

IMAGE CLASSIFICATION OF SKIN DISEASES USING DEEP LEARNING

MONICA DAYALANI AMARNANI



A REPORT SUBMITTED AS PART OF THE REQUIREMENTS FOR THE DEGREE
OF BSc IN COMPUTER SCIENCE
AT THE SCHOOL OF COMPUTING
ROBERT GORDON UNIVERSITY
ABERDEEN, SCOTLAND

May 2021

Supervisor Dr. Eyad Elyan

Abstract

Deep learning is so powerful that it imitates the workings of a human brains. For this project, we used the approach of deep learning to detect different type of skin diseases. Here we make use of convolutional neural network based models as well as pre-trained models such as VGG16 and ResNet50 and we push the performance of these models for them to have good metrics score. Although, this is just the beginning as there is a lot to work on for this project to be used on diagnosis for patients.

Acknowledgements

A special thanks to my supervisor for the guidance, support and advice during this project. I would also like to thank the RGU staff. Lastly to my family for all the support and love they have shown in the past five years.

Declaration

I confirm that the work contained in this BSc project report has been composed solely by myself and has not been accepted in any previous application for a degree. All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.

Signed Monica Dayalani Amarnani

Date 04/05/2021

Contents

Abstract	2
Acknowledgements	3
Declaration	4
1 Introduction	1
1.1 Background	1
1.2 Aims and Objectives	2
1.3 Chapter List	2
2 Literature Review	3
2.1 Outline	3
2.2 Diagnosis & Machine Learning techniques	3
2.3 The recent development (state-of-the-art)	4
2.4 Medical Images	6
2.5 Related to Datasets	6
2.5.1 Skin Diseases Classification using image processing.	7
2.5.2 Results	7
2.6 Computer-Aided Model for skin Diagnosis using Deep learning	11
2.6.1 Results	12
2.7 Analysis Related to Datasets	15
2.8 Summary and Moving Forward	15
3 Requirement Analysis	17
3.1 Functional Requirements	17
3.1.1 Datasets	17
3.1.2 Method	17
3.1.3 Methods comparison	18
3.1.4 Model Improvement & Testing	18

3.2	Non-Functional Requirements	18
3.2.1	Technologies	18
3.2.2	Legal, Social & Ethical Issues	18
3.3	The Application Front End	18
3.3.1	Front End	18
3.3.2	Technologies - Programming Languages and Libraries	18
4	Design & Experiment Setup	19
4.1	Hardware and Software Setup	19
4.2	Dataset	19
4.2.1	Data Pre-Processing	20
4.2.2	Augmentation	22
4.3	Imbalanced dataset and Over the accuracy	22
5	Experiments and Results	25
5.1	Experiments 1 & 2	25
5.2	Experiment 3 & 4	29
5.3	Transfer Learning	31
5.3.1	Experiment 7 & 8	32
5.4	Failed Experiments	33
5.5	Summary of Results	34
6	Evaluation & Conclusion	36
6.1	Overview	36
6.2	Conclusion	37
6.3	Future Work	37
	References	38
A	Project Poster	40
B	Project Proposal & Ethics Form	42
C	Project Log	54
D	Source Code	56

List of Figures

2.1	Showing the process of machine learning image classification using the definition feature/ feature extraction. (Image taken from TowardsData-Science)	4
2.2	LeNet-L convolutional neural network architecture	5
2.3	Design for Multi-layered Perceptron, image taken from https://alexlenail.me/NN-SVG/index.html	5
2.4	Design of Lenet-5, image taken from https://alexlenail.me/NN-SVG/index.html	6
2.5	Cropping to get the region of interest of affected skin	7
2.6	Confusion matrices before fine tune	8
2.7	Confusion matrices after fine tune	8
2.8	Computer Vision Results	9
2.9	Machine Learning Results	9
2.10	The architecture of the system	10
2.11	Feed forward back propagation artificial neural network diagram	11
2.12	Feed forward back propagation artificial neural network diagram	11
2.13	Noises removed from the image.	12
2.14	Melanoma and non-melanoma skin lesions from Dermis Dataset	13
2.15	Melanoma and non-melanoma skin lesions from DemQuest Dataset	13
2.16	: Melanoma Basic cell carcinoma, eczema, and impetigo from Dermnet dataset	13
2.17	Results of the different architectures.	14
4.1	The classes and images in train set.	20
4.2	The classes and images in train set.	20
4.3	The classes and images in train set.	21
4.4	Unrelated images	21
4.5	Augmented Image	22
4.6	Augmented Image	22

4.7	Micro F1 score formula	23
4.8	Confusion Metrics	24
5.1	Model Architecture for Experiment 1 and 2	26
5.2	Accuracy Plot Exp 1	27
5.3	Loss Plot Exp 1	27
5.4	Confusion Matrix Exp 1	27
5.5	Accuracy Plot Exp 2	28
5.6	Accuracy Plot Exp 2	28
5.7	Confusion Matrix Exp 2	28
5.8	Model Architecture	29
5.9	Accuracy Plot Exp 3	30
5.10	Loss Plot Exp 3	30
5.11	Accuracy Plot Exp 4	30
5.12	Loss Plot Exp 4	30
5.13	Confusion Matrix Exp 4	31
5.14	VGG16 Architecture	31
5.15	ResNet50 Architecture	31
5.16	Accuracy plot Exp 7	32
5.17	Accuracy Plot Exp 8	33
5.18	Loss Plot Exp 8	33
5.19	Confusion Matrix Exp 8	33
5.20	Accuracy plot of 23 Classes	34
5.21	Accuracy Plot Exp 5	34
5.22	Accuracy Plot Exp 6	34
5.23	Accuracy Plot Exp 10	34

Chapter 1

Introduction

1.1 Background

Medical images Classification is a very important research topic for health and for machine vision. It has attracted extensive research efforts in the last decade. Dermatology is a sector of medicine that uses medical imaging to diagnose patients. The types of medical imaging range from X-rays, CT (computed tomography) scan, MRI (magnetic resonance imagining), ultrasound, nuclear medicine imaging. Furthermore, some solution has been found to classify skin diseases using machine learning algorithms which is SVM (Support Vector Machine) such as skin cancer Melanoma [1]. Convolutional Neural Networks has advanced masses where lot of research has been published in image recognition for medical scans in the past years. This thesis portrays the applying of Convolutional Neural Network algorithm to diagnose skin disease in medical imaging. This projects aim is whether a medical image can predict a skin disease. Skin disease is one of the most common illnesses in human daily life. It pervades all cultures, happens in any respect ages, and influences among 30 The work done for the thesis will be using a publicly available dataset for skin diseases which will include images of normal patients and patients with some skin problem. The images in this dataset are taken from the public portal which is the largest dermatology source. Because of a shortage of dermatologists, most cases are seen instead by general practitioners with lower diagnostic accuracy. A new deep learning system could diagnose around 16.000 de-identified skin conditions from tele dermatology. Although 80mentioned previously due to the shortage of trained dermatologists the NHS has just seen its waiting list growing and patients waiting for more than 6 weeks to have medical imaging perform on them. The importance of this project is to demonstrate that with deep learning techniques, medical images can be successfully diagnosed at an acceptable rate

1.2 Aims and Objectives

Aim: the aim of the thesis is to predict whether a patient has skin disease based on medical images using deep learning models.

Objective 1: Literature review on medical imaging, deep learning, convolutional neural network and methodologies.

Objective 2: Implement technologies to build machine learning workflow.

Objective 3: Test and evaluate the model performance.

Objective 4: Validate the algorithm using the existing data.

Objective 5: Find improvements for the existing technologies.

1.3 Chapter List

Chapter 2 Literature Review This chapter reviews literature from previous papers related with the project aims and objectives.

Chapter 3 Requirement Analysis This chapter is about the specification we need to meet and the specification that will help improve this project.

Chapter 4 Design & Experiment Setup. This chapter has specification on the hardware and software implementation as well as on the dataset and data cleaning. .

Chapter 5 Experiments & Results. This chapter gives specific information on the experiments carried out throughout this paper.

Chapter 6 Evaluation & Conclusion. This chapter evaluates the overall project and the conclusion of the thesis is presented.

Chapter 2

Literature Review

This chapter evaluates the literature that is related to this project aims and objectives. In particular, it will review the relevant literature around deep learning and the different types of machine learning that have advanced during the years. Apart from this, it will review previous projects that have been successful and are linked to the thesis.

2.1 Outline

Medical imaging technology is mainly used by Radiologists and in this case by dermatologists as well. Radiologists are medical doctors that specialize in diagnosing and treating injuries and diseases using medical images procedures within the human body and dermatologist are skin doctors that also require medical images procedure to diagnose patients. There are various types of images with vary between CT scan (computed tomography scan) ultrasounds and X-rays [2]. Furthermore, a big issue within the radiology department is that they are need in staff to produce the scans. This causes the patients at risk since the NHS does not have sufficient radiologists to meet the diagnostic imaging obligations [3][4].

2.2 Diagnosis & Machine Learning techniques

Medical images are classified within image classification. From past few years, image classification is enhanced by using the tools of deep learning and convolutional neural networks. The process of image classifications implies teaching a model to understand the features form the images and the test on unseen images to compute accuracy [5]. Machine learning has developed masses during the last years but prior to having the expansion into deep learning other methods of image classification were applied. As

shown in Figure 2.1, feature extraction/defining the features was the key step of the machine learning method before creating any model, in this case before classifying any data.

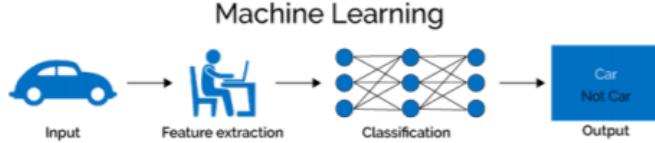


Figure 2.1: Showing the process of machine learning image classification using the definition feature/ feature extraction. (Image taken from TowardsDataScience)

Furthermore, to a computer, images are a number, in this case, they are defined as pixel and they can be represented in a 2D matrix of pixels values. To build the machine learning pipeline we can see in the image above that classification is the key, so, in order to classify an image the machine learning pipeline needs to be able to tell what is unique about each image. Classification can also be taught of involving detection of the features in a given image and opting if a feature is for a class label are present in an image we can then predict that the particular image is of that class with a higher probability [6].

2.3 The recent development (state-of-the-art)

Convolutional Neural Networks be an artificial neural network that has a specialization for being able to pick out or detect a pattern and <https://www.overleaf.com/project/5fbfb99cb1d164014099b1set> a prediction. Also, CNN are a unique type of Multi layered Perceptron (MLP). MLP utilizes a supervised learning technique called backpropagation for training. CNN differs from MLP, where CNN can understand spatial relation (relation between nearby pixels of image) between pixels of images therefore, CNN will perform better than MLP. Whereas MLP will perform good for simple image classification. The differences are seen in the design which are shown in Figure 2.2 and Figure 2.3. Convolutional Neural Networks was originally used for identification of handwriting but has been also used for image and speech recognition [7] , text detection and recognition [8], object tracking this means it has been set to detect different type of patterns [9]. Moreover, Figure 2.2 is fully connected layer as the image low-level feature by training CNN. Extracting this image feature is a part of the image retrieval feature. This process of extracting the image features by the CNN is the continuous learning process of all the images in the image dataset and deep learning network architecture is process of extraction and abstraction layer by layer to the original image data. The high-level feature is

more abstract and more efficient semantic features, it can be better able to express the information contained in the image [8]. Commencing with the figure 2.4, convolutional neural networks have structure of convolutional layers that are followed by ReLU operation, max-pooling and then a fully connected layer. There are alternatives of this design in different image classification tasks. One of them is ImageNet that is the one that has stood out the most, this had 1.2 million high-resolution images that were put in 1000 different classes. On the test data, it was achieved top-1 and top-5 error rates of 37.5% and 17.0% which takes it to be better than the previous state of art [10].

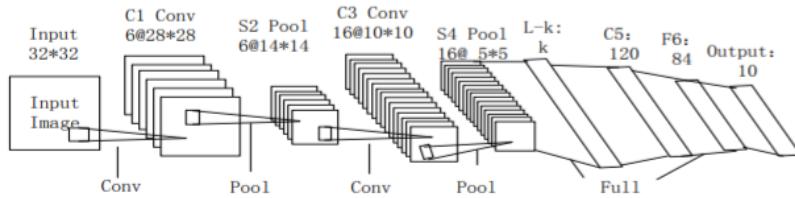


Figure 2.2: LeNet-L convolutional neural network architecture

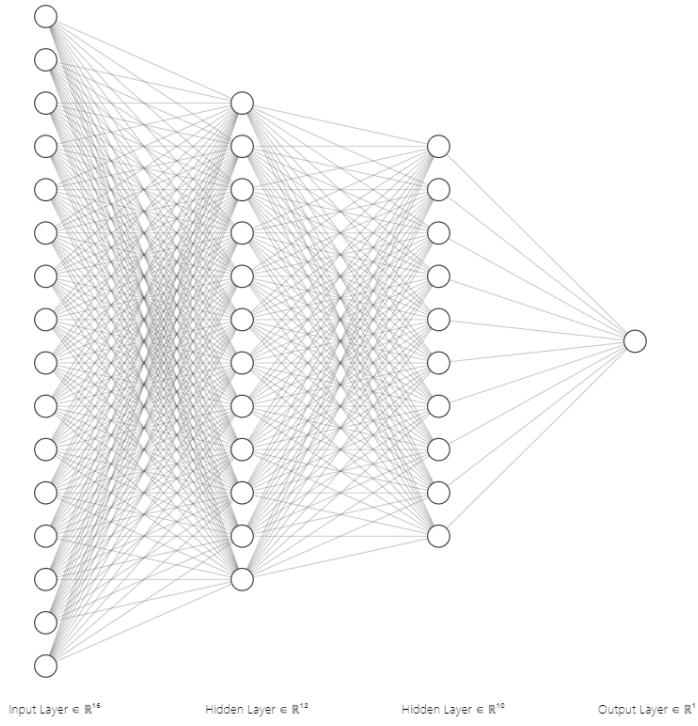


Figure 2.3: Design for Multi-layered Perceptron, image taken from <https://alexlenail.me/NN-SVG/index.html>

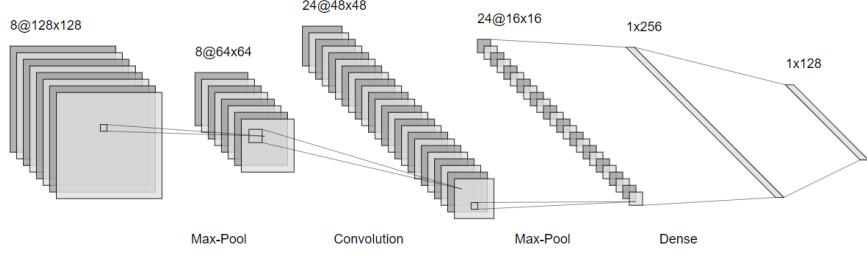


Figure 2.4: Design of Lenet-5, image taken from <https://alexlenail.me/NN-SVG/index.html>

2.4 Medical Images

As mentioned in the outline, the dermatology and radiology department in the UK had not enough dermatologists' professionals to attend the large number of patients or radiologist to meet imagining and diagnostic demands warned by the Royal College of Radiologists. Dermatologists treat over 2000 types of skin conditions, each year over 54% of the population is affected. Over the recent years skin cancer has been increasing which takes dermatologist and radiologist to be needed. The paper has also cited that the British Association of Dermatologist said that there were 813 dermatology specialists which include consultants, trainees, associates and a total of 729 consultant posts, 75 of which were vacant and 98 of which were occupied by locums. For such big population (61.8m), the level of consultants would result in 989 full time recommended the royal college. Furthermore, when coming to radiologists 97% agreed that they were not able to meet their diagnostic reporting during their working hours. This was mentioned in the paper "This points to an insufficient number of radiologists to meet the increasing demand for imaging and diagnostic services". This report also cites that there was a particular shortage in Scotland where consultant workforce grew a 7% during the years but demand on CT, MRI and other scans went up by 10%. This kind of impacts of shortage risks the patient's treatment.

Moreover, in Radiology and in medical imagining CAD had become a major research topic since with CAD radiologist can use it as a second opinion and to make final decision, in fact CAD systems have been used to assist physicians in the detection of cancer such as mammograms [11].

2.5 Related to Datasets

In the domain of machine learning in medical image analysis, there have been a vague number of successful papers published on every type of skin disease. The first step is to

consider a dataset that would help to be used to construct a predictive model. In this case, the model could predict whether there is an existing disease or not. Furthermore, in this section we will review work related to the thesis title.

2.5.1 Skin Diseases Classification using image processing.

In this paper, researchers classify diseases of 10 different classes such as Herpes, warts, Melanoma skin cancer, this contains 5500 images obtained from the Dermnet dataset. These images are based on skin diseases that can look different in different parts of the body, hence, to prevent this it should be taken in count at the beginning of the disease. Thus, their proposed system consisted of image processing that deals with image augmentation (this is because deep learning algorithm learns best when it has access dealt to data) and removal of unwanted elements as observed on Figure 2.5 (will allow to extract the feature in a better manner) as well as transfer learning of training of dermatological images [12]. Transfer training deals with feature extractions and fine tuning of pretrained convolutional neural network VGG16 model.

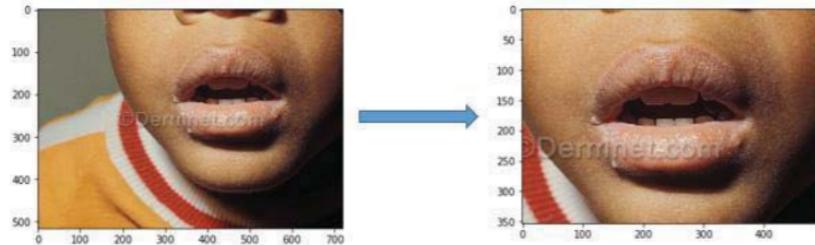


Figure 2.5: Cropping to get the region of interest of affected skin

2.5.2 Results

Results showed that after training the CNN with Dermnet dataset images. Pre-trained weights were initialised from a model by using ImageNet pretrained models of VGG16. Moreover, in Figure 2.6 and 2.7 we can observe the 2 confusion matrices in which the colour of each cell is representing the probability of the prediction, the number on each cell gives the occurrences of the prediction. The higher diagonal values of the confusion matrix indicate many correct predictions. We can see from the two figures the high confidence predictions are located along the diagonal. In conclusion the confusion matrix of the new CNN found to have higher values in its leading diagonal elements.

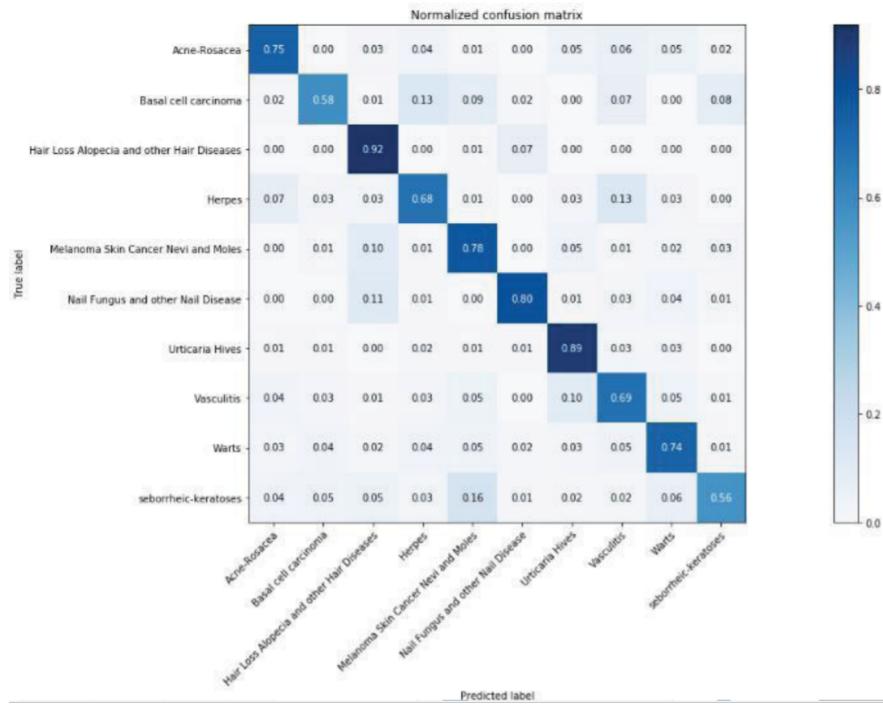


Figure 2.6: Confusion matrices before fine tune

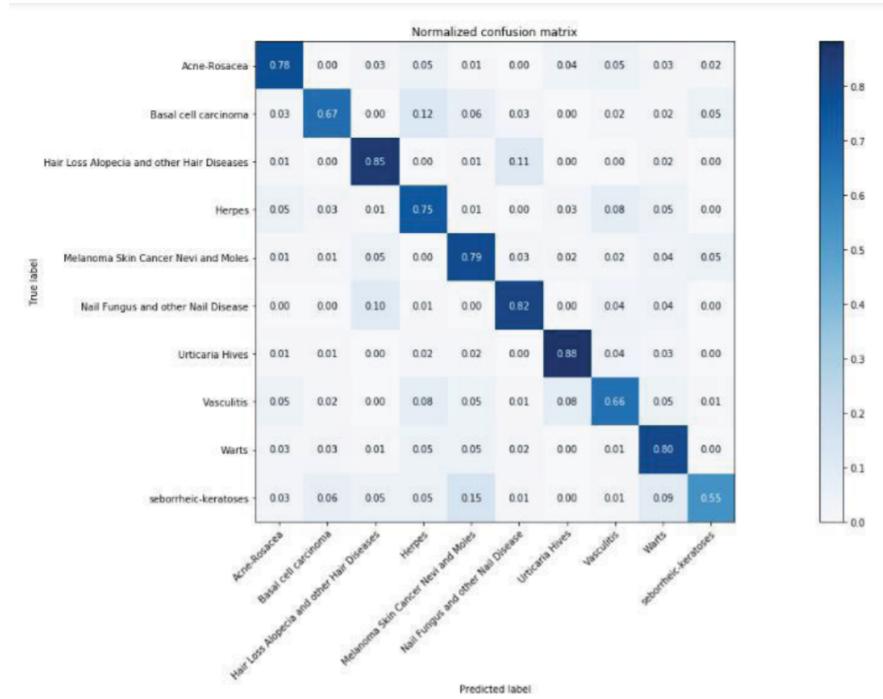


Figure 2.7: Confusion matrices after fine tune

On the other hand, there is a similar paper where various experiments were undertaken to show the performance of the model [13]. This research uses a dual stage approach which combines computer vision and machine learning. Computer vision has two phases. In the first phase is image processing, and in the second phase is making use of the features extracted from the image processing and identifying the disease using multiple algorithms such as Maximum Entropy Model (MaxEnt) and Artificial neural network (ANN). Moreover, in the stage 2 they use various machine learning techniques. This stage of prediction is made available to medical professionals who have access to histopathological attributes such as exocytosis or hyperkeratosis. The attributes are taken as inputs in the system from the user and gives a better classification of the disease. The testing of the system is done using Decision Trees, neural networks and KNN (Kth Nearest Neighbour) Model. This paper is comparing six diseases, them using both approaches has given an accuracy up to a 95%. In Figure 2.8 and 2.9 we can look at the results computer vision and machine learning has given. In Figure 2.8 using the Maximum Entropy model (MaxEnt) allowed them to get a better accuracy whereas in Figure 2.9, non-linear models like ANN and Decision Trees learns the underlying pattern and gives better accuracy.

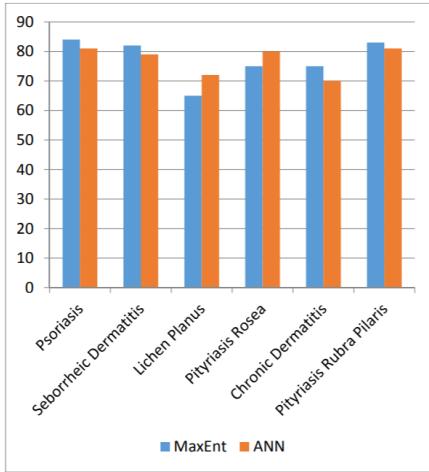


Figure 2.8: Computer Vision Results

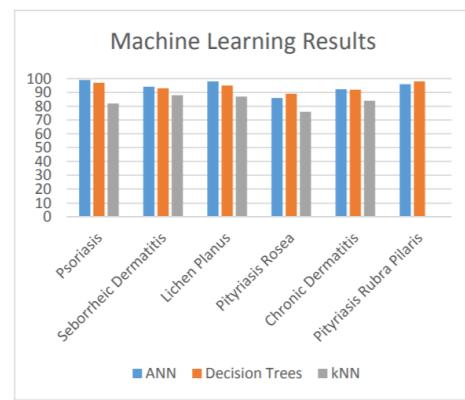


Figure 2.9: Machine Learning Results

There are many ways to detect a skin disease, in this paper similar to the two above the detection of dermatological disease using image processing and Artificial Neural Network is done [14]. In this paper, the purpose is to build an automated skin disease detection system. Here, the detection of nine different types of skin disease conducted. There are 8 different types of algorithms used such as YCbCr, grey image, sharpening filter, etc. for image pre-processing to find visual pattern and significant features like average colour code of infected area or infected area size in case of pixels and shape. The system will take 10 different features from the image pre-processing results and

user inputs that are listed in Figure 2.10. Then, these user inputs values are trained along the colour skin image extracted features to train and test into a feed forward back propagation ANN to identify dermatological disease. We can observe this process in the Figure 2.10.

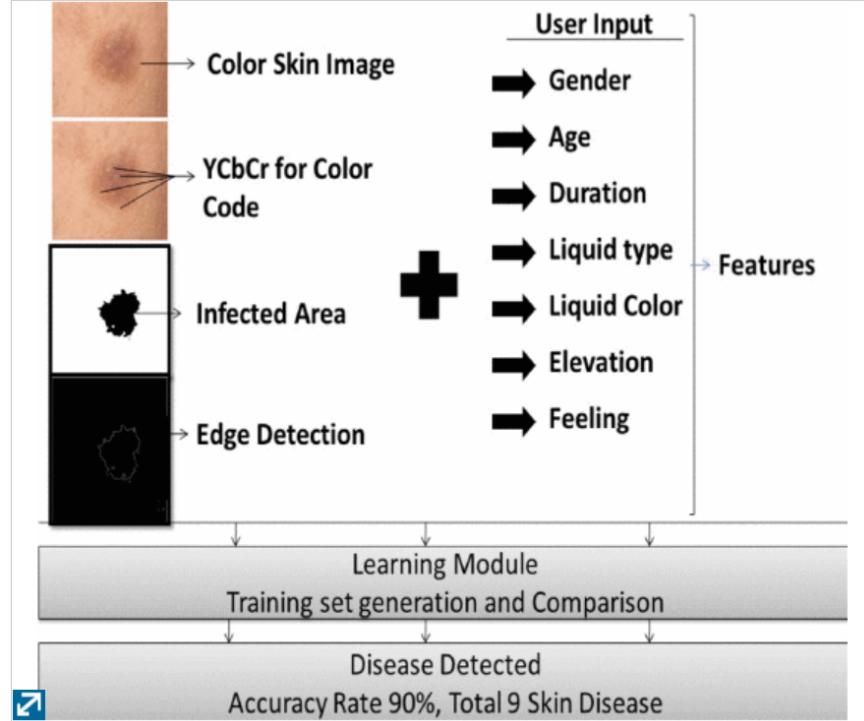


Figure 2.10: The architecture of the system

As mentioned above, feed-forward back-propagation neural network training is used. The validation and testing of the system is done by using the tenfold cross validation. This is used since there are no overlapping of the test data and training data, which makes the testing results more viable and dependable. In Figure 2.11 we can observe an example of this.

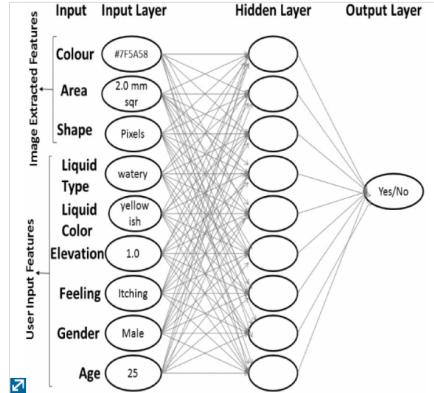


Figure 2.11: Feed forward back propagation artificial neural network diagram

Lastly, this system detected 9 different dermatological diseases with an accuracy of 90%. The supervised system was more accurate than the rest. The supervised system is at 90% of accuracy whereas semi-supervised at 88% and unsupervised at 85%. In Figure 2.12 we can see different detection rate for 9 different diseases. In the detection rate of diseases that have low elevation in the infected area like psoriasis, foot ulcer and vitiligo are very high like 91%, 97% and 97%, whereas the detection rate that have high elevation in the infected area comparatively low that is around 85-88%. Out of all these diseases the system shows the best accuracy rate for foot ulcer and vitiligo that are to a 97%.

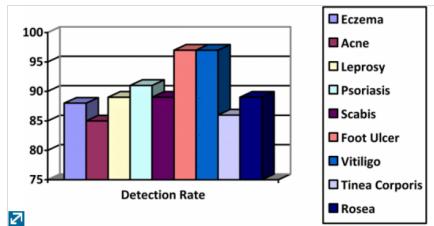


Figure 2.12: Feed forward back propagation artificial neural network diagram

2.6 Computer-Aided Model for skin Diagnosis using Deep learning

For this paper, researchers conduct experiments on three datasets. From these datasets, researchers introduced an enhanced automated computer-aided model which integrates an enhanced segmentation phase for locating the infected lesion of the skin. In this a CNN is designed as a feature extractor [15]. The implementation of the model in this paper started with pre-processing that is used to remove background noises from the skin image as shown in Figure 2.13. Then, the model passes to image processing. In

this paper, Region of Interest (ROI) Segmentation [16] is part of the image processing. Image segmentation can be classifying in three categories: Edge-based, Region-Based and Pixel-based direct classification methods. In this paper, pixel-based is chosen for the system. Pixel-based has 3 major steps that are converting colour space, feature extraction and clustering. Furthermore, feature extraction is another very important phase of image processing which requires vast domain knowledge to help in classification phase. The CNNs used in the proposed model by using a pre-trained model and adapt it to the system.



Figure 2.13: Noises removed from the image.

Moreover, the researchers also had a classifier model that was designed based on multiclass linear support Vector Machine (SVM) that is trained with CNN. In this paper, after the feature extraction, images have been classified using linear support vector machine. SVM is a supervised learning technique that seeks an optimal hyper-plane to separate two classes of samples [1]. Also, SVM have some remarkable characteristics such as, its ability to learn independently of the dimensional of the feature. This is chosen due to its robustness, simplicity and does not tend to over fit training data.

2.6.1 Results

The experiments were carried out to validate the efficiency of the model. Therefore, to measure the performance of the classification different sets of experiments were conducted. The performance of the classification is evaluated in terms of sensitivity, specificity, and accuracy from the confusion matrix of classification. As mentioned, the model was tested on three different datasets. In Figure 2.14 we can see that the model is able to identify all patients with non-melanoma disease. Additionally, is proficient to diagnose 87.5% patients with melanoma. The accuracy of the correctly diagnosis patients is 93.75%. In Figure 2.15, the system can identify 94.12% patients with non-melanoma and can diagnose 94.12% of patients with melanoma in the accuracy classification. In Figure 2.16 we can see that the model can produce high accuracy when there are multiple skin diseases proposed. The highest accuracy is seen in Melanoma disease. Furthermore, these were also compared to other proposed model to several

state-of-the-art skin diagnosis such as Robert Amelard were the accuracy was 83.59% for melanoma diagnosis and 94% this model.

Skin disease	Accuracy	Specificity	Sensitivity
Melanoma	93.75%	100%	87.5%
Non Melanoma	93.75%	87.5%	100%

Figure 2.14: Melanoma and non-melanoma skin lesions from Dermis Dataset

Skin disease	Accuracy	Specificity	Sensitivity
Melanoma	94.12%	94.12%	94.12%
Non Melanoma	94.12%	94.12%	94.12%

Figure 2.15: Melanoma and non-melanoma skin lesions from DemQuest Dataset

Skin disease	Accuracy	Specificity	Sensitivity
Melanoma	98.04%	100%	88.89%
Basal Cell Carcinoma	96.15%	96%	88.89%
Eczema	94.12%	100%	77.78%
Impetigo	91.42%	88.46%	100%

Figure 2.16: : Melanoma Basic cell carcinoma, eczema, and impetigo from Dermnet dataset

Overall, using CNN as representative and discriminate feature extractor allowed the model to represent its diagnosis in the effective solution for automated recognition of skin diseases.

Similarly, to the paper above, another comparative study of deep learning on melanoma detection has been conducted. In this paper, they have used five deep learning architectures that are Xception architecture which stands for extreme inception it has depth-wise separable convolutions [17]. AlexNet is the first breakthrough in the architecture of CNN applied to large datasets [18]. VGGNet it achieves one of the top performances in the ImageNet large scale visual recognition [19]. Finally, they make the use of evaluation metrics to compare the Deep convolutional neural network (DCNN) for skin lesion predictions. Furthermore, the experiment for this paper was evaluating the performance of these different architectures. The dataset used is ISIC as well as ImageNet for pre-trained weights as this helps them to get a robust performance. In Figure 2.17, we can observe the results that were obtained from DCNN. The asterisks are the results from the model that was trained using data augmentation and pre-processing steps. The results show that ResNet50* architecture achieves the best accuracy rate in most of the validation processes, and it has the highest average F-score of 92.74%, and accuracy of 92.08%.

Methods	Precision	Recall	F-Score	ACC
AlexNet	0.7717	0.7873	0.7726	0.7853
AlexNet*	0.8421	0.8125	0.8231	0.8045
ResNet50	0.8652	0.8663	0.8537	0.8637
ResNet50*	0.9373	0.9253	0.9274	0.9208
VGGNet16	0.8442	0.8447	0.8433	0.8436
VGGNet16*	0.907	0.9032	0.9061	0.8836
VGGNet19	0.8468	0.8457	0.8436	0.8461
VGGNet19*	0.8855	0.8882	0.8838	0.8870
Xception	0.4472	0.6633	0.5346	0.6629
Xception*	0.9019	0.9057	0.9041	0.9030

Figure 2.17: Results of the different architectures.

2.7 Analysis Related to Datasets

Considering in reviewing the related work, we can see a wide array of methods have been applied to produce several State-of-the-art. Results on medical science such as herpes, eczema, melanoma and other classification of skin diseases. In the first two papers from section 2.5.1 we can observe the classification/detection of skin diseases using image processing and CNN. Image processing is a crucial part of the process in building a system. There are various types of processing this: first one by removing unwanted element and clearing up the image by removing noises and then training the CNN. Whereas on the second paper a dual stage approach was used, there they first made use of computer vision for image processing and features extraction by using MaxEnt and ANN. Then, in the second stage applying the machine learning techniques and testing it with ANN, KNN and Decision trees. Both approaches give a result with a good accuracy. Furthermore, the two papers on section 2.5.2, in the first one, the research was based on melanoma skin cancer. The process starts with image processing, ROI segmentation which is also an important part of image processing. Additionally, they make use of SVM model that is trained with CNN. Finally, the experiment of model is running, where the model was tested with three different datasets and the accuracy for all of them was above 90%. Moreover, in the second paper of this section a comparative study of deep learning on melanoma is done. Here, five architectures are being compared which are AlexNet, Xception, VGGNet, ResNet50 and DCNN. From here the best accuracy was given by ResNet50 that was trained using data augmentation and pre-processing steps. Lastly, from the five papers, we have assessed that the progress has allowed more advanced methods to help in the detection and combining different techniques can increase the detection of skin diseases.

2.8 Summary and Moving Forward

From the review, we can observe sophisticated methods being applied such as Deep convolutional neural network to medical images that has given good results on diseases such as melanoma skin cancer. Other boundaries have been made in the field of deep learning and specially in CNN architectures. Moreover, having a big concern in the NHS with a shortage of dermatologists as well as radiologists' solutions need be found to solve the issue of increased backlogs. Finally, concluding on how the rest of the paper will be implemented and the proposed problem will be approached. The objective of this project is to employ the CNN architectures to build an accurate image classification pipeline on forecasting the presence of skin diseases in images. To train the model, a public dataset will be used which has also been used in previous papers. If the results of this initial pipeline show high performance around a couple of evaluated metrics,

then the model can be used to classify more than one disease included in the dataset.

Chapter 3

Requirement Analysis

In this project there is a number of requirements to match. Therefore, this chapter will be discussing the important requirements of the project as well as the methodology and design behind it. This chapter is divided into two sections where one discuss about functional requirements and the other about non functional requirements. These requirements were produces to meet the aims and objectives as described in the previous chapter.

3.1 Functional Requirements

3.1.1 Datasets

The dataset selected for this project is Dermnet, this dataset contains image data for 23 categories of skin diseases, where the total number of images are around 19,500, out of which 15,500 are split in training set and the rest in the test set. Furthermore, these images will be pre-processed as this will allow in the improvement of the image, in the pixel values and in other features so that the model can benefit from this improved data. Other than that, the reason of this chosen dataset is that is a very common among the dermatology section, also, as we can see it as been immensely used in the literature, therefore it will allow to build a strong model and the comparison of diseases would be straightforwardly.

3.1.2 Method

In this project, Convolutional Neural Network will be used to build a model using image classification. It can also be observed that CNN is the most common model one used in previous papers as shown in Chapter 2.

3.1.3 Methods comparison

As demonstrated in other literature reviews, comparison of different algorithms could do to compare the prediction of those applied.

3.1.4 Model Improvement & Testing

Model's improvement should be done by trying on different formats in models parameters and then test the result of the new improvements.

3.2 Non-Functional Requirements

3.2.1 Technologies

For this project, the method's that will be used can be realised using Python, Tensorflow, Kheras and sci-kit learn as would help in a quicker development of CNN model.

This will be developed on google collab since it has

3.2.2 Legal, Social & Ethical Issues

The dermatology data used in this project is all from publicly available datasets.

3.3 The Application Front End

3.3.1 Front End

The idea of this front end is, if the model gives good accuracy and the performance is good, these result should be used for a mobile application or perhaps for a web application. This application should detect the disease displayed in the image.

Furthermore, as per user input, the user using this application should be able to click a picture and upload the image to the application.

3.3.2 Technologies - Programming Languages and Libraries

The programming languages used for a mobile application must be Java and Android Studio. On the other hand, for web application the languages used should be HTML, CSS, JavaScript.

Chapter 4

Design & Experiment Setup

This chapter explains the process of the experimental setup before doing the experiments. In this chapter, we discuss the hardware and software setup. Besides, we give an overview of the dataset, explaining the details of the dataset and the preprocessing that was done before starting the main experiments. Then a brief explanation of augmentation is given. Lastly, we explain the evaluation metrics and how we will read our results from the experiments that will be successful or failed.

4.1 Hardware and Software Setup

The computer used for this paper is an HP Pavilion having a processor of Intel(R) Core(TM) i5-7200U CPU @ 2.50GZ 2.71GHZ and a RAM of 8.00GB. Although to run the experiments, Goggle Collaborative was used.

Furthermore, the software used for this paper includes Python 3. The experiment uses Keras, a neural network library written in Python that runs on top of Tensorflow as its back-end. Moreover, Keras is a deep learning API, and it is developed to implement deep learning models in a fast and easy manner for research and development. Keras supports an API that is Tensorflow Distribution API, which allows easily to run models on large GPU clusters.

4.2 Dataset

The dataset used for this paper for training and testing the algorithms is the Dermnet dataset. This dataset has been obtained from a public repository called the dermnet website that is the largest dermatology online source for online medical education.

Furthermore, the dataset contains two folders: the training and testing folders; inside, we can find 23 sub-folders consisting of 23 different skin diseases. The training folder includes some 15557 images and the test folder 4002 images of skin diseases. In Figure 4.1 and 4.2 we can find the pictures distributed between the 23 classes and all different the conditions found in the dataset. It is important to notice this distribution as this will be defining the performance of the model.

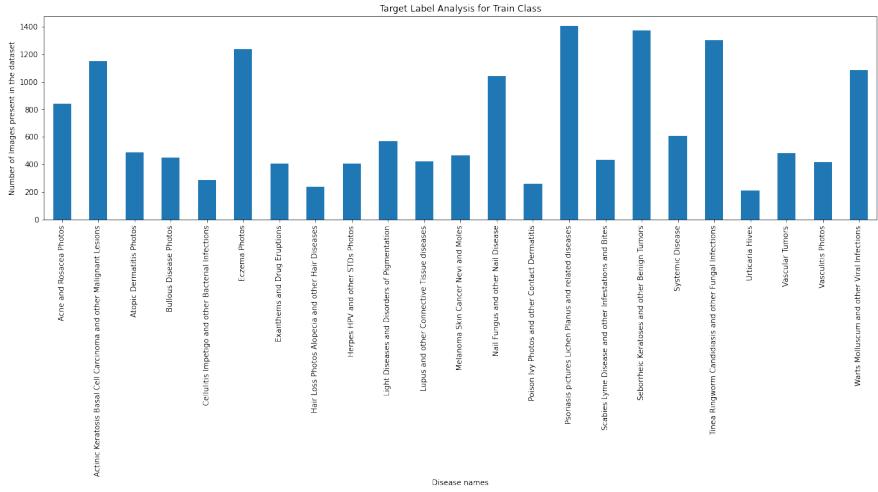


Figure 4.1: The classes and images in train set.

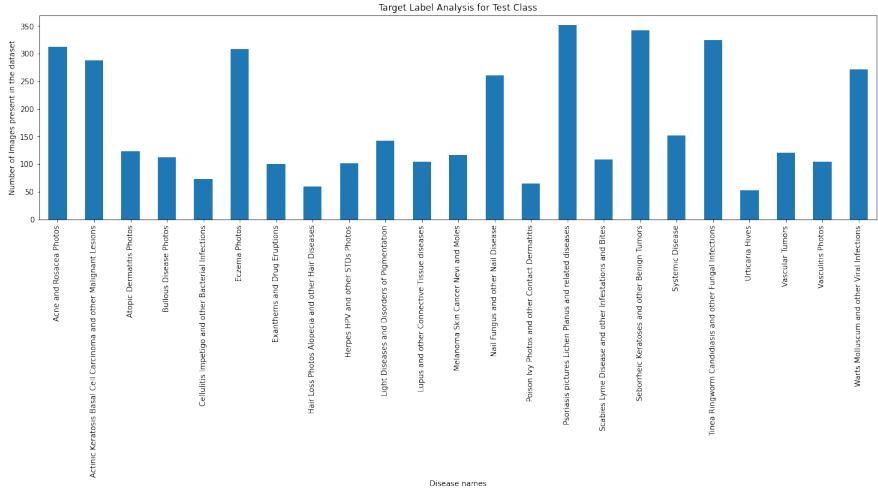


Figure 4.2: The classes and images in train set.

4.2.1 Data Pre-Processing

While exploring the dataset for the preprocessing part, we realised that the images shown in the dataset were all different sizes. This could be due to the diverse type of diseases based on a distinct part of the body. Therefore, before applying any CNN

methods to the learning algorithms, we needed to make sure that all the images were of the same size. The images were resized to 256 x 256 from the original size, where some photos scaled up and others scaled-down.

Furthermore, while continuing the exploration of this dataset, we found that the dataset contained the same images of one disease in various folders, as shown in figure 4.3 the same image has been duplicated in two different classes. We removed the duplicate image and saved it in one folder since that would help us build our model with better efficiency.

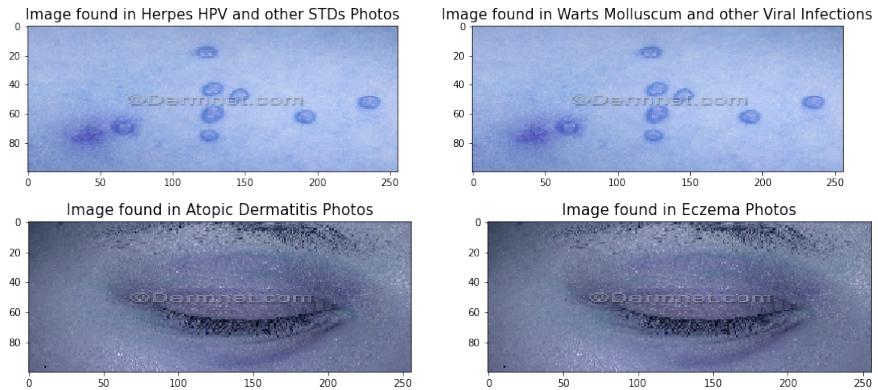


Figure 4.3: The classes and images in train set.

Moreover, another related obstacle the dataset had, was unrelated images or graphical images of skin diseases that would confuse the learning algorithms; a snippet of these are displayed in figure 4.4.

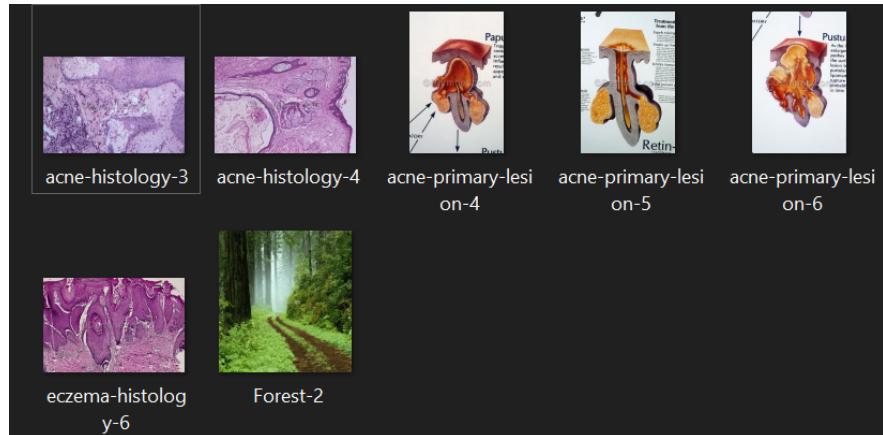


Figure 4.4: Unrelated images

The link for the data pre-processing can be found here: [Data Cleaning.ipynb](#)

4.2.2 Augmentation

Data Augmentation is an important part of the process, this is a technique that creates new training data from the already existing one, therefore, the images will belong to the same class as the original image but as a transformed version. These transformed versions include zooming, rotation, horizontal flip, brightness, etc. these augmentation type were used for this project as well. We can see in figure 4.4 the different type of augmented images.

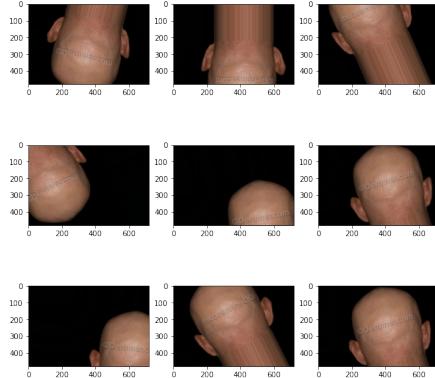


Figure 4.5: Augmented Image

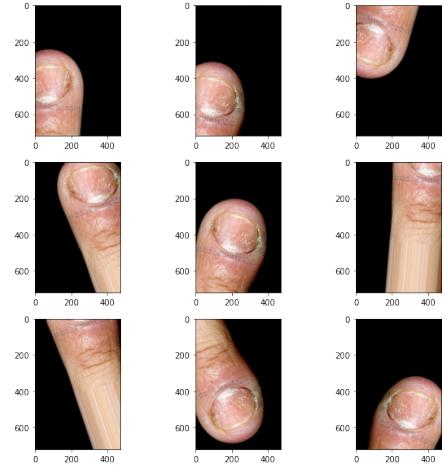


Figure 4.6: Augmented Image

Furthermore, data augmentation is not only used for transforming the image with different type of techniques but it also helps in over-fitting by providing different forms of views and angles to a single image, therefore, the model is able to generalize correctly.

4.3 Imbalanced dataset and Over the accuracy

As shown in figure 4.1 and 4.2 we can observe that the dataset for the 23 classes is highly imbalanced; the distribution of this ranges between 200 to 1400 images. Therefore, going with standard accuracy is not enough; we needed to go beyond accuracy. The aim was to classify all the images accurately and have a good model performance.

Furthermore, as per our aim to detect different types of skin diseases, we considered micro F1 score (confusion matrix) for the evaluation of the results. Micro F1 score asses the multi-class classification problems. In figure 4.7 we can observe the values that this study will examine precision, recall and accuracy.

Precision is defined as the quality of the model showing the returned results of a precise class. Furthermore, when we are having a multi-class classification problem,

$$\text{Micro F1-score} = 2 * \frac{\text{Micro-precision} * \text{Micro-recall}}{\text{Micro-precision} + \text{Micro-recall}}$$

Figure 4.7: Micro F1 score formula

then we add the True positive(TP) and False Positive(FP) with the number of classes we have.

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

Recall also referred as sensitivity, is the fraction of retrieved instances among all relevant instances. Similarly, like precision, for multi-class binary problem, we add the TP and FN depending to the number of classes we have.

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

Specificity shows the degree to which the model identifies the skin diseases as per disease. Also, if multi-class classification problem were to calculated then, we would add the TN and FP according to the number of classes displayed.

$$Specificity = \frac{TN}{TN + FP}$$

Accuracy shows the total percentage of correctness in the classification out of test set.

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

All of these equations is deduced to this confusion matrix shown in the figure 4.8 this is a example that will be used to view or analyse our results. The confusion matrix will be build upon the number of classes presented. Furthermore, True positive (TP) is the sample that are correctly identified by the model with any sort of disease. False Negative (FN) is the sample that are incorrectly identified by the model with any sort of disease. True Negative (TN) sample that are correctly identified with normal samples by the model. False Positive (FP) sample that are incorrectly identified with normal samples by the model.

Furthermore, in multi-class classification, we check the log loss. Log loss, also called as Cross-entropy loss for multi-class problems, measures a classification model's performance whose output is a probability value between 0 and 1.

		Predicted class	
		<i>P</i>	<i>N</i>
<i>P</i>	<i>P</i>	True Positives (TP)	False Negatives (FN)
	<i>N</i>	False Positives (FP)	True Negatives (TN)

Figure 4.8: Confusion Metrics

Chapter 5

Experiments and Results

In this chapter, we explain all the experiments that were carried out for this project in detail, this includes the successful and the failed experiments. Each experiment has an explanation and then their results.

5.1 Experiments 1 & 2

The model architecture for experiment one and two was carried out with a customised model that was a combination of convolutional neural network. The structure for this customised cNN model which consisted of two convolutional layers each having 32 filters and a 3x3 kernel size. The activation that was provided to these layers were "ReLU" activation. After applying all the convolutional layers a max pooling layer was added with a pool size of 2x2. Max pooling reduces the dimensionality, hence, the necessary features are extracted with max operation. Then the weights are flattened, to provide as an input the fully connected layer having a neuron size as same as the number of input images with "ReLU" activation. Finally the model is connected to the output layer with SoftMax activation. In Figure 5.1 we can observe the model architecture used.

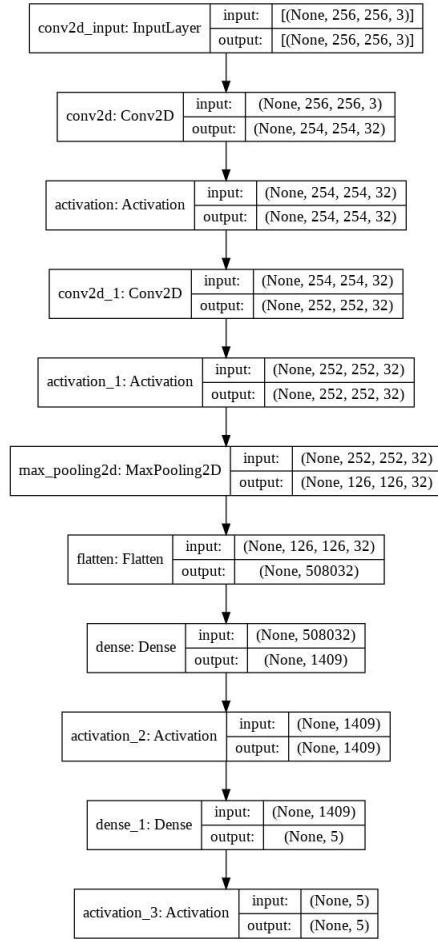


Figure 5.1: Model Architecture for Experiment 1 and 2

Image Classes The Classes used for this experiment were the following:

1. Hair Loss Photos Alopecia and other Hair Diseases - Class 0
2. Herpes HPV and other STDs Photos - Class 1
3. Melanoma Skin Cancer Nevi and Moles - Class 2
4. Nail Fungus and other Nail Disease - Class 3
5. Urticaria Hives - Class 4

Hyper-Parameters The Hyper Parameters used for this experiment are listed below:

- Imbalanced train and test dataset
- Batch size - 32
- Channels - 3 (RGB)

- Height of image - 256
- Width of image - 256
- Epochs - 6/30
- Adam optimiser with learning rate - 0.001

Results

In this experiments we tested them with and without augmentation, although model architecture was the same for both of them. Experiment 1 was performed without augmentation.

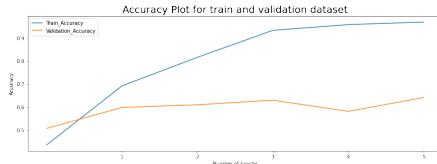


Figure 5.2: Accuracy Plot Exp 1

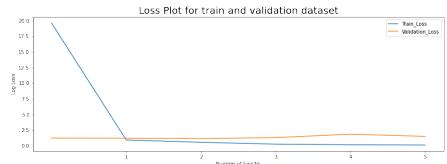


Figure 5.3: Loss Plot Exp 1

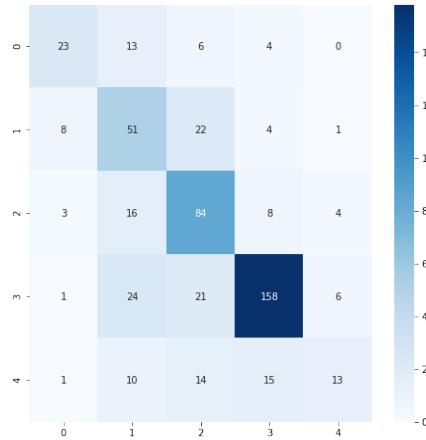


Figure 5.4: Confusion Matrix Exp 1

In figure 5.3 we can observe the loss plot and in figure 5.2 we can view the accuracy plot between train and validation data. The model accuracy for this experiment shows that the training and validation lines are unstable during training process and they are also highly over-fitting. However, the loss lines are stable and not decreasing. The training accuracy score is 96.56% where as validation accuracy is 64.20%. Furthermore, in Figure 5.4 we can view a confusion matrix which shows that the model is biased towards the classes that have higher number of images such as class two and three, which they are able to classify them easily, whereas class 5 is not performing very well and it is unable

to achieve at least a 50% of f1 score. The possible reason of such over-fitting could be due to the classes being highly imbalanced and complex in nature. Augmentation is also not provided to prevent over-fitting.

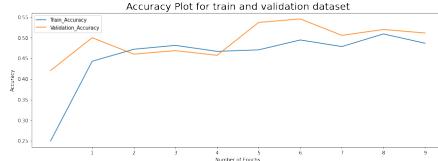


Figure 5.5: Accuracy Plot Exp 2

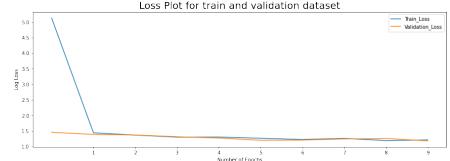


Figure 5.6: Accuracy Plot Exp 2

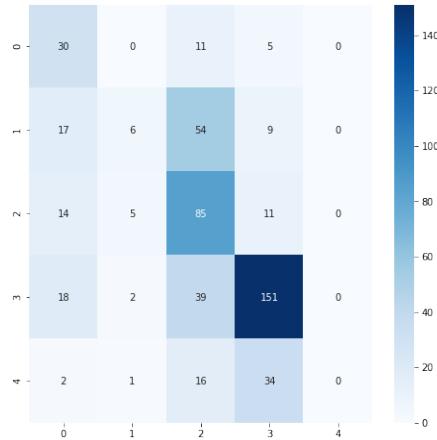


Figure 5.7: Confusion Matrix Exp 2

On the other hand, experiment 2 was performed with augmentation. We can observe the loss plot and we can view the accuracy plot in figure 5.6 for training and validation dataset. The model accuracy for this experiment shows that the training and validation lines are stable within a range of percentage during training process. In this experiment we can also see that the loss lines are more stable than in the previous experiment, although accuracy is much lesser, therefore, the model is under-fitting. Moreover, in figure 5.7 the confusion matrix shows that the model is biased towards class three whereas in class four 0 images were detected and the other classes are less than a 50% except class two that is coming to a 53%. The possible reason of such accuracy could be adding some more layers to the model architecture.

Here we can find the link for Experiment 1 & 2:

[Experiment 1.ipynb](#); [Experiment 2.ipynb](#)

5.2 Experiment 3 & 4

In this experiment the aim was to improve the model architecture, therefore, this model consisted of two convolutional layers each having 32 filters and a 3x3 kernel size. Then a Max pooling layer was added with a pool size of 2x2. Then again a single convolutional layer was added with a max pooling layer of 2x2, which was repeated twice. Also, we used "tanh" activation between the convolutional layers to make the weight conversion slower and to check how the model would perform in the case of activation being "tahn". Then the weights were flattened, to provide as an input the fully connected layer having a neuron size as same as the number of input images with "ReLU" activation. Finally the model is connected to the output layer with SoftMax activation.

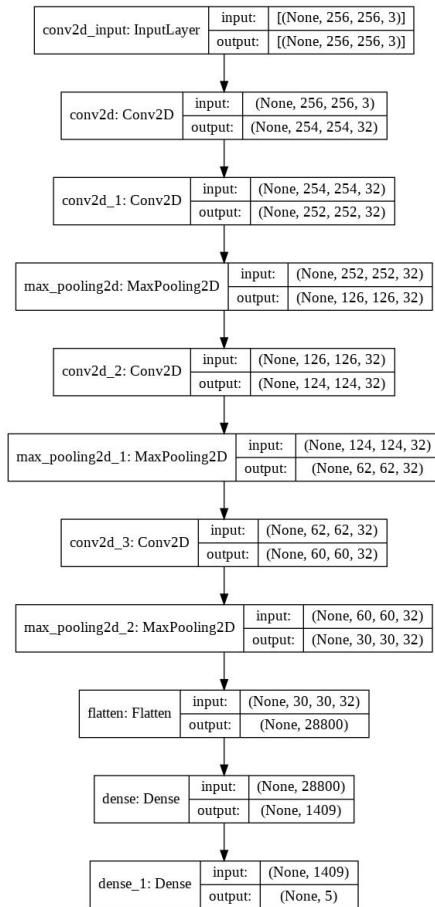


Figure 5.8: Model Architecture

Image Classes The image classes used for this experiment where the same as in experiment one and two as this was to see if we found any improvements.

Hyper-Parameters The Hyper Parameters used for this experiment are listed below:

- Imbalanced train and test dataset
- Batch size - 32
- Channels - 3 (RGB)
- Height of image - 256
- Width of image - 256
- Epochs - 23
- Adam optimiser with learning rate - 0.001

Results In experiment three which was ran without augmentation, we can view the figure 5.9 being the accuracy plot and figure 5.10 being the loss plot. In these plots we can observe that the lines are going up to a good accuracy although they are again over-fitting, therefore we went for experiment four to see if it would yield a better accuracy with the new model and with augmented images.

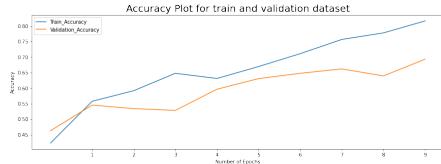


Figure 5.9: Accuracy Plot Exp 3

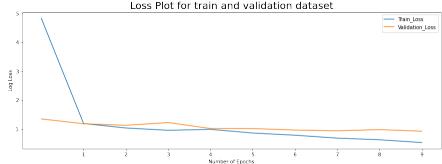


Figure 5.10: Loss Plot Exp 3

Furthermore, in Figures 5.11 and 5.12 we can observe the accuracy and loss plots. Here we can see the accuracy lines are slightly unstable, whereas the loss lines are more stable. We can observe that the accuracy hasn't yield a high number. In the confusion matrix figure 5.13 we can see that class three being Nail disease class yield up to a 72% of f1-score, although the rest of the classes were below or up to a 50% of f1-score. Therefore, get a better performance we decided to use transfer learning to see if this will help us improve our model.

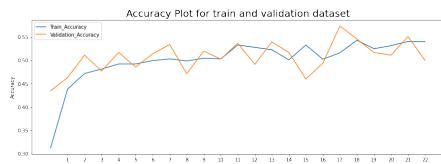


Figure 5.11: Accuracy Plot Exp 4

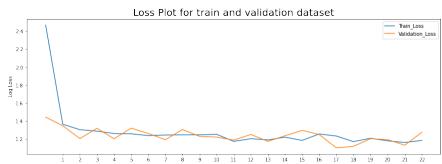


Figure 5.12: Loss Plot Exp 4

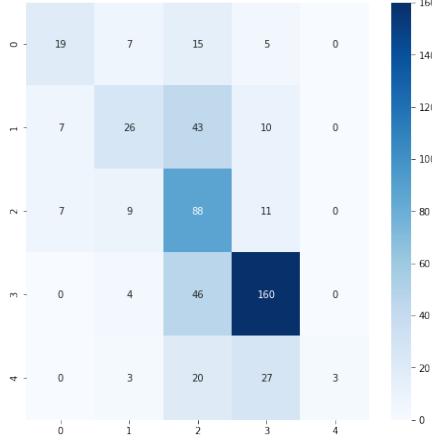


Figure 5.13: Confusion Matrix Exp 4

Here we can find a link to experiments 3 and 4:

[Experiment 3.ipynb](#); [Experiment 4.ipynb](#)

5.3 Transfer Learning

One of the powerful ideas in deep learning is to take the knowledge of a pre-configured model that has been optimised for feature extraction and then apply that models knowledge to another task, this technique is called transfer learning. This learning has shown that it helps to generalize better as well as decreases the training time of a neural network. This has also helped in challenges such as having a limitation of dataset size and training models from scratch [20]. Furthermore, there are various pre-trained models such as VGG16 [19], and ResNet50. In this experiment we have used both pre-trained models to see their performance. In figure 5.14 we can observe the model architecture of 16 layers, and in 5.15 we can observe a 48 convolutional layer with one Max pooling layer and one average Pool layer.

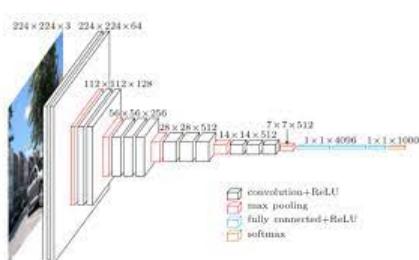


Figure 5.14: VGG16 Architecture

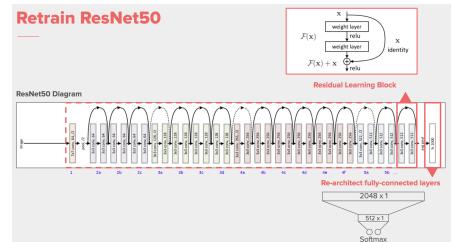


Figure 5.15: ResNet50 Architecture

5.3.1 Experiment 7 & 8

For this experiment, we used pre-trained VGG16 model (imagenet weights) and then we linked a customised fully connected (FC) layer. Moreover, we made all the convolutional layers that were trainable as false and the weight training only happened on the FC layer, which was based of 2 dense layers having a neuron size as same as the number of input images with "ReLU" activation. Finally the model is connected to the output layer with SoftMax activation.

Image Classes The classes used for this experiment are the same five classes used for the previous experiments.

Hyper-Parameters The Hyper Parameters used for this experiment are listed below:

- Imbalanced train and test dataset
- Batch size - 32
- Channels - 3 (RGB)
- Height of image - 256
- Width of image - 256
- Epochs - 30
- Adam optimiser with learning rate - 0.0001

Results In experiment 7 we trained the model without augmenting the input images, this was done to check the performance of the model, again the accuracy was very high but the model was over-fitting, therefore we went for experiment 8. We can view the accuracy plot for experiment 7 in the figure [5.16](#)

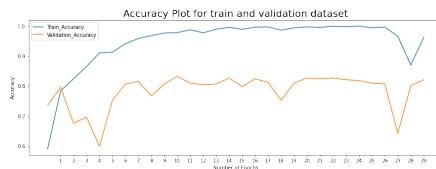


Figure 5.16: Accuracy plot Exp 7

Furthermore, experiment 8 was performed with augmentation. We can view in figure [5.17](#) the accuracy plot and in figure [5.18](#) the loss plot. This is one of the best model we have come across, as we can view in the accuracy plot that the accuracy has improved to a 68% as compared to the other experiments. The loss lines in this model are very stable, therefore, it is allowing the model to be more precise. Furthermore, in figure

[5.19](#) we can find the confusion matrix, which shows that all the classes are generalizing and the f1-score for all is a minimum of 50%. Among all the classes, class three (Nail disease) has the highest precision score that is a 92%, other than that, class one (Herpes and other STDs) has a 69% of precision score, but class four (Urticaria Hives) has still got a lower precision score compared to others which is a 42%, although, it is a very good improvement compared to the other experiments. The reason because the class four was not generalizing correctly could be due to the complexity being high in this specific group.

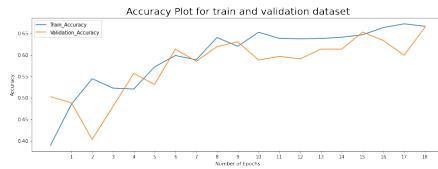


Figure 5.17: Accuracy Plot Exp 8

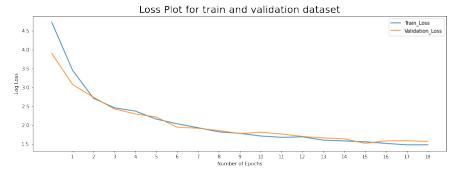


Figure 5.18: Loss Plot Exp 8



Figure 5.19: Confusion Matrix Exp 8

Here we can find experiments 7 and 8 [Experiment 7.ipynb](#); [Experiment 8.ipynb](#)

5.4 Failed Experiments

A number of experiments were tried out before coming to choose 5 classes to test. First experiment ran was the 23 classes using customised CNN model as well as VGG16 architecture (trained from scratch), which did not yield a good accuracy, as we can see in Figure 5.20

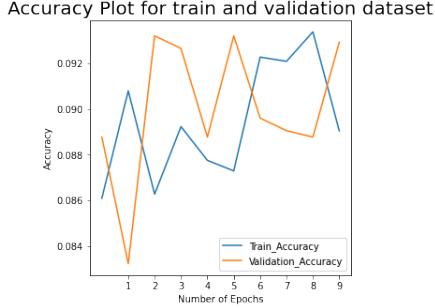


Figure 5.20: Accuracy plot of 23 Classes

Then we decided to go for a 10 classes classification making the use of customised CNN model. We ran this experiment with the input images as well as augmentation. The accuracy was not yielding over a 50%, therefore the model was under-fitting. We can see the accuracy plot in figure 5.21 and figure 5.22

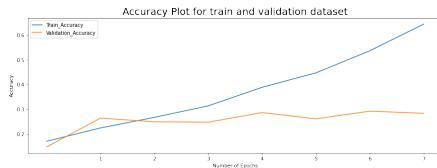


Figure 5.21: Accuracy Plot Exp 5

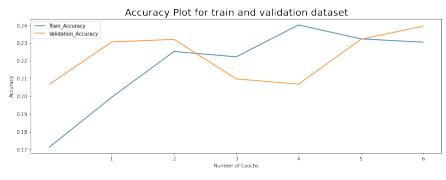


Figure 5.22: Accuracy Plot Exp 6

Moreover, along with VGG16 pre-trained architecture model, we also went with for ResNet50 pre-trained architecture (transfer learning) with the five classes used in the previous experiments. We did this to examine the performance whether it would perform better than experiment 7, but ResNet50 failed us to give a good performance. We can observe the accuracy plot for this experiment in figure 5.23

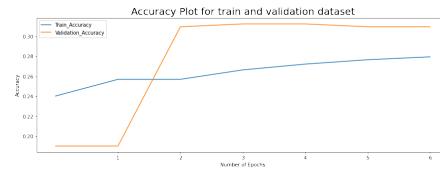


Figure 5.23: Accuracy Plot Exp 10

5.5 Summary of Results

The results for this experiment have been summarised in the table 5.1. As we can observe the final benchmark for our 5 classes is the experiment 8 using VGG16 architecture, although, the observations for all the improving experiments are listed in the

table, this includes the specification of an important metrics score which is F1-score and the accuracy of each one of them.

Experiment	Model	Accuracy	F1 Score
Experiment 1 - no augmentation	2Conv + FC	65%	Class 1- 56% Class 2- 51% Class 3- 64% Class 4- 79% Class 5- 34%
Experiment 2 - with augmentation	2Conv + FC	53%	Class 1- 56% Class 2- 51% Class 3- 64% Class 4- 79% Class 5- 34%
Experiment 3 - no augmentation	4Conv + FC	70%	Class 1- 65% Class 2- 54% Class 3- 68% Class 4- 85% Class 5- 51%
Experiment 4 - with augmentation	4Conv + FC	54&	Class 1- 51% Class 2- 25% Class 3- 50% Class 4- 72% Class 5- 0.07%
Experiment 7 - no augmentation	VGG16	83%	Class 1- 79% Class 2- 74% Class 3- 77% Class 4- 95% Class 5- 74%
Experiment 8 - with Augmentation	VGG16	69%	Class 1- 55% Class 2- 53% Class 3- 59% Class 4- 89% Class 5- 51%

Table 5.1: Summary of Results

Chapter 6

Evaluation & Conclusion

In this chapter we will asses the success of this project regarding the requirements that were described before the implementation process. Then we will provide an overall conclusion of this project displaying the things we learn.

6.1 Overview

The research done in Chapter 2 led us to the requirement Analysis done in Chapter 3. This analysis was using MoSCoW framework which allows prioritizing the tasks. Overall, this analysis was divided into "must" and "should". The requirements for "must" were all carried out. There were CNN models, performing an improvement for these models using fine-tuning of hyper-parameters were a must for the project as stated in Chapter 3. Furthermore, the technologies used were Keras and Tensorflow as they help develop a robust CNN model. Finally, the Legal, Social and Ethical issues related to the dataset were meeting the standards as the dataset was publicly available.

Furthermore, not all the requirements from "should" were achieved. One of the requirement was to have a model to detect all type of diseases, although, most of the time was put in detecting certain types of diseases to have a good model performance, therefore, it was not achieved. Another requirement was that we could implement a website for the model to detect these different type of diseases. This would only be available once the model has been examined and achieved as process of detection. These requirements have not affected the end result of the paper, although having them implemented would have shown an improvement to this project. Also, these were not implemented due to the complexity of the dataset or due to time.

6.2 Conclusion

The aim of this project was to apply deep learning methods to different type of skin diseases images. While the exploration of the dataset, it was apparent that the dataset was highly complex due to every disease having all different part of body, therefore, it was hard for the model to generalize. The dataset also had anomalies, as mentioned in the dataset section. Also, having imbalanced data can cause the model to be biased towards one class than other, as it has happened in some of the previous experiments.

Furthermore, after cleaning the data in chapter 4, we ran various experiments as shown in chapter 5, we observed that the transfer learning approach using VGG16 outperformed compared to the other models. The CNN model was used to see the performance. The VGG16 model was used to extract features by setting as false the convolutional layer as we added our own fully connected layers. In conclusion, the transfer learning model was our benchmark as it outperformed in all metrics compared to the customised model for this specific project.

6.3 Future Work

There is still a lot to improve in this paper, as the aim is to add more diseases for detection with deep learning models. We can experiment with more CNN layers and more pre-trained architectures such as EfficiNet and Inception model. We could also make use of different techniques to balance the classes.

References

- [1] Dietrich R, Opper M, Sompolinsky H. Statistical mechanics of support vector networks. *Physical review letters*. 1999;82(14):2975.
- [2] Bradley WG. History of medical imaging. *Proceedings of the American Philosophical Society*. 2008;152(3):349–361.
- [3] Rimmer A. Radiologist shortage leaves patient care at risk, warns royal college. *BMJ: British Medical Journal (Online)*. 2017;359.
- [4] Eedy D. The crisis in dermatology. *BMJ*. 2015;350. Available from: <https://www.bmjjournals.org/content/350/bmj.h2765>.
- [5] Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW); 2018. p. 117–122.
- [6] Amini A. Introduction to deep learning. MIT. 2019;6:S191.
- [7] Abdel-Hamid O, Deng L, Yu D. Exploring convolutional neural network structures and optimization techniques for speech recognition. In: *Interspeech*. vol. 11; 2013. p. 73–5.
- [8] Wang T, Wu DJ, Coates A, Ng AY. End-to-end text recognition with convolutional neural networks. In: *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*. IEEE; 2012. p. 3304–3308.
- [9] Guo T, Dong J, Li H, Gao Y. Simple convolutional neural network on image classification. In: *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*(. IEEE; 2017. p. 721–724.
- [10] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee; 2009. p. 248–255.
- [11] Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*. 2007;31(4-5):198–211.
- [12] Sah AK, Bhusal S, Amatya S, Mainali M, Shakya S. Dermatological Diseases Classification using Image Processing and Deep Neural Network. In: *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. IEEE; 2019. p. 381–386.
- [13] Kumar VB, Kumar SS, Saboo V. Dermatological disease detection using image processing and machine learning. In: *2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR)*. IEEE; 2016. p. 1–6.
- [14] Yasir R, Rahman MA, Ahmed N. Dermatological disease detection using image processing and artificial neural network. In: *8th International Conference on Electrical and Computer Engineering*; 2014. p. 687–690.

- [15] Shoieb DA, Youssef SM, Aly WM. Computer-aided model for skin diagnosis using deep learning. *Journal of Image and Graphics.* 2016;4(2):122–129.
- [16] <https://towardsdatascience.com/understanding-region-of-interest-part-1-roi-pooling-e4f5dd65bb44>.
- [17] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1251–1258.
- [18] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Neural Information Processing Systems. vol. 25; p. 3065–3075. 2012.
- [19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014.
- [20] Yu Y, MJWXGHZZ Lin H. Deep Transfer Learning for Modality Classification of Medical Images. arXiv preprint arXiv:14091556. 2017.

Appendix A

Project Poster

Image Classification of Skin Diseases Using Deep Learning.

Monica Dayalani Amarnani & Dr Eyad Elyan

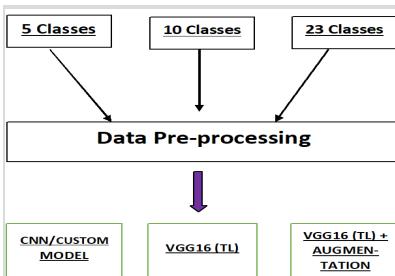
Introduction

Skin diseases are generally among the most common diseases. In 2001, in sub-Saharan Africa, mortality rates were around 20000, and skin diseases were among the high reasons for deaths, reports WHO (World health organization). Moreover, there is a shortage of elementary skills in managing skin diseases, which tends to be a problem. Also, several studies show that assessing success in the management of skin diseases in primary care that treatment failure is above 80%. Therefore, with so many resources available, a tool that diagnoses different skin diseases would be the perfect fit for these problems.

Project Aim

The Project Aim is to implement a skin disease classification pipeline that will be using different techniques of deep learning. Furthermore, this project aims to apply other techniques such as data augmentation, transfer learning, and the use of different models such as VGG16 and ResNet50. These will help to deal with the highly complex and imbalanced dataset.

Methods



For this project, different Convolutional Neural Network(CNN) methods are used, which are present in figure 1. CNN is one of the best deep learning techniques to do feature extraction. Furthermore, the dataset used for this project is highly imbalanced and complex. Therefore, this became challenging since all the images needed to be one size to classify them into the CNN model.

Figures and Results

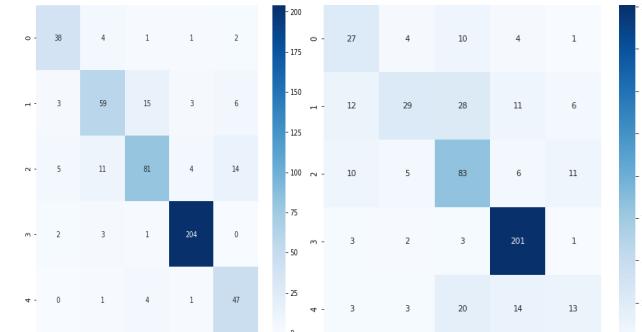


Figure 1: Confusion metric without Aug (left) & with Aug (right)

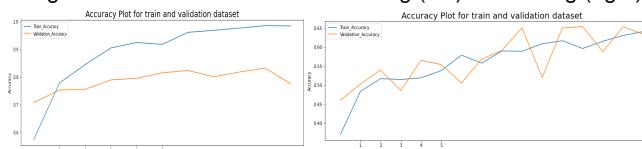


Figure 2: Accuracy plot without Aug (left) & with Aug (right)

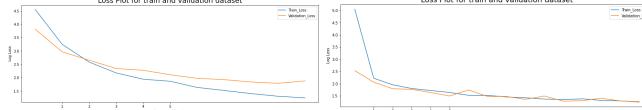


Figure 3: Loss plot without Aug (left) & with Aug (right)

After running several experiments and doing hyper-parameters optimization, we ran an initial model, and it did not show good results. Therefore, we chose to go for pre-trained (transfer learning) architecture such as VGG16 and ResNet50. Furthermore, after running the pre-trained model with augmentation, we obtained an accuracy of **67%** on the test set. Although, we have to say that the model was **low in precision** due to the highly complex and imbalanced dataset. Moreover, in figure 1, the confusion matrix shows that the model is slightly biased towards classes two and four out of those five classes. In figure 2 and 3, we can find the accuracy and loss plots of the models trained with and without augmentation. We can see that the accuracy plot without augmentation is overfitting.

Model	F1 Score	Accuracy	Overfitting
Custom Model / CNN	0.54	0.57	No
VGG16 (Transfer Learning) Without Augmentation	0.84	0.84	Yes
VGG16 (Transfer Learning) With Augmentation	0.69	0.67	No

Table 1: Summary of Models Performance

We can observe in table 1 the accuracy, F1-score, and whether it is overfitting, a performance for each of the algorithms is evaluated. Furthermore, three different algorithms are compared. In this case, transfer learning is used and shows an optimal result compared to other algorithms. We can view that the CNN model built of convolutional layers is showing very low accuracy. However, using the model VGG16 shows a better result than the custom model; VGG16 with augmentation comes up to 67%, which we could say is high for this complex dataset. Although using the model without augmentation is given a higher accuracy, but it is overfitting.

Conclusion

In conclusion, various Convolutional Neural Network models are trained and tested on different skin diseases as they appear in the dataset. We can observe that using methods such as transfer learning through the VGG16 model performed better than a custom model built with convolutional layers. Moreover, other techniques are used, such as data augmentation, which helped with the imbalanced classes in the dataset. Overall, there is a brief overview of the deep learning techniques applied to medical images to detect a different kind of skin diseases. In future work, we hope to build a more generalized classification pipeline on more than five diseases at once that can be held higher in accuracy too.

Acknowledgments

I would like to thank my supervisor Dr Eyad Elyan for all the support and help he has provided throughout this project. I would also like to thank my family and friends for all the support during this year.

References

Hay R, Bendeck SE, Chen S, et al. Skin Diseases. In: Jamison DT, Breman JG, Measham AR, et al., editors. Disease Control Priorities in Developing Countries. 2nd edition. Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2006. Chapter 37. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK11733/> Co-published by Oxford University Press, New York.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, pp.1097-1105.

Appendix B

Project Proposal & Ethics Form

Detailed Project Proposal

First Name:	Monica
Last Name:	Dayalani Amarnani
Student Number:	1611872
Supervisor:	Dr. Eyad Elan

Defining your Project

1.1 Project title

Help: a brief statement about what you are actually going to do.

Using Deep Learning to read medical images. Can deep learning diagnose skin diseases?

1.2 Background

Help: Provide the background to your project. This section should highlight the main topics in the area you are going to research. Essentially what is the project about, what has been done before and why is this project important? ~500 words

Medical images Classification is a very important research topic for health and for machine vision. It has attracted extensive research efforts in the last decade. Dermatology is a sector of medicine that uses medical imaging to diagnose patients. The types of medical imaging range from X-rays, CT (computed tomography) scan, MRI (magnetic resonance imagining), ultrasound, nuclear medicine imaging. Furthermore, some solution has been found to classify skin diseases using machine learning algorithms which is SVM (Support Vector Machine) such as skin cancer Melanoma[1].

Convolutional Neural Networks has advanced masses where lot of research has been published in image recognition for medical scans in the past years. This thesis portrays the applying of Convolutional Neural Network algorithm to diagnose skin disease in medical imaging. This projects aim is whether a medical image can predict a skin disease. Skin disease is one of the most common illnesses in human daily life. It pervades all cultures, happens in any respect ages, and influences among 30 % and 70 % of individuals. The work done for the thesis will be using a publicly available dataset for skin diseases which will include images of normal patients and patients with some skin problem. The images in this dataset are taken from the public portal which is the largest dermatology source.

Because of a shortage of dermatologists, most cases are seen instead by general practitioners with lower diagnostic accuracy. A new deep learning system could diagnose around 16.000 de-identified skin conditions from tele dermatology. Although 80% of the cases were seen in primary care. Thus, as mentioned previously due to the shortage of trained dermatologists the NHS has just seen its waiting list growing and patients waiting for more than 6 weeks to have medical imaging perform on them.

The importance of this project is to demonstrate that with deep learning techniques, medical images can be successfully diagnosed at an acceptable rate.

1.3 Motivation

Help: To whom is this project important? A project must address a question/problem that generates a small piece of new knowledge/solution. This new knowledge/solution must be important to a named group or to a specific client (such as a company, an academic audience, policy makers, people with disabilities) to make it worthwhile carrying out. This is the **motivation** for your project. In this section you should address who will benefit from your findings and how they will benefit. ~300 words

Example 1: If you intend to demonstrate that a mobile application that automates class registers at RGU will be more efficient than paper-based registers - the group who would be interested in knowing/applying these findings would be both academic and administrative staff at RGU and they would benefit by time saved and a reduction in their administrative workload.

Example 2: You are demonstrating that a particular 3D model design increases realism in 3D environments. The group that would be interested would be games designers or developers of 3D virtual environment applications. They would benefit from producing more realistic environments that could increase sales of their products.

Example 3: You have designed a new network topology for IrishOil plc's new Aberdeen headquarters. The interested group would clearly be IrishOil. They would benefit from easier maintenance and improved security of their computer network.

It is often asked if dermatologist will be replaced by artificial intelligence? Dermatologist Harald Kittler mentioned that "Our little experiment shows AI could widen our point of view and help us make new connections" This means this project can help them with identifying the problem and seeing in a different way.

The motivation is to help in the diagnoses of patients. In circumstances where the dermatologist is not present, and the practitioners need a help or where there is a disagreement on whether the patient should be diagnosed with certain disease. In these cases, deep learning models can help to make a final decision.

Furthermore, this model can be helpful for more than one sector of medicine since medical imaging is used for other areas as well.

1.4 Aim & Objectives

Help: Outline what are the main things your project is going to do and what steps or milestones will be used to achieve this aim. The Aim is unlikely to change throughout your project; however, the objectives are likely to adapt to your ongoing research and development. In particular it is highly likely that you may wish to split objectives into sub-objectives as work progresses. A good clear set of objectives give you something to evaluate your final project against.

Example : For the timetable app outlined above

Aim: To create a functioning attendance application that efficiently automates the taking of class registers.

Objective 1: study existing register system in place at RGU and identify weaknesses

Objective 2: research existing automation technology's and identify and evaluate those that may be appropriate to taking in class registers

Objective 3: Implement chosen technologies to create prototype application

Objective 4: Conduct user trials to evaluate capabilities of prototype application

Objective 5: Create a refined application incorporating feedback from user trials

Aim: the aim of the thesis is to predict whether a patient has skin disease based on medical images using deep learning models.

Objective 1: Literature review on medical imaging, deep learning, convolutional neural network and methodologies.

Objective 2: Implement technologies to build machine learning workflow.

Objective 3: Test and evaluate the model performance

Objective 4: Validate the algorithm using the existing data.

Objective 5: Find improvements for the existing technologies.

1.5 Key Techniques

Help: Perform some initial research into the area and outline what techniques you may research in further detail here. The techniques you cover here should include references to the papers where you have sourced

the information. The techniques mentioned here are very likely to become the section headers in your literature review.

The key techniques is using deep learning and convolutional neural networks which falls in to deep neural networks and mainly in to computer vision.

1.6 Legal, Social, Ethical, Professional and Security issues

Help: Here you should discuss any legal, social, profession and security issues that you believe may occur during the course of your project. It is not acceptable to write none in this box, all projects, regardless of focus will have to address issues in one, or more, of these categories. This is an extremely important part of your honours project to which there is no correct answer, this section must be fully discussed with your Honours Supervisor.

Example 1 : In the class register example above – there would be a Legal and Security issue with the gathering and storage of student data. There may be a social constraint as you may be relying on a user to have access to a specific technology. There will need to be consideration of user accessibility.

Example 2 : A 3D model design may have ethical considerations in its evaluation. What if your model made users feel nauseous. Social constrains may again be access to technology or accessibility issues.

Example 3 : You network design need to adhere to specific company policies. You would need to consider the possibility that your design could be wrong, compromising the company's security.

This project has the possibility to be used in the real-world. There could be a possibility where the program fails or provides a wrong diagnosis which will be given to a patient. To prevent this sort of issues it is important for an expert to give the final decision. The responsibility would lie to healthcare professional to use that software correctly.

1.7 Project Plan

Help: This is the project plan as to how you will go about achieving the objectives of the project.

Example: In the class register example above the research plan may involve:
Collecting and analysing paper-based registers in a given class on five occasions.
Identifying the error rate average on these occasions
Researching existing automation techniques
Designing and implementing a mobile application that automatically records attendance in class.
Deploying the application in the class on five occasions.
Identifying the error rate average of the mobile application on these occasions.
Comparison of data and summary of findings.

1. Information of important research
 - 1.1 Literature Review
 - 1.2 Evaluation of technologies that can be used
2. Review previous work
 - 2.1 Examples of previous work using CNN
 - 2.2 Information on these examples
3. Design implementation
 - 3.1 Collection of requirements

- 3.2 Non-functional and functional requirements
- 4. Implementation
 - 4.1 Build model against the data provided
 - 4.2 Build interface for medical images to be uploaded
- 5. Testing
 - 5.1 Test model against test set and report
- 6. Findings
 - 6.1 Explain the findings when reporting test results with visualisations
- 7. Evaluate project
 - 7.1 Changes and achievements made from initial project plan
- 8. Conclusion

1.8 Ethics Form

You must include in your signed ethics form in this submission or you will not be able to continue the project.

References

- [1] N. V. Kumar, P. V. Kumar, K. Pramodh and Y. Karuna, "Classification of Skin diseases using Image processing and SVM," 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 2019, pp. 1-5, doi: 10.1109/ViTECoN.2019.8899449.
- [2] Roderick J. Hay, Nicole E. Johns, Hywel C. Williams, Ian W. Bolliger, Robert P. Dellavalle, David J. Margolis, Robin Marks, Luigi Naldi, Martin A. Weinstock, Sarah K. Wulf, Catherine Michaud, Christopher J.L. Murray, Mohsen Naghavi, The Global Burden of Skin Disease in 2010: An Analysis of the Prevalence and Impact of Skin Conditions, *Journal of Investigative Dermatology*, Volume 134, Issue 6, 2014, Pages 1527-1534, ISSN 0022-202X, <https://doi.org/10.1038/jid.2013.446>.



STUDENT PROJECT ETHICAL REVIEW (SPER) FORM

The aim of the University's *Research Ethics Policy* is to establish and promote good ethical practice in the conduct of academic research. The questionnaire is intended to enable researchers to undertake an initial self-assessment of ethical issues in their research. Ethical conduct is not primarily a matter of following fixed rules; it depends on researchers developing a considered, flexible and thoughtful practice.

The questionnaire aims to engage researchers discursively with the ethical dimensions of their work and potential ethical issues, and the main focus of any subsequent review is not to 'approve' or 'disapprove' of a project but to make sure that this process has taken place.

The *Research Ethics Policy* is available at
www.intranet.rgu.ac.uk/credo/staff/page.cfm?pge=7060

Student Name	Monica Dayalani Amarnani
Supervisor	Dr. Eyad Elyan
Project Title	Image Classification of Skin Diseases using Deep Learning
Course of Study	BSc Computer Science
School/Department	School of Computing

Part 1 : Descriptive Questions			
1	Does the research involve, or does information in the research relate to: (a) individual human subjects (b) groups (e.g. families, communities, crowds) (c) organisations (d) animals? Please provide further details:	Yes	No
		X	X
		X	X
		X	X
2	Will the research deal with information which is private or confidential? Please provide further details: The dataset that the model will be run against is a public dataset of skin diseases. It does not contain patient health information(PHI).	Yes	No
		X	X

Part 2: The Impact of the Research			
3	In the process of doing the research, is there any potential for harm to be done to, or costs to be imposed on	Yes	No
	(a) research participants?		X
	(b) research subjects?		X
	(c) you, as the researcher?		X
	(d) third parties?		X
	Please state what you believe are the implications of the research:		
4	When the research is complete, could negative consequences follow:	Yes	No
	(a) for research subjects		X
	(b) or elsewhere?		X
	Please state what you believe are the consequences of the research:		

Part 3: Ethical Procedures			
5	Does the research require informed consent or approval from:	Yes	No
	(a) research participants?	<input checked="" type="checkbox"/>	X
	(b) research subjects	<input checked="" type="checkbox"/>	X
	(c) external bodies	<input checked="" type="checkbox"/>	X
	If you answered yes to any of the above, please explain your answer:		
6	Are there reasons why research subjects may need safeguards or protection?	Yes	No
	<input checked="" type="checkbox"/>	X	
	If you answered yes to the above, please state the reasons and indicate the measures to be		
7	Has PVG membership status been considered?	<input checked="" type="checkbox"/>	X
	(a) PVG membership is not required.	<input checked="" type="checkbox"/>	X
	(b) PVG membership is required for working with children.	<input checked="" type="checkbox"/>	X
	(c) PVG membership is required for working with protected adults.	<input checked="" type="checkbox"/>	X
	(d) PVG membership is required for working with both children and protected	<input checked="" type="checkbox"/>	X
	If you answered yes to (b), (c) or (d) above, please give details:		
8	Are specified procedures or safeguards required for recording, management, or storage of data?	Yes	No
	<input checked="" type="checkbox"/>	X	
	If you answered yes to the above, please outline the likely undertakings:		

Part 4: The Research Relationship			
9	Does the research require you to give or make undertakings to research participants or subjects about the use of data?	Yes	No
			X
	If you answered yes to the above, please outline the likely undertakings:		
10	Is the research likely to be affected by the relationship with a sponsor, funder or employer?	Yes	No
			X
	If you answered yes to the above, please identify how the research may be affected:		

Part 5: Other Issues			
11	Are there any other ethical issues not covered by this form which you believe you should raise?	Yes	No
			X

Statement by Student			
I believe that the information I have given in this form is correct, and that I have addressed the ethical issues as fully as possible at this stage.			
Signature	M. Dayalani Amarnani	Date	07/10/2020

If any ethical issues arise during the course of the research, students should complete a further Student Project Ethical Review (SPER) form.

The *Research Ethics Policy* is available at
www.intranet.rgu.ac.uk/credo/staff/page.cfm?pg=7060

Part 6: To be completed by the supervisor			
12	Does the research have potentially negative implications for the University?		No
	If you answered yes to the above, please explain your answer:		
13	Are any potential conflicts of interest likely to arise in the course of the research?		No
	If you answered yes to the above, please identify the potential conflicts:		
14	Are you satisfied that the student has engaged adequately with the ethical implications of the work? [In signifying agreement, supervisors are accepting part of the ethical responsibility for the project]	Yes	
	If you answered no to the above, please identify the potential issues:		
15	Appraisal: Please select one of the following		
	The research project should proceed in its present form – no further action is required	<input checked="" type="checkbox"/>	
	The research project requires ethical approval by the School Ethics Review Panel		
	The research project needs to be returned to the student for modification prior to further action		
	The research project requires ethical review by an external body. If this applies please give details		
	Title of External Body providing ethical review		
	Address of External Body		
Anticipated date when External Body may consider project			

Affirmation by Supervisor			
<p>I have read the student's responses and have discussed ethical issues arising with the student. I can confirm that, to the best of my understanding, the information presented by the student is correct and appropriate to allow an informed judgement on whether further ethical approval is required.</p>			
Signature	Eyad Elyan	Date	16.10.2020

Appendix C

Project Log

Weekly Meetings	Meetings and Tasks
30 th September	Discussion of the project and supervisor gave suggestions on how and where to start.
8 th October	Change of project to skin diseases project & understand latex to work on it
15 th October	Going through the Ethic form and Project Proposal before submission
16 th October	Project Proposal and Ethics form submission
22 nd October	Start of Literature Review with outline and machine learning techniques
29 th October	Continuation of Literature Review and other Coursework submissions
5 th November	No meeting as other coursework's were priority
13 th November	Updating supervisor Eyad with the literature review progress, he expected more to be done from what was shown
26 th November	A quick overview about the literature review before submission
27 th November	Submission of Literature Review
3 rd December	Working on Coursework's as deadline was approaching
10 th December	Briefing on Requirement analysis
17 th December	Working on Final Courseworks
24 th December	Christmas holiday
1 st January	Working on Requirement Analysis
8 th January	Looking into dataset and checking requirement analysis
12 th January	Going through requirement analysis and Next steps of the project
18 th January	Submission of Requirement analysis
28 th January	Data preprosessig
4 th February	Blood test appointment, couldn't attend meeting
11 th February	Problems with the dataset as program was crashing everytime it was ran
18 th February	Work in progress and learning more about multi class
25 th February	Supervisor Eyad gave a briefing on how the final project was to be submitted
4 th March	Couldn't attend meeting as I was travelling back home
11 th March	Bried Supervisor Eyad with the progress
18 th March	Working on courseworks
25 th March	More experiments were done and Supervisor Eyad gave a briefing on Demo and Poster
1 st of April	Easter Week
8 th April	Briefing on Poster and Report Structure
15 th April	Working on Coursework and report
22 nd April	Comments on Jupyter notebook and writing report
27 th April	Finish up Report
3 rd April	Project Finished

Appendix D

Source Code

[Click Here](#) to go to the source code.