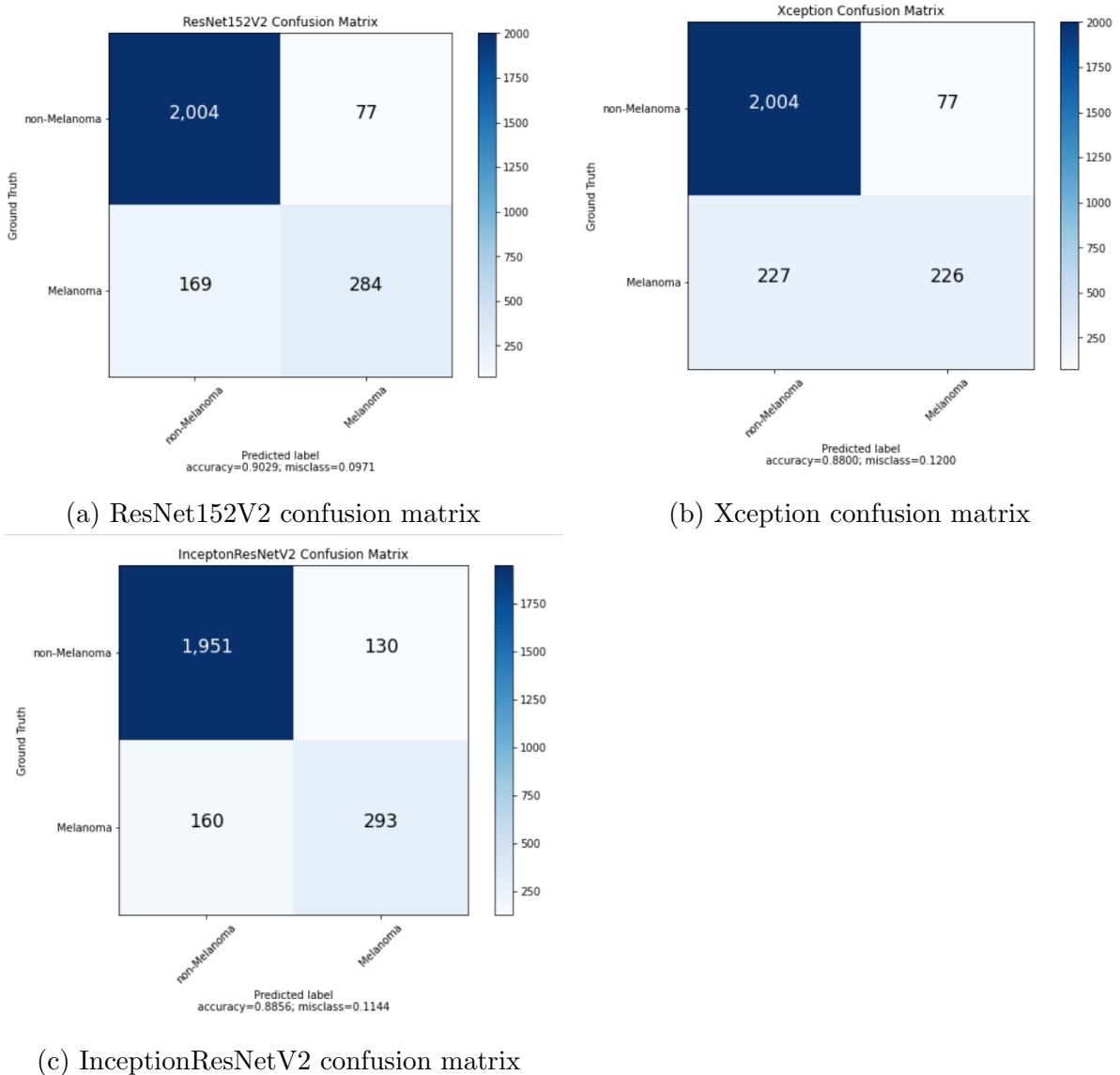


Figure 4.1: Confusion Matrices for all 3 models

4.2.2 Further Experimentation

Data augmentation was applied on the training dataset in order to produce five more models from the InceptionResNetV2 model. The original training dataset contained 16,647 Benign images and 3,617 Melanoma images, a ratio of 4.6 to 1 roughly. For Model 2, random image augmentation, such as rotation between 40 degrees clockwise and anti-clockwise, horizontal flipping and shearing in the x and y directions was applied to each Melanoma image and the produced image was added to the dataset, thus, doubling their amount. Benign images were left untouched. This produced 16,647 Benign images and 7,234 Melanoma images, a ratio of 2.3 to 1 roughly. For model 3, Melanoma images were

left untouched. On the other hand, out of the 16,647 Benign images, 3,617 images were chosen randomly. The obtained dataset consisted of 3,617 Melanoma images and 3,617 Benign images, a ratio of 1 to 1. As for the fourth model, random data augmentation was applied to each Melanoma image as described in model 2. However, 3,617 Benign images were randomly selected from the 16,647 available images, producing a dataset consisting of 3,617 Benign images and 7,234 Melanoma images, producing a ratio of 1 to 2. As for the fifth model, random data augmentation was applied to each Melanoma image as described in model 2. However, in this model, out of the 16,647 Benign images, 7,234 were randomly selected. The final dataset consisted of 7,234 Melanoma images and 7,234 Benign images, a ratio of 1 to 1. Finally, the sixth model consisted of 10,851 randomly-chosen Benign images and 7,234 Melanoma images, a ratio of 1.5 to 1. All six models' datasets are summarized in table 4.2. Each of the models was trained using the exact same approach used on the original model. The accuracies, sensitivities and specificities of all 4 models are compared. Balanced Accuracy was used to choose the best model out of the 6. The results are shown in table 4.3. In addition, confusion matrices were generated for all five new models, as seen in figure 4.2.

Model	Benign Images	Melanoma Images	Ratio
1(Original)	16,647	3,617	4.6
2	16,647	7,234	2.3
3	3,617	3,617	1
4	3,617	7,234	0.5
5	7,234	7,234	1
6	10,851	7,234	1.5

Table 4.2: All six datasets

Model	Accuracy	Specificity	Sensitivity	Balanced Accuracy
1(Original)	91.0%	97.0%	63.6%	80.3%
2	90.5%	95.5%	67.6%	81.6%
3	85.1%	87.0%	76.6%	81.8%
4	81.1%	80.8%	81.0%	81.0%
5	86.2%	87.8%	78.6%	83.2%
6	89.3%	91.7%	78.1%	84.9%

Table 4.3: All six models compared

4.2. RESULTS

27

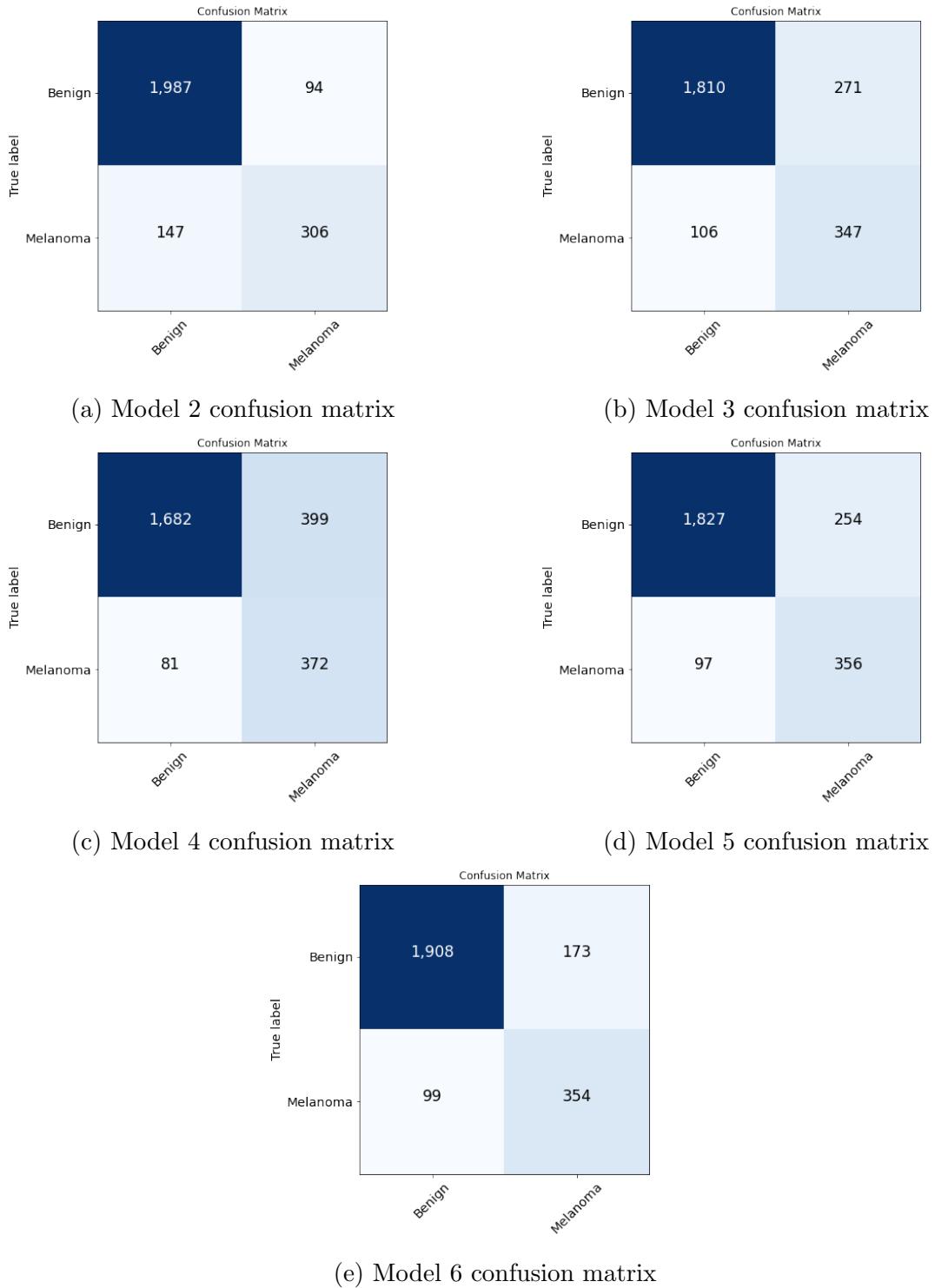


Figure 4.2: Confusion Matrices for the five new models

Model 4 produced the highest sensitivity, while model 1 produced the highest specificity and the highest accuracy. Model 6 produced the highest Balanced Accuracy. As

a result, model 6 was chosen to be the best one out of all six. Therefore, the model was trained for 30 more epochs at a learning rate of 0.00001, and the best model was obtained. In addition, the Xception and ResNet152V2 models were trained again using the same approach, with the dataset used for model 6, which was chosen as the best one. The results for all 3 final models are compared in table 4.4, and their confusion matrices are shown in figure 4.3.

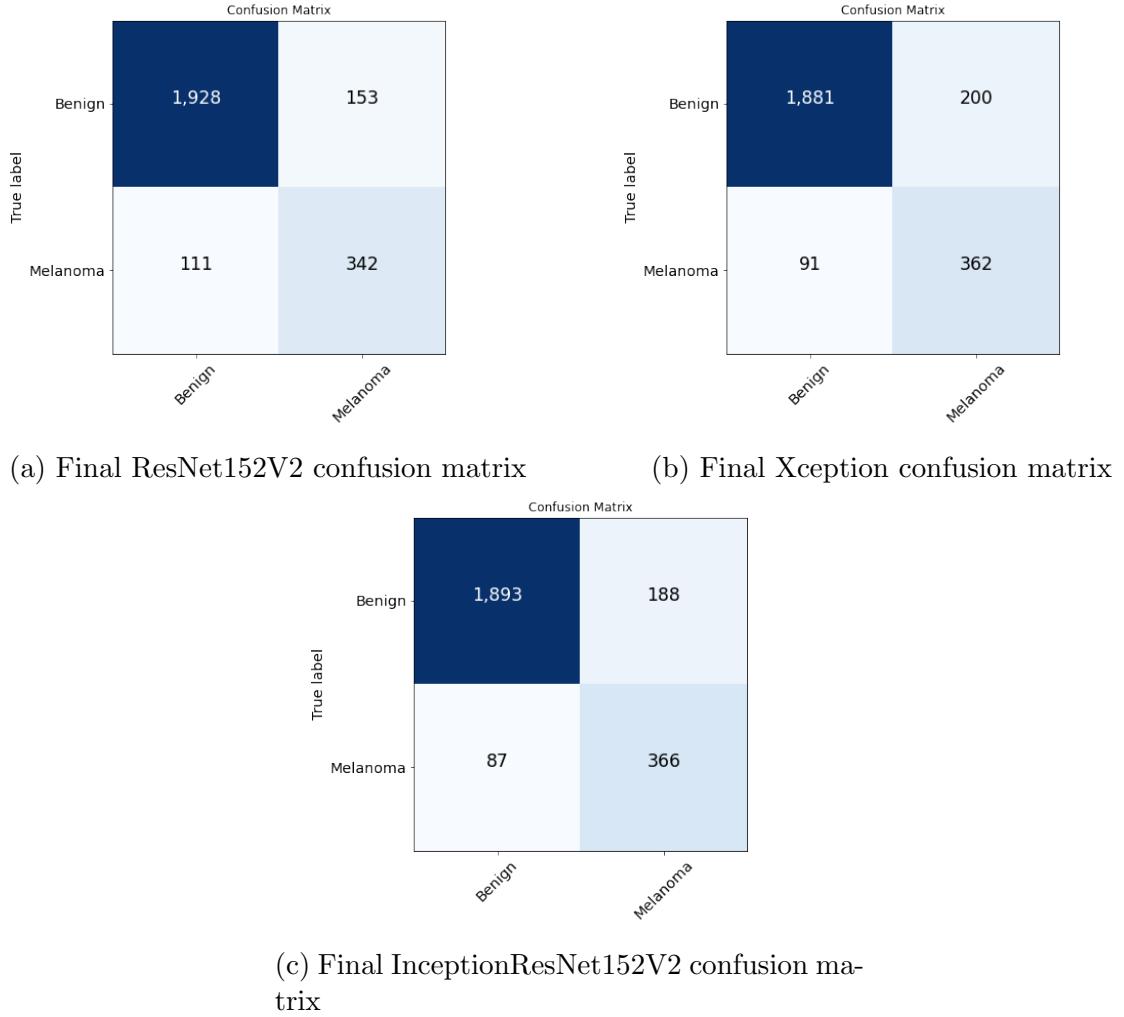


Figure 4.3: Final Confusion Matrices

Base Model	Accuracy	Specificity	Sensitivity	Balanced Accuracy
ResNet152V2	89.6%	92.7%	75.5%	84.1%
Xception	88.5%	90.4%	79.9%	85.2%
InceptionResNetV2	89.2%	91.0%	80.8%	85.9%

Table 4.4: Final results of all 3 models compared

The results for all three models were very similar. However, Out of the three models, the InceptionResNetV2 model produced the best results. Therefore, it was chosen as our main and final model. The model is then compared to the results of the ISIC 2019 challenge’s top 3 teams in the third task, which was lesion diagnosis. In the challenge they were required to classify seven different types of skin cancer. However, in this study, we are only classifying Melanoma. Therefore, the proposed model in this paper was compared to the Melanoma classification results only of the models in the challenge. The comparison is shown in table 4.5.

Team	Accuracy	Specificity	Sensitivity	Balanced Accuracy
Proposed Model	89.2%	91.0%	80.8%	85.9%
DAISYLab	90.3%	97.6%	54.5%	76.1%
Torus Actions	88.7%	91.1%	77.1%	84.1%
DermaCode	89.3%	93.2%	69.9%	81.6%

Table 4.5: Proposed model compared to the top 3 ranking teams’ models in the ISIC 2019 challenge: Task 3

Even though the accuracy of the proposed model is very similar to the accuracies the teams’ models, it massively outperformed the other models, in terms of Balanced Accuracy and Sensitivity. The Specificity of the model was lower than those of the teams’ models. However, as explained before, the sensitivity metric is more important than the Specificity metric as the consequences of a false negative are much more harmful than the consequences of a false positive, so a slight reduction in specificity for a massive increase in sensitivity is more favourable.

4.3 Experiment setting and Dataset

4.3.1 Tools used

- Keras
- TensorFlow
- Python
- Google Colab Pro
- JupyterLab

4.3.2 Dataset

The dataset used is from The International Skin Imaging Collaboration (ISIC). It is an international effort to improve melanoma diagnosis, sponsored by the International Society for Digital Imaging of the Skin (ISDIS). The ISIC Archive contains the largest publicly available collection of quality controlled dermoscopic images of skin lesions. Two datasets were used in this paper. For the first part, the ISIC 2018 challenge dataset was used. As for the second part, the ISIC 2019 challenge dataset was used.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

The detection of Melanoma from images is a significant problem that has attracted a lot of research in the past years. Different methods have been proposed in the past, which have provided good results, despite their limitations. However, recently researchers have been focusing on the use of deep learning methods due to the increasing amount of intuitive datasets available and the dip in performance of the other models when the lesions lack contrast with the surrounding skin. Convolutional Neural Networks have produced excellent results, and continue to do so with the introduction of newer networks, which provide outstanding results in the classification of objects, and increasing availability of images for training. This experiment explores a new approach; instead of training the networks with the available datasets directly, the images are cropped first around the region of interest to eliminate objects which can interfere with the classification abilities of the networks using Mask R-CNN. Then, the cropped images are used to train three modern networks (Resnet152V2, Xception and InceptionResNetV2) and apply Transfer Learning concepts in order to get better classification results. Then, as an effort to increase sensitivity and improve the models, image transformation techniques were used to down-sample and up-sample the Melanoma and non-Melanoma images and experiment with various datasets to train InceptionResNetV2 model (which achieved the highest sensitivity out of the three tested models). The best results were obtained using a training dataset having 1.5 Benign images for each Melanoma image. The dataset contained 10,851 Benign images and 7,234 Melanoma images. The model was trained for 30 further epochs and the dataset was used to train the Xception and RestNet152V2 models using the same approach. Out of the three models, the InceptionResNetV2 model obtained the best results: An accuracy of 89.2%, a Specificity of 91.0%, a Sensitivity of 80.8% and a Balanced Accuracy of 85.9%. The model obtained a 26.3% increase in Sensitivity over the top ranking model in the ISIC 2019 challenge and a 9.8% increase in Balanced Accuracy.

5.2 Future Work

There is still room for improvement and finding models which can have better sensitivities with having significant reductions in specificities. Other state-of-the-art networks will be explored and compared to the three used in this experiment as they get developed. In addition, more datasets will be used for re-training the networks, as they get released. Also, additional data augmentations and ratios between Benign and Melanoma images will be explored.

Bibliography

- [1] “An introduction to deep learning,” *IBM Developer*.
- [2] S. T. X. D. G. . O. 2018, “Artificial intelligence, machine learning and deep learning,” *Xantaro*, Mar 2021.
- [3] “Artificial neural network,” *Wikipedia*, Jul 2021.
- [4] “Convolutional neural networks (cnns) explained.” https://www.youtube.com/watch?v=YRhxdVk_sIs&ab_channel=deeplizard, Dec 2017.
- [5] E. Odemakinde, “Mask r-cnn: A beginner’s guide,” *viso.ai*, Jul 2021.
- [6] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, “Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery,” *Remote Sensing*, vol. 10, p. 1119, 07 2018.
- [7] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, “Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery,” *Remote Sensing*, vol. 10, p. 1119, 07 2018.
- [8] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, “Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery,” *Remote Sensing*, vol. 10, p. 1119, 07 2018.
- [9] L. Tizek, M. Schielein, F. Seifert, T. Biedermann, A. Böhner, and A. Zink, “Skin diseases are more common than we think: screening results of an unreferrals population at the munich oktoberfest,” *Journal of the European Academy of Dermatology and Venereology*, vol. 33, no. 7, pp. 1421–1428, 2019.
- [10] A. C. Society, “Cancer facts & figures,” *American Cancer Society*, 2016.
- [11] “Melanoma warning signs and images,” *The Skin Cancer Foundation*, Jun 2021.
- [12] “Early detection,” *The Skin Cancer Foundation*, May 2021.
- [13] K. Møllersen, H. M. Kirchesch, T. G. Schopf, and F. Godtliebsen, “Unsupervised segmentation for digital dermoscopic images,” *Skin Research and Technology*, vol. 16, no. 4, pp. 401–407, 2010.

- [14] B. Erkol, R. H. Moss, R. Joe Stanley, W. V. Stoecker, and E. Hvatum, “Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes,” *Skin Research and Technology*, vol. 11, no. 1, pp. 17–26, 2005.
- [15] M. Emre Celebi, H. A. Kingravi, H. Iyatomi, Y. Alp Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz, *et al.*, “Border detection in dermoscopy images using statistical region merging,” *Skin Research and Technology*, vol. 14, no. 3, pp. 347–353, 2008.
- [16] A. Green, N. Martin, J. Pfitzner, M. O’Rourke, and N. Knight, “Computer image analysis in the diagnosis of melanoma,” *Journal of the American Academy of Dermatology*, vol. 31, no. 6, pp. 958–964, 1994.
- [17] C. Sagar and L. M. Saini, “Color channel based segmentation of skin lesion from clinical images for the detection of melanoma,” in *2016 IEEE 1st international conference on power electronics, intelligent control and energy systems (ICPEICES)*, pp. 1–5, IEEE, 2016.
- [18] M. Aljanabi, Y. E. Özok, J. Rahebi, and A. S. Abdullah, “Skin lesion segmentation method for dermoscopy images using artificial bee colony algorithm,” *Symmetry*, vol. 10, no. 8, 2018.
- [19] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, “Automated melanoma recognition,” *IEEE Transactions on Medical Imaging*, vol. 20, no. 3, pp. 233–239, 2001.
- [20] F. Peruch, F. Bogo, M. Bonazza, V.-M. Cappelleri, and E. Peserico, “Simpler, faster, more accurate melanocytic lesion segmentation through meds,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 557–565, 2014.
- [21] R. Melli, C. Grana, and R. Cucchiara, “Comparison of color clustering algorithms for segmentation of dermatological images,” in *Medical Imaging 2006: Image Processing*, vol. 6144, p. 61443S, International Society for Optics and Photonics, 2006.
- [22] P. G. Cavalcanti, J. Scharcanski, and G. V. Baranowski, “A two-stage approach for discriminating melanocytic skin lesions using standard cameras,” *Expert Systems with Applications*, vol. 40, no. 10, pp. 4054–4064, 2013.
- [23] M. F. J. Acosta, L. Y. C. Tovar, M. B. Garcia-Zapirain, and W. S. Percybrooks, “Melanoma diagnosis using deep learning techniques on dermatoscopic images,” *BMC Medical Imaging*, vol. 21, no. 1, pp. 1–11, 2021.
- [24] G. Venkatesh, Y. Naresh, S. Little, and N. E. O’Connor, “A deep residual architecture for skin lesion segmentation,” in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, pp. 277–284, Springer, 2018.

- [25] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [26] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [27] A. A. Adegun and S. Viriri, “Deep learning-based system for automatic melanoma detection,” *IEEE Access*, vol. 8, pp. 7160–7172, 2019.
- [28] E. Nasr-Esfahani, S. Samavi, N. Karimi, S. Soroushmehr, M. Jafari, K. Ward, and K. Najarian, “Melanoma detection by analysis of clinical images using convolutional neural network,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1373–1376, 2016.
- [29] C. Yu, S. Yang, W. Kim, J. Jung, K.-Y. Chung, S. W. Lee, and B. Oh, “Acral melanoma detection using a convolutional neural network for dermoscopy images,” *PloS one*, vol. 13, no. 3, p. e0193321, 2018.
- [30] X. Zhang, “Melanoma segmentation based on deep learning,” *Computer assisted surgery*, vol. 22, no. sup1, pp. 267–277, 2017.
- [31] M. Attia, M. Hossny, S. Nahavandi, and A. Yazdabadi, “Skin melanoma segmentation using recurrent and convolutional neural networks,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 292–296, IEEE, 2017.
- [32] L. T. Thao and N. H. Quang, “Automatic skin lesion analysis towards melanoma detection,” in *2017 21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES)*, pp. 106–111, 2017.
- [33] M. A. Al-Masni, M. A. Al-Antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, “Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks,” *Computer methods and programs in biomedicine*, vol. 162, pp. 221–231, 2018.
- [34] W. Abdulla, “Mask r-cnn for object detection and instance segmentation on keras and tensorflow.” https://github.com/matterport/Mask_RCNN, 2017.