Classification of Melanoma using Transfer Learning and Deep Learning Neural Networks

MSc Research Project
MSc in Data Analytics

Vishal Satishbhai Goyal
Student ID: x18180540

School of Computing
National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Vishal Satishbhai Goyal
Student ID:	x18180540
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Catherine Mulwa
Submission Due Date:	17/08/2020
Project Title:	Classification of Melanoma using Transfer Learning and Deep Learning Neural Networks
Word Count:	6872
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Vishal Satishbhai Goyal
Date:	26th September 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Classification of Melanoma using Transfer Learning and Deep Learning Neural Networks

Vishal Satishbhai Goyal
X18180540
MSc in Data Analytics

26th September 2020

Abstract

The world is suffering from many fatal skin disease, one such disease causing fatalities is Melanoma skin cancer which is being researched from the past many years. Different Deep learning Neural Networks have achieved optimistic results in identifying the Melanoma. However, due to the similar optics of malignant and benign tumor images, it becomes very difficult to get precise results and so a huge improvement is needed. This project helps the automatic classification of melanoma images using deep learning and transfer learning models. Image Augmentation and pre-processing are performed on the images for the model to achieve better accuracy and at last, all the applied models are evaluated based on accuracy and loss. After comparison, it was observed that MobileNet, LeNet, and ResNet achieved better accuracy as compared to CNN and AlexNet. Highest accuracy of 75.38% was achieved by MobileNet.

Area: Deep Neural Network, AlexNet, LeNet, ResNet-50, MobileNet, CNN, Image Processing, BlackHat Filter, Adam Optimization.

1 Introduction

1.1 Background

The skin condition is one of the major causes of fatalities worldwide and it poses one of the greatest threats to human lives. One such condition is skin cancer which is very common worldwide. World Health Organization(WHO) in 2018 press release an estimated 18.1 million new cancer cases and around 9.6 million deaths. Asia alone estimated over half the cases due to a large population followed by Europe, America and Africa. Melanoma is one serious kind of skin cancer that is caused by the turbulent growth of pigment-producing cells. The prevalence of skin cancer is affected by both regional and age-related influences, with melanoma being the most prevalent condition in regions such as Australia, New Zealand, Northern Europe, and North America Siegel et al. (2019). It can mainly be caused by exposure to UV rays from the light which increases the chances of normal benign mole converting to melanoma. Other possible factors include genetics and immune deficiency conditions Guy Jr et al. (2015). However, it can be treated if detected in the early stages by performing the surgery, but advanced stages can spread to internal organs of the body from the skin.

For most of the cases, a biopsy is needed by dermatologists to decide if a tumor is in an early stage or higher stage. Although this course of action involves a certain amount of expense and dejection, automatic detection techniques are considered fast, efficient, and reliable for skin cancer identification Hoshyar et al. (2011). Dermatology imaging experts claim that skin melanoma diagnosis can be streamlined based on certain anatomical characteristics and colour details typically considered as types of various skin cancer. The main diagnostic of melanoma has been found to be the vertical thickness, 3-Dimensional shape/size, and lesion colour Karargyris et al. (2012).

Much research has been done in the past for the image classification of Melanoma lesion images where pigment network-based skin cancer detection approach with Artificial Neural network and different CNN based architecture(U-NET, FCRN and DRN) Vinay et al. (2020). Also, image segmentation techniques like the K-means clustering algorithm and the Fuzzy C-means clustering algorithm provided effective results Aishwarya et al. (2020). This project uses various image augmentation techniques like Hair removal, Angle Rotation, Flipping, GaussianBlur filter to enhance the image for the model to learn well.

1.2 Purpose of Study

It is clear that the techniques used for image classification for skin lesions need to make use of the semantics of the Melanoma images by learning the shape, dimensions, color, and also the patient's biological history like age, gender, etc. Previously research shows promising results for the classification and detection of Melanoma but still, there is some room for improvement. This research considers many useful image preprocessing techniques and compares different models results in the classification of Melanoma.

1.3 Research Question

"Can different augmentation techniques on skin cancer images improve the model capabilities to learn efficiently for binary classification tasks with smaller datasets?" Augmentation techniques like hair removal, image flipping, angle rotation, resizing, Gaussian filter is implemented and then images are fed to different CNN architecture models and transfer learning models.

1.4 Objectives and Project Structure

Section 2 is entirely dedicated to the previous research done in the field. **Section 3** provides a detailed explanation of the research methodologies that are used to achieve the objective and complete the project. **Section 4** discusses how the project methodology was implemented. **Section 5** discusses evaluation techniques used to evaluate the performance of the model and comparison between based on accuracy. Then the project is concluded in this **section 6** along with future work. Figure 1 shows the objectives set for the project.

Objectives	Description	Techniques	Evaluation
1	Critical Review Of Literature for Melanoma classification(2010-2020)		
2.1 2.2 2.3 2.4	Data Collection Exploratory Data Analysis Data Pre-Processing Feature Extraction	Image-Resizing, Image-Flips, Image-Rotations, Blackhat-Filter, Threshold-Filter, Inpainting Filter, Grey-Scale Filter, Hair Remove filter, Image -Data Generator	
3.1 3.2 3.3 3.4 3.5	Implementation Using MobileNet Implementation Using AlexNet Implementation Using LeNet Implementation Using CNN Implementation Using ResNet50		Adam Optimizer, Accuracy, Loss
4	Evaluation Of Different Models		

Figure 1: Project Objectives

2 Related Work

Many organizations have devoted time and resources to develop the early screening process due to the importance of early detection of melanoma. In the past, many researchers have worked on learning the semantics of the cancer cells based on the shape, size, and color of the skin lesions. Also, many supervised and unsupervised learning methods were used for the same. With the improvements in the methodologies used for classification, Neural Network-based approaches have provided very good results. The review for the work done in the past has been discussed below.

2.1 Traditional Methods For Image Classification

Ballerini et al. (2012) proposed an algorithm for the diagnosis of non-melanoma skin lesions using a novel hierarchical grouping K-Nearest Neighbors (K-NN) approach. The classification task was performed on five common classes of non-melanoma skin lesions. With their hierarchical approach, the image classes were divided into two groups, one containing the BCC, AK, and SCC, and the second group consists of classes SK and ML(Melanocytic Nevus/Mole). All these are dermatologist's defined diagnosis classes. The selection of features was encoded in the hierarchical framework which selects the most important feature subsets at each node of the hierarchical structure. Features like color and texture were then extracted from lesions. The key contribution of this research was they covered the majority of the skin lesion types and the results were based on color images, unlike the dermoscopy images.

Irregular streaks are valuable indicators for the diagnosis of melanoma (a highly dangerous type of skin cancer) using images of dermoscopy. A three-way classification method for the detection of streaks in a pigmented skin lesion namely Absent, Regular, and Irregular was introduced by Sadeghi et al. (2013). Additionally, to detect the underlying pattern, the directional pattern of observed streaks is analyzed to obtain their orientation features. For identifying the valid streaks, the directional pattern is analyzed to extract the features to detect the underlying pattern. A set of 18 features were proposed namely

STR (streaks), inclusive of 3 Structural, 3 Geometric, 6 orientation, and 6 chromatic characteristics. Then, the method considers a graphical visualization to model the geometric pattern of genuine streaks. With the help of features and the color and texture of the entire lesions, they were able to achieve around 85 percent accuracy based on 940 images.

A similar approach was proposed by Mhaske and Phalke (2013) using supervised and unsupervised learning algorithms. Only 150 images were considered for their research, where the images were first converted to greyscale. Then some segmentation on the images was performed to obtaining the region of interest. A Multi-Layer Perceptron was used which performs well with the dermoscopy images. According to them the cancer features resides around the border of the lesions. So to obtain important feature 2-D wavelet where the original image is divided into 4 parts at first level decomposition and then in second decomposition each of the remaining 4 parts is again divided into 4 parts making it to 16 parts. Then a high and low pass filter is applied where the low pass filter output gives an approximation of the original image. Features are calculated based on mean, median, mean-variance, standard deviation, etc on the image vector. Three different algorithms were used for classification namely Backpropagation neural network, K-means Clustering algorithm, and support vector machine out which the support vector machine algorithm outperforms the other two in terms of classification accuracy.

Linsangan et al. (2018) proposed the use of geometric features to detect melanoma. Their methodology includes fetching images from the repository, then the images were resized along with the application of bright and contrast modification to compensate for the non-uniformity in the illumination images. Gaussian blur was then applied for removing the background noise from images. Further to this, the images were grayscaled and diluted function was applied to enlarge the interested boundary. This helps increases the foreground boundary pixels and decreases the holes pixels present within the foreground. Then segmentation of the images was done where images went through thresholding where the grey-scaled images were transformed into binary form. Then edge detection was performed were change in brightness occurs sharply based on pixel values. On the extracted features parameters like area, perimeter, diameter, and irregularity index were obtained to get the shape uniformity in the image. The final extracted features were passed to the k-Nearest neighbor algorithm. The methodology worked well with the smaller image set.

Kamboj et al. (2018) proposed the same traditional methodology involving image preprocessing, segmentation, feature extraction, and classification. To remove the unwanted artifact methods of preprocessing like hair removal using Dull-Razor and thresholding technique for removing the reflective artifacts were used. After that median filter was used for smoothening the image. The images were then converted to greyscale and the active contour method was used for the segmentation purpose. The mask obtained was enhanced using morphological operations to obtain the bounding box and from that, the region of interest is cropped and resized. Twelve different features were extracted from each HSV and YCbCr color space and joined to make a feature vector. For classification between Benign and Malignant class three different classifiers namely kNN, Naive Bayes, and Decision Trees were used from which Decision Tree outperformed compared to the other two classifiers. Authors also discussed the future work where they suggest texture, shape, and border features can increase the accuracy by a good amount.

2.2 Support Vector Machine For Image Classification

Maurya et al. (2014) proposed a framework that involved texture feature extraction using the GLM(Gray Level Co-Occurrence Matrix) where distinct features like auto-correlation, Energy, Entropy, Contrast, and homogeneity from each cancer class category were extracted. These extracted features were then passed as input to the multi-class SVM(Support Vector Machine) classifier. Their methodology using both GLCM and MSVM(Multi-Class Support Vector Machine) provides good results for the classification of 4 different types of skin cancer.

Suganya (2016) discussed the method for the classification of melanoma as well as non-melanoma skin lesions. According to the author, the feature extraction and segmentation techniques are a very important part in classification tasks. As a part of image processing, the median filter method was used along with smoothing to remove artifacts like hair instead of the Dull razor software which only works well with thin hair artifacts. The border detection method using CLAHE(Contrast Limited Adaptive Histogram Equalization) was implemented to remove the illumination around the tumor area. K-means clustering was used for the segmentation process. The pixels obtained within the detected border are further divided or classified into 3 areas, where 30 percent of the whole tumor area was named as peripheral and the inside most part is the main tumor. Then feature like color, sub-region, and texture where extracted from the peripherals, central tumor, and the whole tumor using the Wilkis lambda method. Then 2 step classification was performed for the classification of Melanocytic skin and Non-Melanocytic skin lesions using the SVM(Support Vector Classifier) and was evaluated using the metrics like sensitivity, accuracy, and specificity. The results set the benchmark for accuracy over 90 percent. A similar methodology was used by Mane and Shinde (2018) compared to the different functions used by the SVM classifiers and concluded that SVM linear function provided better results in comparison to bayesnet and radial functions in the classification of Melanoma classes benign and malignant.

Alquran et al. (2017) research was focused more on the feature extraction part where the research aims to correct the images which don't seem to be homogeneous due to the illumination during image capture. Various pre-processing techniques like RGB to grayscale, noise filtering, contrast enhancement, median filter, and histogram equalization were implemented to remove the unwanted noise from the dermoscopy images and enhance it to makes the segmentation process as accurate as possible. Otsu thresholding was used to get the region of interest and then image filling was used to remove the background pixels which was outside the region of interest. Further image opening, masking the ROI, converting the image to grayscale, and histogram equalization was performed one after the other. Four-step of feature extraction techniques using GLCM, ABCD was performed for differentiating the skin lesion types based on color, shape, and texture. These features were passed to the Principal component analysis to drop the ineffective features. The final results were fed to the SVM model for the classification of the lesions into benign and malignant. The SVM model was also used by Masood and Al-Jumaily (2017) to deal with the unlabelled data. In the research, they proposed a semi-advised training and classification algorithms which can effectively use the minimal labeled data along with the large un-labeled data. This methodology first extracts the features from the labeled data and then feds smaller subsets of the unlabeled data to train the model based on the matching features. As there is a good risk of having misclassified data items, the impact of misclassified data is regulated by collecting advice weights values based on discrepancies in classification and using these weights in combination with SVM decision values. Those weights also enable the algorithm to eliminate the outliers.

A new algorithm to distinguish the dermoscopic images into malignant and benign is introduced by Nezhadian and Rashidi (2017). The images were initially segmented with the use of an active counter model and two features like texture and color components were extracted. ABCD features were extracted based on color and texture. Along with this 2-D wavelet transformation was also applied to get the infected boundaries. All the extracted features were then fed to the SVM classifier to classify between the skin cancer type classes.

Lynn and Kyu (2017) suggested a systematic framework for classifying dermoscopic images of pigmented skin lesions. The skin images were first used to eliminate unnecessary hair and noise, and then the segmentation process to isolate the affected area. The mean-shift algorithm was implemented which segments the lesion from the whole image to detect melanoma skin cancer. The separation of features is then carried out following the basic principles of ABCD dermatology using the RELIEF which the feature extraction method that works on feature weight estimation. Also, PCA was applied to reduce the features having collinearity. After extracting the lesion characteristics, the feature selection algorithm was used to provide optimized characteristics to provide the classification stage. The classification algorithms like kNN, SVM, and decision trees were then used to classify the selected optimized features.

A segmentation technique for Detection of Melanoma using the SVM classifier and Snake model was developed by Bumrungkun et al. (2018). This methodology was developed as a pre-processing step to obtain the area or border of the cancer image excluding the outside area which is the background. The three types of initial symmetry curves are set which are circle, ellipse, and rectangular. Further, the images of different shapes, sizes, and colors are feed for the SVM classifier to classify them into the defined template. Further, based on the color, shape, and size features, these images are passed into the snake model where the symmetry of the images is learned and the model tries to predict the border of the image where cancer mole is present. Similarly, when a new image is passed in the SVM classier, the model compares it to the already stored N templates and finds a match and the snake model predicts the borders.

2.3 Artificial Neural Network Based Image Classification

Jain and Jain (2012) concentrated on creating a screening method for skin cancer that can be used by non-experts in general practice to distinguish usual from unusual cases. This method consists of feature detection and classification techniques. They used the Matlab function to read the image and edge detection function was used to draw a contour plot around the image. The discrete 2-D wavelet transformation was performed on the decomposed image and approximation coefficients matrix cA was composed. This composed matrix was then fed to the neural network classifier which was a probabilistic neural network that classifies whether the image represents the cancerous or non-cancerous class. The model was tested for only 40 skin cancer images and as the sample size was increasing the accuracy for their model was getting reduced. This can be considered as a drawback to their research.

A new methodology was developed by Alfed et al. (2015) using Artificial Neural Network. The dermoscopy images were first converted to blue component images and then some artifacts were removed from the image which distorts the image in some kind. Hair detection and removal were a crucial part of their processing as extracting the pigment network from images becomes difficult because the hair has the same linear shape as the pigment network. Gaussian filter was also applied to the images to remove the background effect of the images. These features were then fed to the two-layered ANN model to perform the binary classification where one output layer and one hidden layer with 120 neurons and Radial Bias as an activation function was used. The classification accuracy was taken as an evaluation matrix and all true positive, false negative, false positive, and true negative cases were extracted.

A Deep CNN based architecture for the efficient identification and segmentation of melanoma lesions was introduced by Adegun and Viriri (2019) in this article. This design adopts an improved deep CNN connected to a series of skip pathways. It consists of a lower-size encoder-decoder network that aims to minimize the consumption of computational resources. The multi-stage strategy overcomes the drawback of certain deep CNN in generating coarsely segmented outputs while processing complex images of skin lesions. Under this method, the entire network is split into steps, with each step covering specific learning and extraction of features. A new method for classifying skin lesions was derived from the outcome of the softmax classifier. The process adapts dice loss function, which learns and measures losses from the variance between the expected performance and the reality mark into a pixel-wise classification softmax classifier. The program was designed to reduce the difficulty of profound learning architecture in melanoma identification. This also seeks to build an integrated framework for the treatment of melanoma cancer that will fulfill the challenge of medical diagnosis in real-time.

2.4 Transfer learning Models For Image Classification

Setiawan (2020) proposed a very different and simple methodology using the Convolutional Neural Network architecture. The size of the original images was checked and the lowest resolution image size was considered to resize all the images. Then popular image enhancement techniques CLAHE and MSRCR were used. The original image was split to get 3 channel colors. Then the three color channel images were merged to form an RGB color-enhanced image. The resulting image was a more enhanced version of the original image was the lesion was visible. Another technique MSRCR was applied to the original image to get an enhanced RGB image. For the classification purpose, the VGG16 a CNN architecture was used. This study involves training and testing on two different datasets, one is the CLAHE based enhanced version and the second is the MSRCR based RGB enhanced version. 50 epochs were performed to train the model where it was observed that the validation dataset results were almost the same for original, CLAHE, and MS-

RCR enhanced image. CLAHE is more suited for use with CNN for early diagnosis of skin cancer in color image enhancement. This key contribution is that the image contrast improvement is not essential for testing for skin cancer. This will also lower the load on the computation.

Goyal et al. (2019) proposed an ensemble-based approach for the segmentation of skin lesion images. The images had different background color and the size of the images was large, so to reduce the computational cost and to give a similar color to the image the images were converted to gray-scale using the shades of gray technique, and the images were resized to 500 x 375. The DeeplabV3+ encoder-decoder model was trained using the pre-trained PASCAL VOC. The semantic label lesion was assigned to every pixel in the dermoscopic image. Another segmentation was performed using the Mask RCNN which was fine-tuned with Res-Net Inception V2 model om MS-COCO dataset. Basic post-processing was performed to remove unwanted artifacts. Two ensemble techniques were used namely Ensemble ADD and Ensemble-comparison. If DeeplabV3+ doesn't provide an output then output prediction from Mask RCNN is considered and another way around. Both the results are combined and the evaluation is done. This research outperformed the previous state of art model in sensitivity, specificity and accuracy.

The study done by Ahmad et al. (2020) discusses a model that uses Deep Convolutional Neural Networks with triplet loss function to enhance the classification of skin disease. Fine-tuning of all layers of ResNet152 and InceptionResNet-V2 was executed to tackle the issues of images of facial skin condition. Next, the 128-Dimensional features were extracted to Euclidean space from the training samples, then L-2 distances were determined using learned embedding between the corresponding images. Thereafter, the task of classification of skin disease was accomplished given the gap of L2 between images. The experiment was performed on four types of skin lesions and a dataset of about 12000 images. The methodology outperformed the previous state-of-art techniques in terms of accuracy.

3 Research Methodology

The project uses a modified methodology with the underlying structure of Knowledge Discovery(KDD).

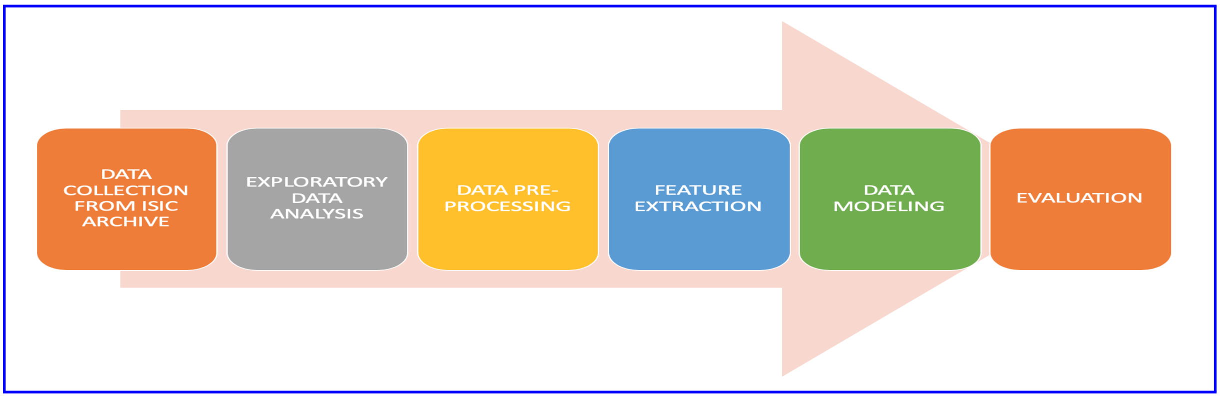

Figure 2: Process Flow Diagram For Melanoma Classification

The main objective of this project is to classify between the Melanoma and Benign skin lesion images from a very biased image dataset taken from the International skin imaging collaboration (ISIC) Archive database which is publicly available skin lesion images for research work. Figure 2 shows the process flow diagram which consists of different phases starting from data collection, data exploratory analysis, data preprocessing, feature extraction, model building, classification of skin lesions, and evaluation to complete the project.

3.1 Data Collection

The dataset chosen for the project is taken from an International skin imaging collaboration (ISIC) Archive database [1] which is publicly available skin lesion images of Benign and Malignant categories for research work. The dataset was produced by the International Skin Imaging Collaboration. The images were prepared by Hospital Clínic de Barcelona, Medical University of Vienna, Memorial Sloan Kettering Cancer Center, Melanoma Institute Australia, The University of Queensland, and the University of Athens Medical School. The dataset consists of images in different formats namely DICOM format which is a medical image format and the other format is of JPEG. The DICOM image format is of fixed size 1024*1024, while the JPEG images are of different sizes. Figure 3 shows some examples of the Melanoma skin lesion images and figure 4 shows the benign category.The dataset is highly imbalanced with 98% benign cases and just 2% malignant cases. After closely studying the dataset, it was found that there are multiple records for a single patient in the dataset. Some of the patients have more than 250 images for the benign class. Due to this, the dataset is biased towards the benign class. The malignant images are less in comparison to benign class. Therefore, for the model to work well only down-sampling was performed, where the malignant cases with count 584 were taken as minority class and the benign class was downsampled to make it equal to the number of malignant cases. There were around 1,168 unique images were considered for the project. The malignant skin lesions are asymmetrical, have uneven borders, double shade, etc. But the benign are symmetrical, single-colored, even borders look like a small mole Adegun and Viriri (2019).

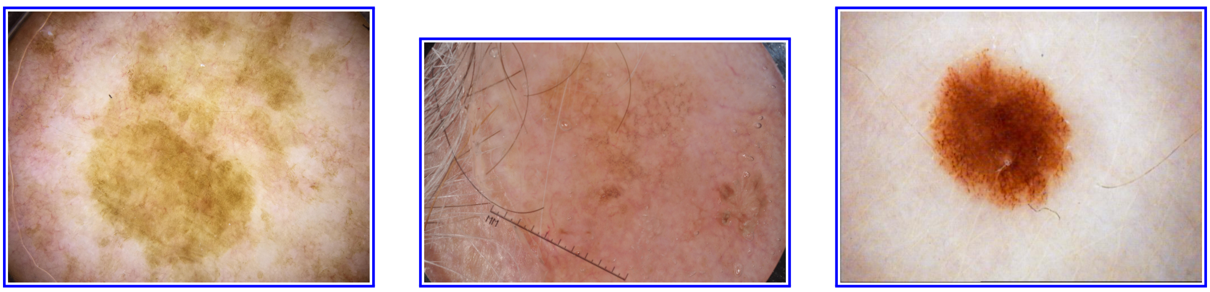

Figure 3: Melanoma Skin Lesion Images

[1] https://challenge2020.isic-archive.com/

Figure 4: Benign Skin Lesion Images

3.2 Exploratory Data Analysis

The data exploratory analysis was carried out to understand the data and the metadata attached to it. During the analysis, various findings were gathered which played a vital role in processing the images. In the dataset the ratio of Male to Female having benign and malignant cases was discovered and from figure 6, it is evident that males are more exposed to melanoma as compared to females. There were six different body parts like the torso, head/neck, lower extremity, upper extremity, palm/sole, and oral-genital where the cancer was present. Among these body parts as shown in figure 6 cancer is mostly present on torso followed by lower and upper extremity. Cancer is less present on head/neck and genitals.

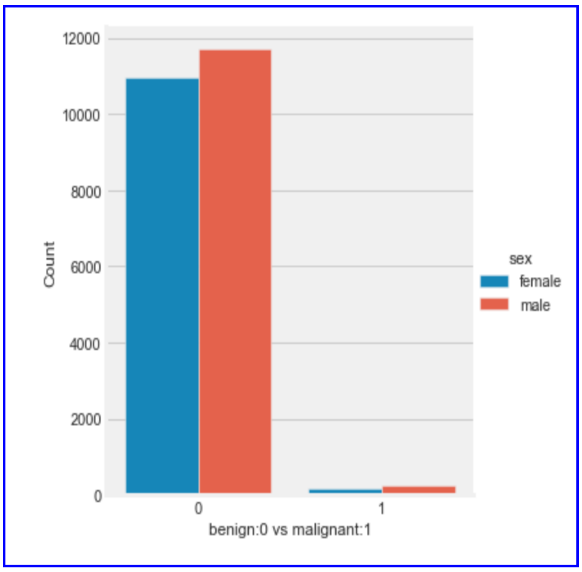

Figure 5: Melanoma Case w.r.t Sex

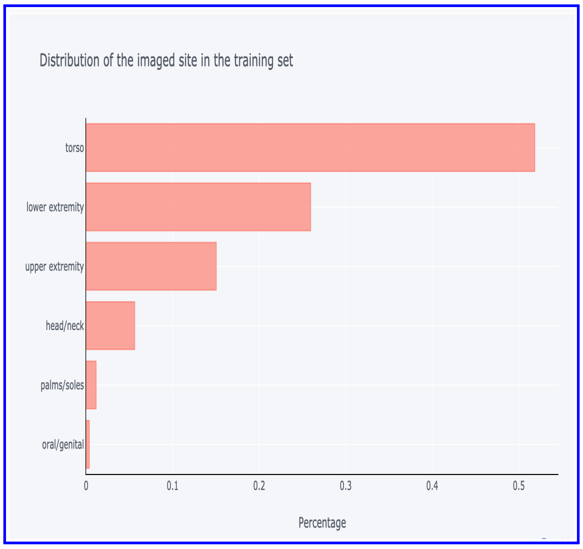

Figure 6: Body Part With Cancer

From the figure 7 it is clear that males are more exposed to cancer on the torso, head/neck, and palm/soles as compared to females. In the case of lower and upper extremity females has more count than males. Age is also considered to be one of the factors which contribute to cancer stats where it is clear from the figure 8 that malignant cancer is more common between the age group of 50 to 75. Also, the benign cases are mostly from the age group between 40 to 70. For most of the cases, the cancer category is unknown which amounts to 81% of the whole dataset, the next category is nevus which amounts to 18% of the dataset and rest all categories like melanoma, seborrheic keratosis, lentigo NOS, lichenoid keratosis, solar lentigo, and atypical melanocytic proliferation

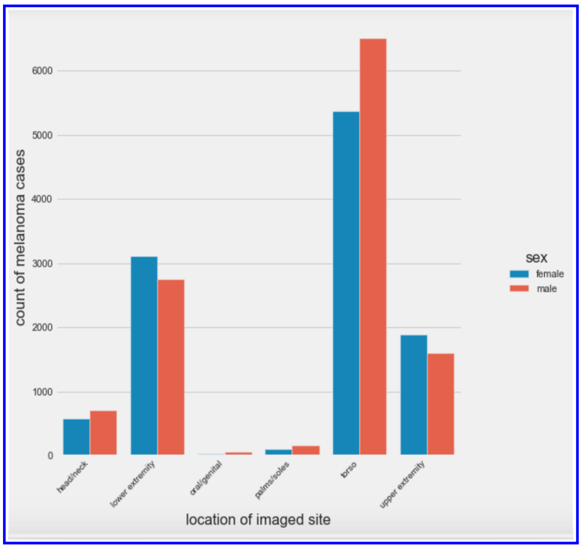

Figure 7: Location of Cancer w.r.t Sex

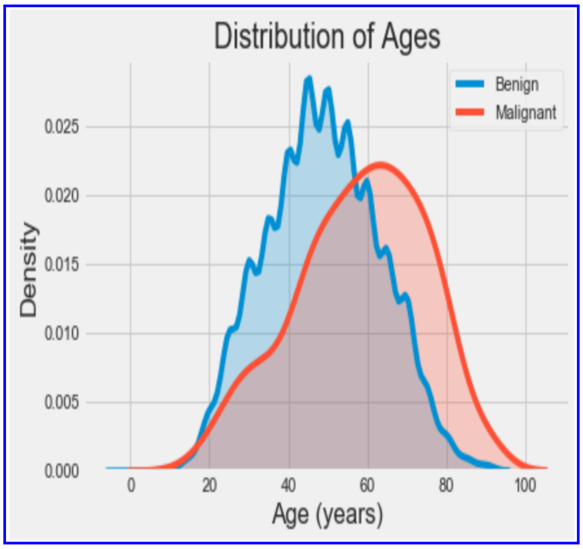

Figure 8: Age w.r.t Cancer Category

amounts to 1% of the cases as shown in figure 9. The images in the dataset are having different dimensions and have different brightness and contrast. The benign cases have a higher intensity of red color pixel followed by green and blue (10) while in case of malignant cases the red color has high intensity followed by blue and then green (11).

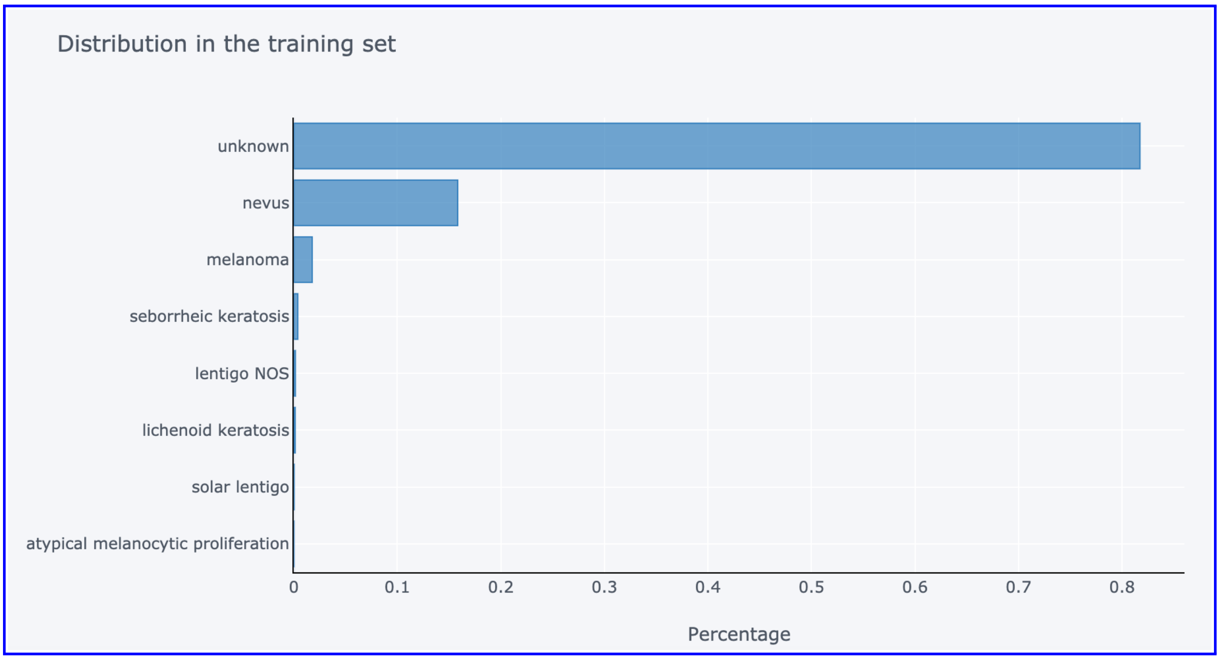

Figure 9: Type of Cancer

3.3 Data Pre-Processing

The images from the downsampled dataset containing 1,168 were split into training and testing sets. Then, the images were divided into two different folders based on the malignant and benign cases. As the image was of different sizes, all the images were resized

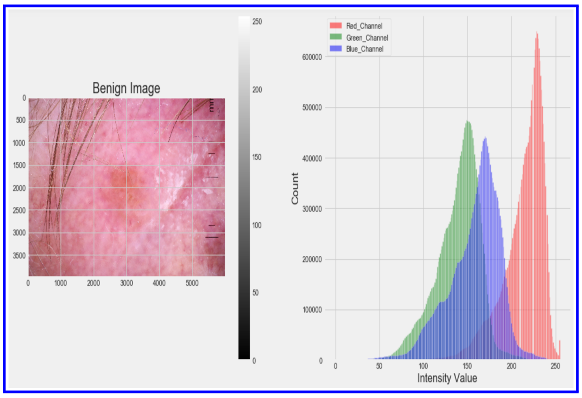

Figure 10: Benign Color Intensity

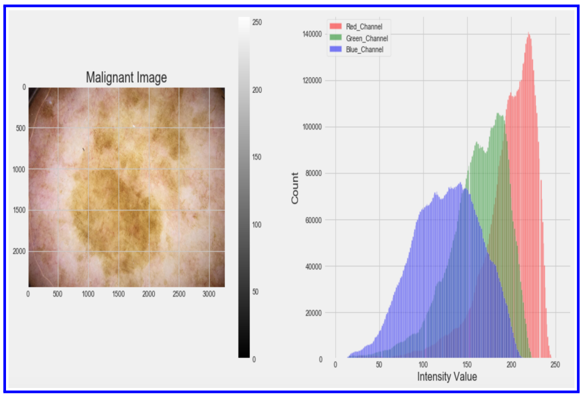

Figure 11: Malignant Color Intensity

to 224*224*3 for maintaining the same size for each image. Various image filters were used to remove the unwanted artifacts from the image like hairs from skin, the photograph bars in some of the images, and noise. To remove noise and hairs from the images

BlackHat morphology and Total variation were used from the work done by Adil et al. (2020). The images were first converted from RGB format to a single Grayscale format. In this transformation, the red channel images are converted to gray ones. After this, the BlackHat filter was applied to enhance the regions of the images through opening closing. This method extracts the components which are darker than their surroundings. The darker artifacts which are found from these images are then passed through the binary thresholding transformation which increments the intensity of the artifacts detected. Now to retrieve the original image without hair, the total transformation called in-painting is applied using the original image and the artifacts counter-image. The recovered image obtained from this process is free from unwanted artifacts(Hair). Figure 12 shows the whole transformation process.

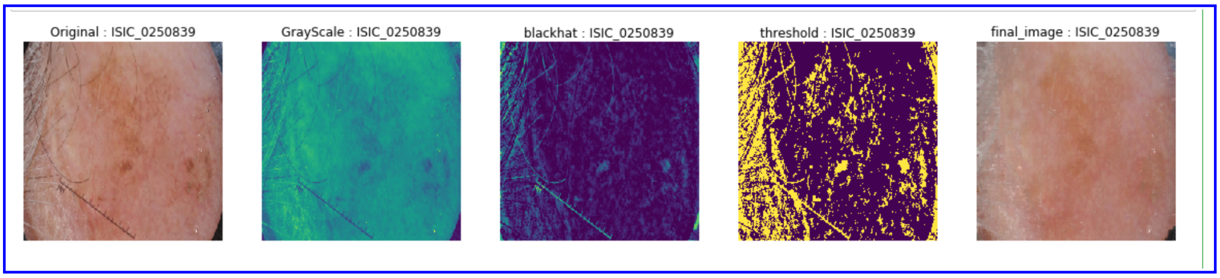

Figure 12: Image Pre-processing Process

After preprocessing the image, it was necessary to augment the data to train the model efficiently. Neural network models often require a large amount of data for training. Data augmentation was performed using the Keras Image data-generator. It is a technique that helps to expand the variety of data required for training models massively, without explicitly gathering additional data Chugh et al. (2020). For this project, many parameters were used inside the image data-generator like Flipping, horizontal flips, vertical flips, height shifts, width shifts, and images were rotated at 20 degrees angle and the fill mode was set to the nearest. The Keras flow_from_directory method was used to read the images from the directory where the images were already divided into the malignant and benign category groups.

3.4 Feature Extraction

The Pillars of Image classification relates to the identification of basic shapes and geometry of objects. That is a mechanism containing image normalization tasks, image segmentation, extracting key features, and identifying the classes. The CNN architecture has redefined this area by understanding basic shapes and geometry in the first layer, then mastering the image features in the deeper layers, which results in a more precise classification of the images Jogin et al. (2018). For this project, the CNN architecture takes care of the feature extraction process and trains the models based on these extracted features.

3.5 Modelling

Many different image classification techniques and models are built in the past for skin cancer classification. From the literature, it is clear that transfer learning and deep learning models work well with the image classification tasks as compared to traditional methods. For this project different transfer learning models like MobileNet and ResNet50 along with deep learning models like CNN, AlexNet and LeNet are build to classify the binary class label problem and further comparison between them is being explained.

3.6 Evaluation

The evaluation of the models performing binary classification is evaluated using the metrics of accuracy and loss. Adam optimization is applied to the model to optimize the results of the models. Graphs of Epoch versus Accuracy and Epoch versus Loss is presented for results.

4 Implementation

This section discusses the technologies used for completing the project as well as the implementation details. Figure 13 shows the different technologies adapted from the starting phase till the results phase.The project implementation is coded in python language and the IDE used for running the code was Jupyter Notebook on Anaconda Platform. Anaconda is a free and open-source platform where huge data processing, predictive analysis, data mining, etc can be performed using Python and R programming. It is a preferred platform as it takes care of all the library management and dependencies. Jupyter notebook is also an open-source application that runs on the web and allows the creation and sharing of documents consisting of live code, visualizations, data processing, etc. For plotting the graphs for Exploratory Data Analysis and evaluation matrix MatplotLib was used which is a library for visualization in python. Keras is used as a backend. Keras is indeed an API that's designed for users, not computers. Keras promotes best practices for lowering the cognitive load. It provides reliable and quick APIs, it helps to reduce the number of user interactions required for common usage cases and delivers transparent and supportable error messages. This also provides comprehensive guides and handbooks for users. It sits on top of TensorFlow architecture. Keras gives the flexibility to develop all forms of frameworks; it may be RNN, CNN, simple NN, Deep NN, etc.

For this project, Keras is used for implementing the models because with machine learning models the tasks like feature extraction need to be performed separately but with Keras, the feature extraction is being handled by itself.

Figure 13: Technologies Adapted

4.1 Implementation using MobileNet

MobileNet is used as a model for this project, as its architecture is lightweight. It utilizes depth-wise separable convolutions which generally means that on each color channel it executes a single convolution instead of adding all three and flattening it. It does have a filtering function of the input channels. The element-wise convolution, therefore, adds a 1x1 convolution to integrate the deep convolution's outputs. A generic convolution filters and integrates the inputs into a newer set of outputs in a single step. This is divided into two layers by the depth-separable convolution, a different filter layer, and a second layer for combining. This factorization results in a massive reduction in computation and size of the model Howard et al. (2017). A slightly modified MobileNet architecture is followed where transfer learning is employed by retraining a few of the top layers. The last four layers of the MobileNet model is frozen. Also, pre-trained ImageNet dataset model weights are used and the layer trainable is set to false. For the model to learn many complex features, 4 dense layers are added where dense layer 2, and dense layer 3 has activation function "Relu". Since Relu is almost linear, they maintain most of the features that form linear models simple to optimize with gradient-based methods. For the last dense layer sigmoid is used as an activation function. The key reason for the use of sigmoid is that this function occurs from (0 to 1). Hence it is particularly applicable for models where there is a need to estimate the likelihood as an output. As there is just the possibility of something between the 0 and 1 scales, sigmoid is the correct option.

Now the training and testing data is loaded into the Image DataGenerator by using the path of the directory passed to flow_from_directory function. Here the image is resized to 224*224. The batch size for training is kept to 16 and for the testing set, it is kept to 1. Then, the model is compiled using the loss function as binary cross-entropy, the optimizer used is Adam and the metrics are kept as accuracy.

4.2 Implementation using AlexNet

One of the best deep learning models AlexNet which gave a challenge to CNN in terms of training time was developed by Krizhevsky et al. (2012) which performed well with the image classification tasks. This model was used for this project for the image classification task. Figure 14 shows the architectural diagram for the AlexNet. The AlexNet

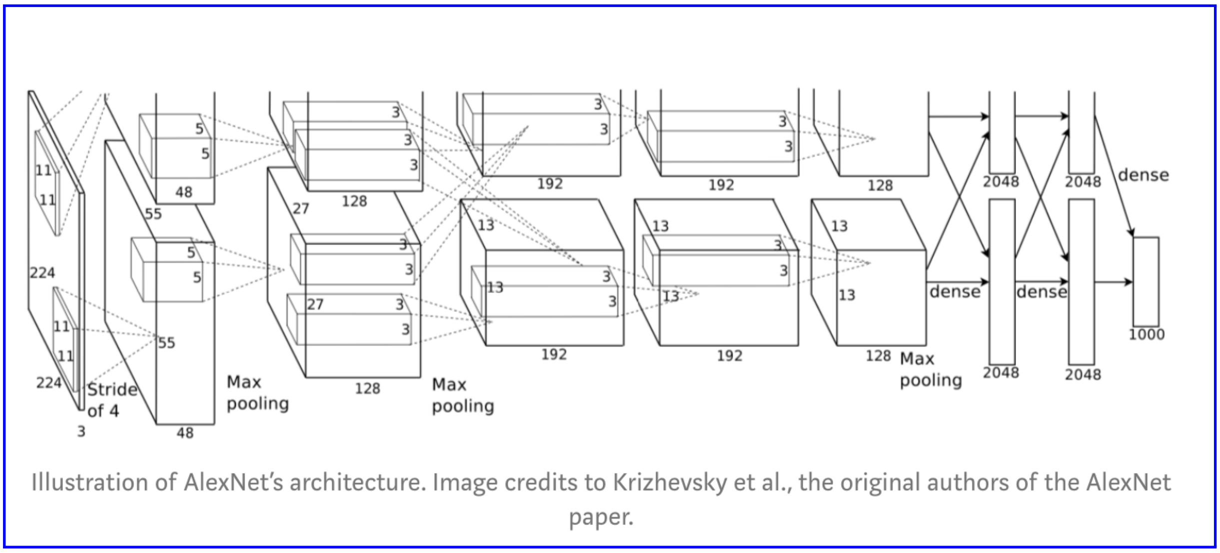

Figure 14: AlexNet Architecture

framework is composed of a total of 8 layers, where five layers are the convolutional layers and the other 3 are the fully-connected layers. AlexNet uses the Relu activation function for the five convolutional layers. Relu is used as an activation function so that the model takes less time to train, it is evident from the research by Krizhevsky et al. (2012) that Relu as an activation function with CNN works six times faster than a CNN using the function "tanh". To reduce the overfitting problem for the model the dropout function with a predetermined 0.5 probability was used in the fully connected layers. This does affect the training time a little. Data augmentation was performed to create variety in the data for the model to train well where horizontal flips, vertical flips, re-scaling, and zooming was performed using the ImageDataGenerator.

4.3 Implementation using LeNet

Another implementation of the CNN networks is the LeNet architecture which was developed by LeCun et al. (1998). Figure 15 shows the architectural design of LeNet basic structure. This model is used for the project which is having seven layers, out of which the first 3 layers are the convolutional layers, two are the subsampling layers and the last 2 layers are fully-connected layers.

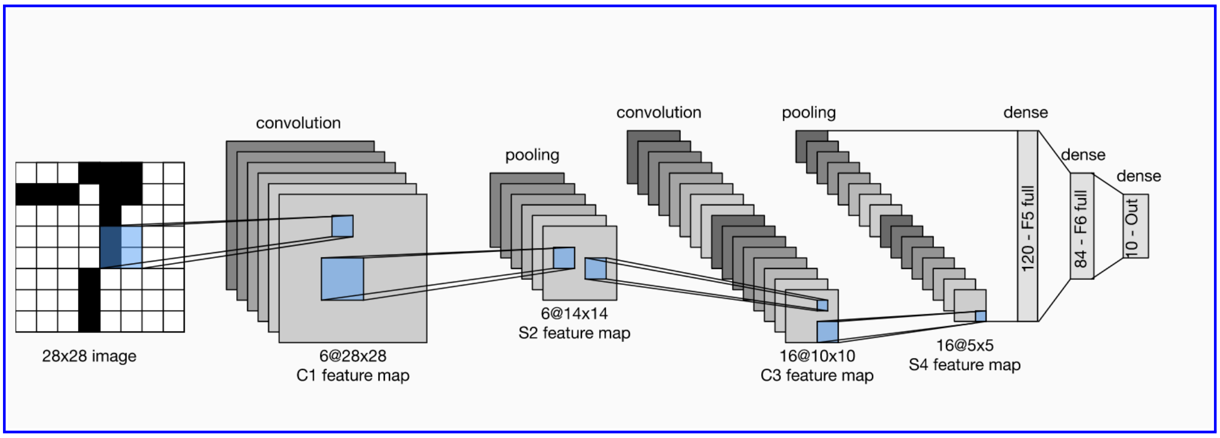

Figure 15: LeNet Architecture

For this project, the LeNet model's first layer i.e. Input layers is designed to accept the image input of size 224x224x3. These input images are transferred to the next layer. The first layer produces an output of 20 features maps with a kernel of size 5x5. The kernel is the window that contains the value of weights that are used while convolution of weight values with input values. Again the Relu activation function is used for creating a non-linearity in the network and for making the output normalize. It provides the ability for the model to learn more complex functions. The softmax activation function is used in the output layer. The rest of the architecture is similar to the basic architecture.

4.4 Implementation using CNN

A custom convolutional neural network was built where there are three convolutional layers and three max-pooling layers which take an input image of size 224x224x3. The activation function Relu is used. There is one flattening layer and 2 dense layers with activation function as Relu and Sigmoid. The optimizer used for this model is Adam and the loss function is selected as binary_crossentropy. The metrics used for the evaluation is accuracy. Image Data-generator is used for augmentation of the images so that the model gets varied images to get trained with complex features. Then accuracy and loss for the validation dataset are performed.

4.5 Implementation using ResNet50

ResNet50 is used as a pre-trained model for transfer learning feature extraction. To incorporate Transfer Learning, the last forecasting layer of the already-trained ResNet50 model has been eliminated and replaced with custom layers for prediction. Pre-trained model weights are fixed, and are not changed during training. The last fully connected layers to load should not be loaded as that serves as the classifier. By using "include top= False," it was achieved. It was necessary as we can add a new fully-connected layer to our task-specific identification alongside the ResNet50 model. By setting the trainable as false the weights were frozen. This avoids any changes to the pre-trained weights during training. We do not want to train ResNet layers because we want to use the knowledge gained by the deep neural network trained from the previous dataset of ImageNet. The image of input size 224x224x3 was passed as an input to the model. Then, the image data generator was used for training and the testing set with vertical and horizontal flips, rotation of the image to 20 degrees, re-scaling, and zooming the images to create varied images to train the model. The model was trained for 50 epochs.

5 Evaluation

The performance of the models based on the accuracy and loss in the training and testing sets has been explained briefly in this part of the project report. Five different models of CNN and transfer learning were used for completing the research project.

5.1 MobileNet Accuracy-Loss Evaluation

The MobileNet model was trained on 835 images and was evaluated on testing with 333 images. Here the image data-generator was used to create a varied amount of images for the model to train well. The model was trained on 50 epochs. Figure 16 shows the Training and validation accuracy and figure 17 shows the Training and validation loss per epoch. From figure 16 it can be seen that as the number of epochs increases, the validation accuracy increases and reaches its peak of 75.38% at 49th epoch. There is a lot of fluctuation between the 9th and 49th epoch but the accuracy remains over 65%. Also, the model here is overfitting as there is a huge difference between the training and validation loss.

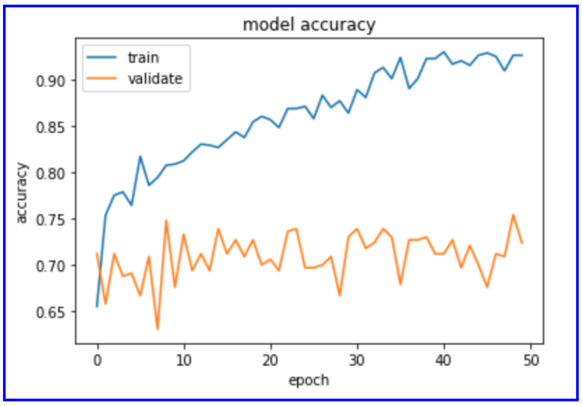

Figure 16: MobileNet Epoch VS Accuracy

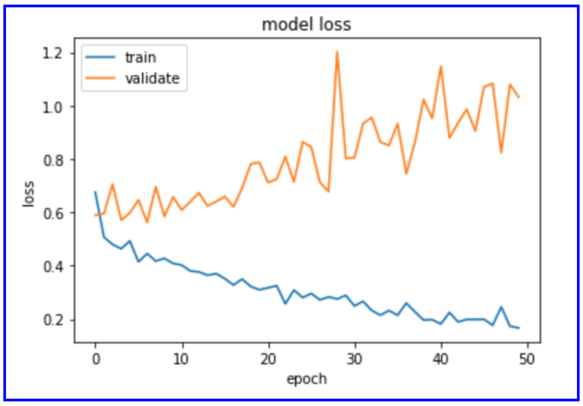

Figure 17: MobileNet Epoch VS Loss

Figure 17 shows that as the number of epochs increases there is a gradual increase in the validation loss. The model achieves the lowest loss of 0.56 at epoch 7th. The model is optimized using Adam optimizer.

5.2 AlexNet Accuracy-Loss Evaluation

AlexNet was used for the implementation of the classification of melanoma. From figure 18 it is can be seen that the accuracy for the validation set is stagnant at 53% till the 17th epoch, the there is a lot of fluctuation of accuracy from 25th epoch to 40th epoch.

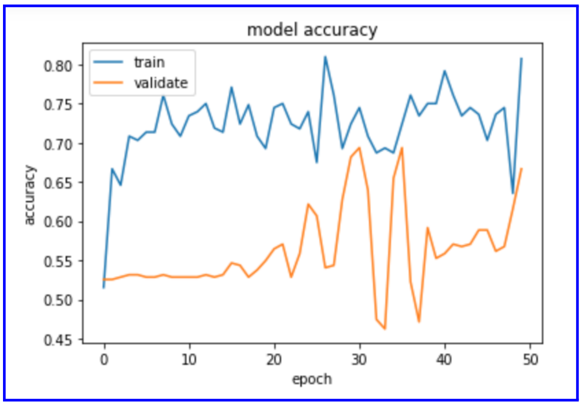

Figure 18: AlexNet Epoch VS Accuracy

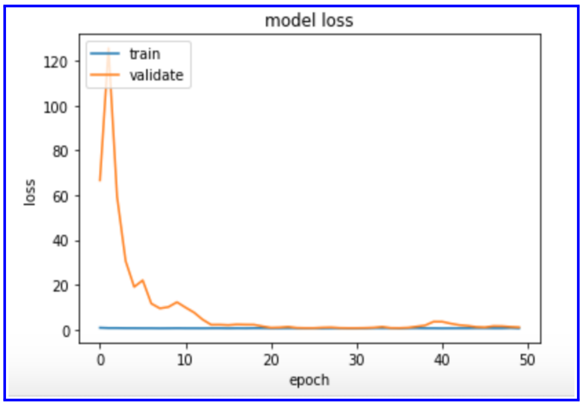

Figure 19: AlexNet Epoch VS Loss

At the 50th epoch, the validation accuracy reaches to 67%. While from figure 19 it is evident that the loss for the training set remained at 0 while but for testing/validation set after the 10th epoch it remained between 0 to 1 for rest of the epochs.

5.3 LeNet Accuracy-Loss Evaluation

LeNet model performed well for the classification task. From figure 20 it is evident that with the increase in epochs the accuracy is increasing steadily.

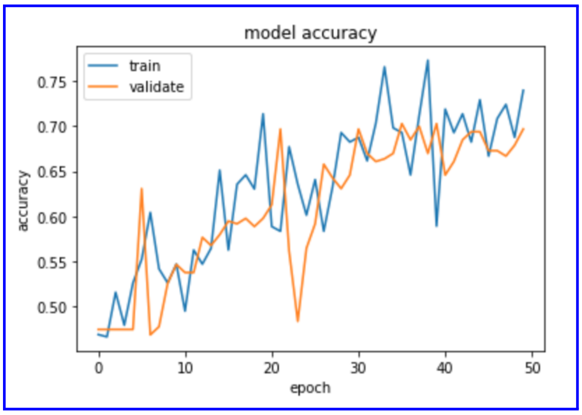

Figure 20: LeNet Epoch VS Accuracy

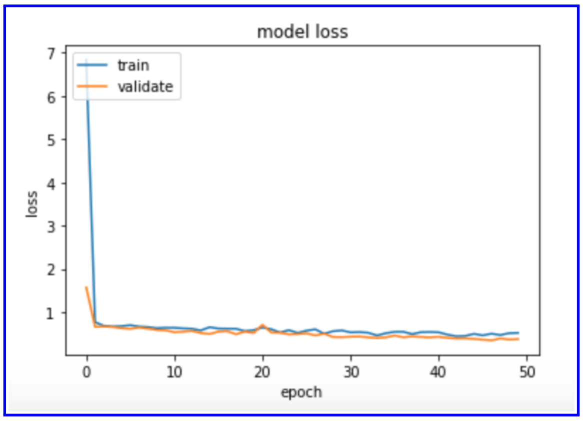

Figure 21: LeNet Epoch VS Loss

The model achieved an accuracy of 70% at the 22nd epoch and decreased sharply to reach 48% at the 24th epoch and then stays steady thereafter with a minor increase and decrease. While from figure 21 it is self-explained that there is a validation loss of 0.6 which is stagnant for every epoch. There is no or minor change in the loss at each epoch.

5.4 CNN Accuracy-Loss Evaluation

CNN was chosen as one of the models for the classification of melanoma as from the past research work it was considered to be one of the effective models in image classification projects. From figure 22 it is can be seen that with the increase in the epochs the accuracy

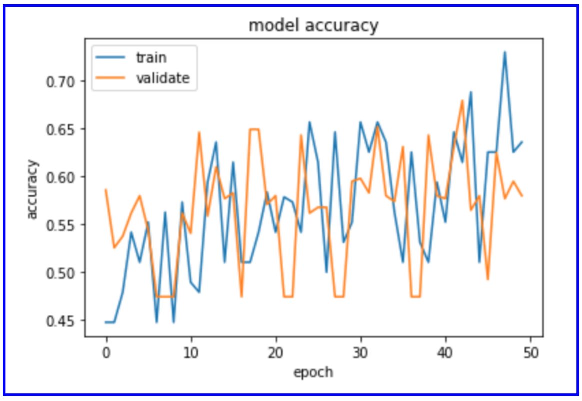

Figure 22: CNN Epoch VS Accuracy

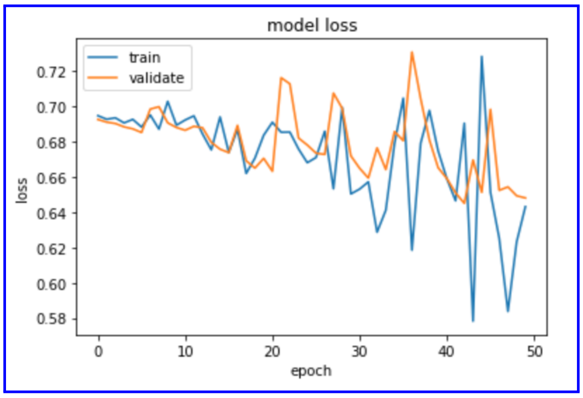

Figure 23: CNN Epoch VS Loss

is increasing and decreasing with lots of fluctuations in between. The model's accuracy was fluctuating as the number of epochs was increasing. The mean accuracy achieved on the validation set is 56.40%. While from figure 23 it is evident that the loss remained at 0.7 till 10 epochs and then started to reduce sharply thereafter reaching a minimum of 0.5 at 47th epoch.

5.5 ResNet50 Accuracy-Loss Evaluation

ResNet50 was used as a transfer learning model for the implementation of the classification of melanoma. From figure 24 it is clear that the validation set best achieves its accuracy of 60.66 at 37th epoch. Similarly, from figure 25 it is evident that minimum loss achieved is 0.66 at the 49th epoch.

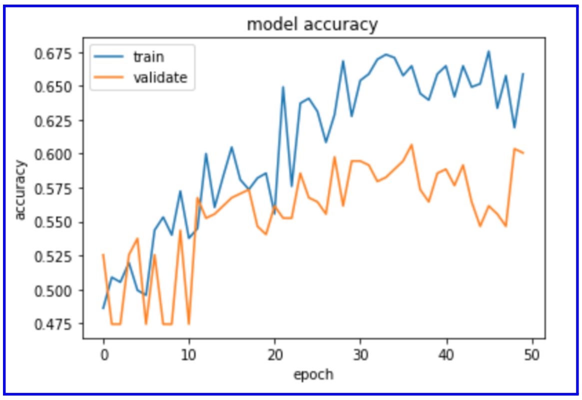

Figure 24: ResNet50 Epoch VS Accuracy

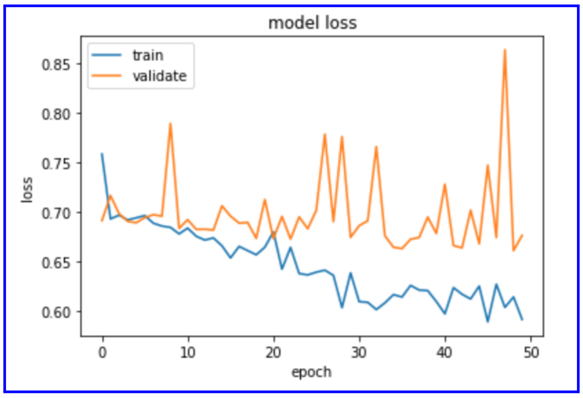

Figure 25: ResNet50 Epoch VS Loss

6 Conclusion and Future Work

Melanoma classification has been one of the active areas which still requires more research. The research done in the past provided using traditional, machine learning, neural networks provided great results on the melanoma classification. The most important findings were the symmetry and semantics of the malignant and benign class of melanoma images. Different image pre-processing techniques were used to remove unwanted artifacts from the images which improved the model's performance. Feature extraction techniques were also used to extract important features from the image. This project made image preprocessing and implemented five different transfer learning and Neural Network models and compared their results. The deep neural network and transfer learning models do not require the feature extraction process to be done separately before feeding the images to the models as this is being handled by the models internally.

Out of five different models implemented the transfer learning models had an issue of overfitting the data. This could be due to the downsampling of the dataset, during which much of the information regarding the benign cases had been a loss. Though the accuracy on the testing set for the mobile model was achieved to be around 75% and for Resnet50 it was around 60.66%. The Neural Network models like AlexNet, LeNet, and CNN achieved an accuracy of 67%, 70%, and 56.40%. MobileNet and LeNet were the two models that performed best on the dataset.

The future work will focus on dealing with the highly imbalanced dataset without the use of oversampling and downsampling techniques. Also, the focus will be on hyper tuning the parameters as well as using the image set with high-resolution images and applying different augmentation techniques for creating the varied images. More complex models will be applied like RetinaNet and CapsuleNet.

References

Adegun, A. A. and Viriri, S. (2019). Deep learning-based system for automatic melanoma detection, *IEEE Access* **8**: 7160–7172.

Adil, A., Iskandar, D., Al-Asad, J. F. and El-Nakla, S. (2020). Classification of skin lesion with hair and artifacts removal using black-hat morphology and total variation, *International Journal of Computing and Digital Systems* **10**: 2–8.

Ahmad, B., Usama, M., Huang, C.-M., Hwang, K., Hossain, M. S. and Muhammad, G. (2020). Discriminative feature learning for skin disease classification using deep convolutional neural network, *IEEE Access* **8**: 39025–39033.

Aishwarya, U., Daniel, I. J. and Raghul, R. (2020). Convolutional neural network based skin lesion classification and identification, *2020 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, pp. 264–270.

Alfed, N., Khelifi, F., Bouridane, A. and Seker, H. (2015). Pigment network-based skin cancer detection, *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, IEEE, pp. 7214–7217.

Alquran, H., Qasmieh, I. A., Alqudah, A. M., Alhammouri, S., Alawneh, E., Abughazaleh, A. and Hasayen, F. (2017). The melanoma skin cancer detection and classification using support vector machine, *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, IEEE, pp. 1–5.

Ballerini, L., Fisher, R. B., Aldridge, B. and Rees, J. (2012). Non-melanoma skin lesion classification using colour image data in a hierarchical k-nn classifier, *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE, pp. 358–361.

Bumrungkun, P., Chamnongthai, K. and Patchoo, W. (2018). Detection skin cancer using svm and snake model, *2018 international workshop on advanced image technology (IWAIT)*, IEEE, pp. 1–4.

Chugh, R. S., Bhatia, V., Khanna, K. and Bhatia, V. (2020). A comparative analysis of classifiers for image classification, *2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pp. 248–253.

Goyal, M., Oakley, A., Bansal, P., Dancey, D. and Yap, M. H. (2019). Skin lesion segmentation in dermoscopic images with ensemble deep learning methods, *IEEE Access* **8**: 4171–4181.

Guy Jr, G. P., Thomas, C. C., Thompson, T., Watson, M., Massetti, G. M. and Richardson, L. C. (2015). Vital signs: melanoma incidence and mortality trends and projections—united states, 1982–2030, *MMWR. Morbidity and mortality weekly report* **64**(21): 591.

Hoshyar, A. N., Al-Jumaily, A. and Sulaiman, R. (2011). Review on automatic early skin cancer detection, *2011 International Conference on Computer Science and Service System (CSSS)*, IEEE, pp. 4036–4039.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications, *arXiv preprint arXiv:1704.04861* .

Jain, Y. K. and Jain, M. (2012). Skin cancer detection and classification using wavelet transform and probabilistic neural network.

Jogin, M., Mohana, Madhulika, M. S., Divya, G. D., Meghana, R. K. and Apoorva, S. (2018). Feature extraction using convolution neural networks (cnn) and deep learning, *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pp. 2319–2323.

Kamboj, A. et al. (2018). A color-based approach for melanoma skin cancer detection, *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, IEEE, pp. 508–513.

Karargyris, A., Karargyris, O. and Pantelopoulos, A. (2012). Derma/care: An advanced image-processing mobile application for monitoring skin cancer, *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, Vol. 2, IEEE, pp. 1–7.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097–1105.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE* **86**(11): 2278–2324.

Linsangan, N. B., Adtoon, J. J. and Torres, J. L. (2018). Geometric analysis of skin lesion for skin cancer using image processing, *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, IEEE, pp. 1–5.

Lynn, N. C. and Kyu, Z. M. (2017). Segmentation and classification of skin cancer melanoma from skin lesion images, *2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, IEEE, pp. 117–122.

Mane, S. and Shinde, S. (2018). A method for melanoma skin cancer detection using dermoscopy images, *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, IEEE, pp. 1–6.

Masood, A. and Al-Jumaily, A. (2017). Semi advised learning and classification algorithm for partially labeled skin cancer data analysis, *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, IEEE, pp. 1–4.

Maurya, R., Singh, S. K., Maurya, A. K. and Kumar, A. (2014). Glcm and multi class support vector machine based automated skin cancer classification, *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, pp. 444–447.

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

Nezhadian, F. K. and Rashidi, S. (2017). Melanoma skin cancer detection using color and new texture features, *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, IEEE, pp. 1–5.

Sadeghi, M., Lee, T. K., McLean, D., Lui, H. and Atkins, M. S. (2013). Detection and analysis of irregular streaks in dermoscopic images of skin lesions, *IEEE transactions on medical imaging* **32**(5): 849–861.

Setiawan, A. W. (2020). Effect of color enhancement on early detection of skin cancer using convolutional neural network, *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, IEEE, pp. 100–103.

Siegel, R. L., Miller, K. D. and Jemal, A. (2019). Cancer statistics, 2019, *CA: A cancer journal for clinicians* **69**(1): 7–34.

Suganya, R. (2016). An automated computer aided diagnosis of skin lesions detection and classification for dermoscopy images, *2016 International Conference on Recent Trends in Information Technology (ICRTIT)*, IEEE, pp. 1–5.

Vinay, B., Shah, P. J., Shekar, V., Vanamala, H. et al. (2020). Detection of melanoma using deep learning techniques, *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, IEEE, pp. 391–394.