Multiclass Skin Lesion Classification with Convolutional Neural Networks

MSc Research Project
MSc. in Data Analytics

Yash Balaji Iyengar
Student ID: x18124739

School of Computing
National College of Ireland

Supervisor: Dr. Vladimir Milosavljevic

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Yash Balaji Iyengar
Student ID:	x18124739
Programme:	MSc. in Data Analytics
Year:	2019
Module:	MSc Research Project
Supervisor:	Dr. Vladimir Milosavljevic
Submission Due Date:	12/12/2019
Project Title:	Skin Lesion Classification with Convolutional Neural Networks
Word Count:	7829
Page Count:	23

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

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

Signature:	
Date:	11th December 2019

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

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

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

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

Multiclass Skin Lesion Classification with Convolutional Neural Networks

Yash Balaji Iyengar
x18124739
MSc. in Data Analytics

Abstract

Skin Cancer is one of the most dangerous diseases to affect human beings. The main cause of this disease is skin exposure to the ultraviolet rays of the sun. Due to these multiple types of abnormalities develop on the body known as skin lesions. These skin lesions in the future may become cancerous. Dermatologists have a hard time in diagnosing and detecting cancerous skin lesions due to the similarity between the types of lesion cells and manual inspection of the images. This study proposes a multiclass classifier which detects seven different types of skin lesions using the method of Convolutional Neural Network with transfer learning. A modified VGG19 architecture has been developed by replacing the classification layers with Global Average Pooling layer, Dropout layer and two fully connected layers with a softmax activation function to classify seven types of skin lesions. The proposed model has been tested on the ISIC 2018 challenge image dataset which consists of 10,015 images of seven different types of skin lesions which were pre-processed using Morphological operations and Histogram Equalisation. The proposed methodology was able to achieve a validation accuracy of 78.68 % and a test accuracy of 74.08 % respectively. This study has been conducted with a motive to assist dermatologists in identifying the type of skin lesion.

1 Introduction

Skin cancer is one of the most frequently diagnosed types of cancer. According to the World Health Organisation, one out of three cancer patients is diagnosed with skin cancer. According to Fan et al. (2019) in the US alone, 99,550 cases of skin cancer were diagnosed in the year 2018. The Human Body consists of skin which is the largest sensory organ according to Kalwa et al. (2019). It has two layers, epidermis which protects the body from injuries, allergies and external toxic radiations and dermis which provides energy and nurtures the epidermis layer and makes it stronger. It is when this organ gets exposed to Ultra Violet rays different types of lesions develop on the skin which may lead to skin cancer. There are three major types of Skin Cancer, basal cell carcinoma, squamous cell carcinoma these two are known as non-melanoma type skin cancer and the third and the most dangerous type is melanoma Chatterjee et al. (2019). If skin cancer is diagnosed in the early stages the patient is given a five-year survival rate, this can be increased up to 90% with proper treatment. In order to provide appropriate treatment and medicine, early diagnosis is very much essential Navarro et al. (2019). Dermoscopy is a high-level imaging method which helps to detect skin lesions. According to Ma and Tavares (2016) detection of skin lesion from these images is performed by a dermatologist and the best-achieved accuracy for detection of skin lesions with the naked eye is 60 %. Also, this takes a lot of time and due to low global availability of these equipment and experts the diagnosis gets delayed, if this process is automated the time required for the diagnosis of skin lesions will be reduced which will help in increasing the survival rate of the patients.

In recent years Convolutional Neural Networks have been able to achieve state of the art results in the field of image recognition and computer vision Marchetti et al. (2018). Neural network architecture is inspired from the way in which the human brain works, once an image is passed as input, each layer extracts some features from the image and learn the patterns in the data to accurately classify the output Pathan et al. (2018). Various studies have been conducted for model-based automated classification of skin lesions by authors like Brinker et al. (2019) and Tschandl et al. (2019) and their models have outperformed dermatologists. Some authors have made use of transfer learning to reduce computing costs at the same time achieve respectable results. Transfer learning is a term that means being able to pass on the knowledge that is gained by learning from the data that was used to train the model on other applications but being able to generalize the model by hyperparameter tuning and using it for classification in different domain. Hosney et al. (2019) used AlexNet architecture for ternary skin lesion classification whereas Tschandl et al. (2019) made use of UNet16, LinkNet152 and LinkNet34 for binary lesion classification.

Accurate classification of skin lesion images depends on the image quality, the amount of noise in the image and the methods used to reduce the effect of noise in the pre-processing stages. The diagnosis of skin lesions from dermoscopic images depends on the images Asymmetry, Border Irregularity, colour Variegation and Diameter of the lesion, this is also known as the ABCD rule Riaz et al. (2019). Authors like Fa et al. (2019), Seeja and Suresh (2019) have used various image segmentation and feature extraction methods to improve the results of the automated diagnosis of skin lesions. Authors like Pathan et al. (2018) and Rahman et al. (2017) have developed a novel optimization algorithm which enhances the parameter tuning in a classifier which has provided better results. Even though abundant research has been conducted in this field most of the study focuses on binary classification or ternary classification of skin lesions. Need for a generalised model was felt while reading the works of literature which identifies multiple types of skin lesions. In this research, a modified transfer learning model is proposed to classify seven different types of skin lesions with pre-processing techniques like Morphological operations and Histogram Equalisation performed on the data to achieve improved results.

2 Related Work

The essence of the literature review section is to give a brief summary of various researches previously conducted in the field of skin lesion detection. The related work section is divided into four subsections they are subsection 2.1, section subsection 2.2, subsection 2.3 and subsection 2.4

2.1 Image Segmentation and Feature selection techniques for skin lesion detection

The manner in which a dataset is pre-processed impacts the performance of a model. In a research conducted by Fa et al. (2019), morphological geodesic active contour segmentation (MGAC) is a technique which is used for pre-processing to automatically detect skin lesions from the images. The dataset used here is from PH2 database and has 200 RGB dermoscopic skin lesion samples. The size of the images ranges from 761 x 768 pixels to 553 x 576 pixels. To reduce computing costs the images were resized to 40% of its original size. While converting the images to grayscale the authors found a high amount of information contained in the blue channel as compared to the red and the green channels. A 9 x 9 size Gaussian filter was applied for noise removals like hair and skin lines. Otsu's threshold was then applied to determine contour patterns. Erosion a pixel blurring method was applied to ignore the body injuries and a mask was applied to the image after erosion. The authors evaluated their model based on ground truth, sensitivity, specificity, F1 score, Jaccard Index and Accuracy. The authors achieved a sensitivity score of 91.72 % and specificity of 97.99% which are higher than than the state of art methods. The only limitation of this study was the small size of the dataset. In future work, the authors have suggested to perform these techniques on datasets with a large number of samples.

Image Segmentation is a process where the image is broken into small parts with a certain amount of pixels in order to remove distortions and noise or to improve the pixelated contrast in the image which might help in the feature extraction process. Seeja and Suresh (2019) have used the above mentioned preprocessing technique in skin lesion classification. The author has used Deep learning method for both image segmentation and classification. The U-Net architecture was used for segmentation of the images in the ISIC 2016 challenge dataset. These segmented images were then passed to VGG16 model for binary classification purpose. The authors compared the results of the VGG16 classifier before and after segmentation and observed that the model was able to get an Accuracy, Sensitivity and Specificity of 83.18 %, 95.53% and 92.96% respectively with Segmentation as compared to 76.08%, 93.45% and 92.96% without segmentation.

Noise and distortion in image effects the machine learning model's ability to detect the disease significantly. In order to understand and weigh the importance of this statement Fan et al. (2019) conducted research on how different types of distortions can affect the model. The model used here was InceptionV3. The dataset consisted of 1300 images of different types of skin lesions. The four different input datasets were prepared which consisted of the original clean dataset, gaussian noise, impulse noise and a mixture of both respectively. The authors then trained the models on all four datasets and evlauated the model with metrics like Accuracy, AUC, t-SNE and Saliency maps. It was observed that the test accuracy of the original dataset was the highest that is 96.50 % and the lowest test accuracy was observed for the dataset with a combination of both noises which was 88.40 %.

A similar study conducted by Rahman et al. (2017) focuses on pre-processing techniques like image resizing, image filtering with Median and Gaussian filters to remove noise. For feature extraction, techniques like Colour Enhancement and Histogram Equalisation were also applied. A technique called DullRazor algorithm was used for removing noise caused due to body hair. Image segmentation was performed with the help of Colour Threshoulding which was also performed by Seeja and Suresh (2019) in their research. These pre-processed images were then used to train KNN and SVM classifiers which detect the type of skin lesion. With this proposed methodology they were able to achieve an average test accuracy of 69 % and an F1 score of 0.89 respectively.

Further emphasis was put on image segmentation and removal of noise due to dermoscopic body hair by Pathan et al. (2018) in their study. A novel approach was developed for automatic detection of body hair and separating the lesion based skin from the normal skin using chroma-based geometric deformable algorithmic model. 2-D Gabor filters were used for detecting the hair masks. The use of this was also advocated by Serte and Demirel (2019) in their study. With the proposed approach they were able to achieve an average Sensitivity of 82.4 %, Specificity of 97.2 %, test Accuracy of 94.6 % and overlap scores of 11.52 %.

2.2 Skin Lesion detection using Transfer Learning

Transfer learning is a method used to reduce computing costs and get good results in image processing. This methodology has been applied by Hosney et al. (2019) in the classification of skin lesions. The authors have used AlexNet architecture on three different datasets consisting of RGB skin lesion images of the classes Melanoma, Melanocytic Nevi and Seborrheic Keratosis. Similar to Mahbod et al. (2019) approach the authors have used augmentation techniques for handling data imbalance in order to avoid overfitting. The author has used ReLU activation function in the five convolutional layers of the AlexNet and striped the output layer and replaced it with two fully connected layers with softmax activation for classification. SGD algorithm has been used for weight initialization of the layers. 10 fold cross-validation was performed to avoid overfitting. The authors used accuracy, sensitivity, specificity and precision to evaluate their model similar to Fa et al. (2019). With the proposed methodology they were able to achieve 96.86%, 97.70% and 95.91% accuracy with binary and ternary classification on the three datasets respectively which is better than the state of art models the only limitation of this methodology is not taking into consideration the other classes of skin lesion which were filtered out of the dataset.

In a study conducted by Tschandl et al. (2019), skin lesion segmentation has been performed using three neural network architectures namely UNet16, LinkNet34 and LinkNet152. The authors have experimented with three weight initialisation techniques like default weights which are He Uniform, ImageNet weights and weights of the model

ResNet34 after fine-tuning. The datasets used for this research are ISIC 2017 challenge and ISIC 2018 challenge which is also known as HAM10000 images dataset. The experiment was performed in a Linux environment with high-performance NVIDIA GeForce 1080Ti GPUs. Due to the availability of high computing power images were resized to 512 x 512 pixels. For post-processing of skin lesion segmentation, conditional random fields (CRF) was used. The authors have used Dice score, Jaccard index and Jaccard TS score as a metric for evaluating their models.In this study the size of the dataset used was bigger as compared to Mahbod et al. (2019). According to the authors, ISIC datasets segmentation masks can be divided into three types, Polygons, pixel-based and line tracings, the HAM10000 dataset consists of line tracings. With multiple experiments, the authors were able to get a high Jaccard score when they used Imagenet weights and fine-tuned their LinkNet152 model as compared to using random weight initialization techniques. The authors were unable to classify certain types of tumour-like actinic keratoses because of lack of pigmentations in the images due to which their model could not extract features and generalize well.

A large amount of research is conducted in the field of medical science, for disease diagnosis using images with the help of Convolutional Neural Networks. In one such work conducted by Mahbod et al. (2019), ternary classification has been performed for the diseases Melanocytic Nevi, Seborrheic Keratoses and Melanoma with the help of an ensemble approach. The authors have used a combination of inter and intra architecture of CNN models. The dataset chosen for this study are images from 2017 skin lesion classification challenge. The test data consisted of 600 images. Colour Standardisation has been used for pre-processing the images along with image normalisation using the Gray World Color Constancy algorithm. The images were resized to 224 x 224 pixels. For pre-training and important feature extraction the authors used AlexNet, VGGNet and ResNet architectures which are also used by Tschandl et al. (2019). In all these architectures the authors stripped the output layers and added two additional fully connected layers with 64 filters each and 3x3 kernels. They used different optimizers like SGDM, RMSprop and Adam similar to Seeja and Suresh (2019). To avoid overfitting the authors up-sampled their images using Image Augmentation techniques. Area Under the Curve (AUC) has been used as the metric for the evaluation of the models. They obtained 87.3% AUC for Melanoma cases detection and 95.5% for Seborrheic Keratoses. Even though this approach is well documented, the sample size chosen is quite small and also the authors do not take into consideration the other types of skin Lesions and focus only on three types.

Alimboyong and Hernandez (2019) have made use of a similar approach for plant seedling identification using transfer learning. They build a custom architecture which was similar to the AlexNet architecture and classify 12 different types of plant seedlings with a dataset obtained from Aarhus University signal processing group. The dataset consists of 4234 images of 12 different types of plants. The model consisted of five convolutional layers with ReLU activation function, padded with max-pooling layers and dropout layers to extract best features and avoid overfitting respectively. Finally, two fully connected layers with softmax activation were used to classify 12 different types of plants. SGD was used to as optimizer to reduce the cost function, similar to Hosney et al. (2019). The model was evaluated based on accuracy, sensitivity and specificity. With the proposed architecture, the authors were able to get a remarkable accuracy of 90.15%, sensitivity of 92.36% and specificity of 91.67%. The only limitation of the research was the size of the dataset and the authors suggested data augmentation in future works to reduce the class imbalance.

MobileNet is also a unique transfer learning architecture which has several use cases in the field of computer vision. Sae-lim et al. (2019) has used this architecture for skin lesion classification. The author has used the HAM10000 dataset which was also used by Tschandl et al. (2019). Dataset consists of RGB dermatoscopic images of 7 different types of skin lesions. To handle the data imbalance in different classes, the data of minority class has been upsampled using image augmentation techniques similar to Mahbod et al. (2019) and Alimboyong and Hernandez (2019). The authors have tweaked the original MobileNet model and stripped the last five layers which consisted of 2 Reshape Layers, Dropout Layer, Convolutional Layer and softmax activation function layer and replaced them with a single Dropout layer and a Fully Connected layer which has activation function as softmax for making predictions of the classes. They did this to avoid overfitting and got excellent results. They used accuracy, sensitivity, specificity and F1-score to evaluate their model performance. They were able to achieve 83.23% accuracy, 87% of sensitivity, 85% of specificity and 82% F1 - score with their proposed methodology.

A very recent study which was published a month ago conducted by Zhang et al. (2019) focuses on an improved Whale optimisation Algorithm instead of the Back Propagation algorithm to reduce the error rates and reinitialize the weights in a neural network due to which efficient classifier was built for skin cancer diagnosis. For training, the model with more data two datasets was fused that is the Dermquest and the DermIS to include 22,000 clinical images. The input size of the image considered was 28 x 28 pixels which are less detailed as compared to other studies conducted in the field. The activation function used was ReLU similar to Tschandl et al. (2019). Due to the WOA algorithm global minima was reached the best values of weights and biases in the algorithm were achieved. After evaluation of the model using index parameters like Accuracy, Sensitivity and Specificity it was observed that the model outperformed the current state of art models.

Yan et al. (2019) have combined the two approaches of using a new optimization algorithm and ensemble learning for creating a generalised model that can detect skin lesions as well as other types of disease. Feature extraction was performed using techniques like Grey Level Run Length Matrix, Local Binary Patterns, Histogram of Oriented Gradients and texture feature extraction. Two deep learning models were developed for feature selection using particle swarm optimization, one being for adaptive acceleration coefficients and second for random acceleration coefficients. The stratified sample split for train, validation and test was 60, 20, 20 which is different as compared to other methods observed. With the proposed methodology they were able to achieve 97.79 % test accuracy.

2.3 Ensemble Learning approach for Skin Lesion Classification

Ensemble learning approaches have obtained better results in computer vision-based applications which can be seen in a recent study conducted by Serte and Demirel (2019). The authors used an ensemble of eight CNN models to create a state of art technique that classifies skin lesion types. The dataset used was ISIC 2017 challenge image dataset which consisted of 2000 images of 3 types of skin lesions. The images were pre-processed using a Fourier Transform application called Gabor Wavelet. The images were rotated in seven different directions by providing different degrees of rotation. Each output was provided to a CNN model for training. A basic CNN model was also built which was trained on Images without any Gabor Wavelet preprocessing. The eight CNN models were built using the AlexNet and the ResNet18 architectures. Imagenet weights were used for training the models. The output probabilities of these eight models were fused using the probability sum rule to determine the type of skin lesion in the test image. The metrics Accuracy, Sensitivity, Specificity and AUC were used for evaluating the ensemble model. The proposed methodology was able to achieve 0.96 AUC, and 0.83 Accuracy respectively. The only limitation of this approach was the size of the dataset and in future work, the authors suggested the model should be trained on more samples.

Harangi (2018) approached the skin cancer classification with a similar approach, the author combined the output of four CNN architectures namely the GoogleNet, ResNet, VGGNet and the AlexNet to build a ternary lesion classifier. The convolutional layers of these models used SGD algorithm for optimization. The ensemble approach was used to compensate for the small dataset which consisted of 2000 training images and 600 test images. IT was artificially upsampled using augmentation techniques to 14,300 images. The output of the four different classifiers was combined using the sum of probabilities. Even though the approach was similar the results fell as they were able to achieve an average AUC of 0.891 which was less than Serte and Demirel (2019) methodology.

In previous cases, we saw the use of augmentation as a technique to handle class imbalance and transfer learning as a technique to detect skin cancer. A research conducted in Moscow by Gavrilov et al. (2019) has used similar methodologies along with an ensemble approach to detect and classify malignant skin lesions. The Inceptionv3 architecture was used on the dataset obtained from International Skin Imaging Collaboration. To increase the training samples of the dataset, image augmentation was used but at a very large scale. The original dataset consisted of 10,000 images. The ideology used was similar to Mahbod et al. (2019) and Sae-lim et al. (2019) and the original image size was 300 x 300 pixels. Five insceptionv3 models were trained with hyperparameter tuning in each model was different weight initialization techniques. These CNN models are were then combined into an ensemble model and the final classification was decided by majority voting between the models. The model was evaluated based on the metrics accuracy, sensitivity, specificity and area under the ROC curve. The authors were able to get 91% accuracy on the test set and area under the ROC being 0.96. The sensitivity obtained was 85% and specificity was 92%.

2.4 Humans vs Computer performance comparison for Skin Cancer Lesion Detection

To diagnose the type of skin cancer in a patient completely based on a machine learning models output is still considered risky as it can't always be a hundred per cent correct all the time. So to overcome this concern a recent study conducted by Hekler et al. (2019) diagnosed a patient by a combination of the opinion of dermatologist and an output of a CNN model. In this study, the HAM10000 dataset was chosen which has been used by many authors in their studies like Tschandl et al. (2019) and Sae-lim et al. (2019). There were 5 types of skin lesions in the dataset and the class imbalance was handled by augmenting the images. The final dataset consisted of 12,336 images out of which 300 images were chosen for test purposes that are 60 images from each class were chosen randomly. These images were given to 112 dermatologists in the form of a questionnaire and were asked to classify into one of the 5 types of skin lesion. The remaining 12,036 images were used to train the CNN model which was built using the ResNet50 architecture. The 300 test images were also tested on this model. The results of the dermatologist were subject to biases and therefore statistical outliers were removed. Then the results of the CNN image classifier and the dermatologist questionnaire were fed as an input to XGBoost algorithm and the hyperparameters were tuned using Randomised searchCV. Accuracy, Sensitivity and Specificity were used to determine the performance of the final model. This proposed methodology was able to achieve a sensitivity score of 86 % and a specificity score of 81.5 %. The limitation of this method would include data collection bias and willingness to participate in the survey.

A similar study conducted by Brinker et al. (2019) lead to a conclusion that the Deep Neural Network classifier provided better output as compared to a dermatologist. A ResNet50 model was trained with 4202 images each of class Melanocitic Nevi and Melanoma. MC Nemar's test was used to determine both the results. The model, as well as the dermatologists, were tested on 804 images and the results showed 67.7% and 62.2% sensitivity and specificity for the dermatologists whereas the DNN classifier was able to achieve a significantly better score for both that is 82.3% sensitivity and 77.9% specificity. To build upon the previous study Maron et al. (2019) wanted to compare the results of a multi-class skin lesion classification model with the dermatologists. For this purpose, HAM10000 dataset was chosen and duplicate images were filtered out. The class imbalances were handled by downsampling the majority class and upsampling the minority class. The total dataset consisted of 12,000 images out of which 300 images were chosen for testing with the help of random selector to avoid selection bias. The ResNet50 model was trained to identify five types of skin lesion and was tested on the 300 test images. At the same time, a survey was conducted and 112 dermatologists from 13 German hospitals participated. After comparing the results it was observed that the model outperformed the dermatologists and was able to achieve a specificity of 98.8 % as compared to the 89.2 % achieved by the dermatologists.

Diagnosis of skin cancer in a patient takes effort and skill from a dermatologist and a measurable amount of time is required for proper diagnosis. A study conducted by Kalwa et al. (2019) proposed a smartphone application which detects benign or malignant melanoma within a time frame of one second with respectable accuracy. Preprocessing techniques include image Gaussian filtering, segmentation of the images using real-time curve evolution algorithm which takes Asymmetry, Colour Variegation, Diameter of the lesion and Border Irregularity into consideration.SMOTE was used to handle the class imbalance in the dataset. Then the segmented images are aligned appropriately along the axes. These images are then used to train SVM classifier which classifies as Melanoma or non-Melanoma. They were able to achieve an average Sensitivity, Specificity, Accuracy and AUC of 80 %, 90 %, 88 % and 0.85 respectively.

3 Methodology

Using Deep Learning Algorithms for finding meaningful insights from the data and applying it to solve a business problem has been a common practice. For generalizing this process so that it can be followed across different industries and domains it has been divided into two methodologies namely Knowledge Discovery in Database (KDD) Fayyad et al. (1996) and Cross-Industry Standard Process for Data Mining (CRISP-DM) Wirth and Hipp (1995). In section 2 we saw that Alimboyong and Hernandez (2019) advocated the use of the CRISP-DM as it is independent of the industry as well as the technology used and can be used in any domain.

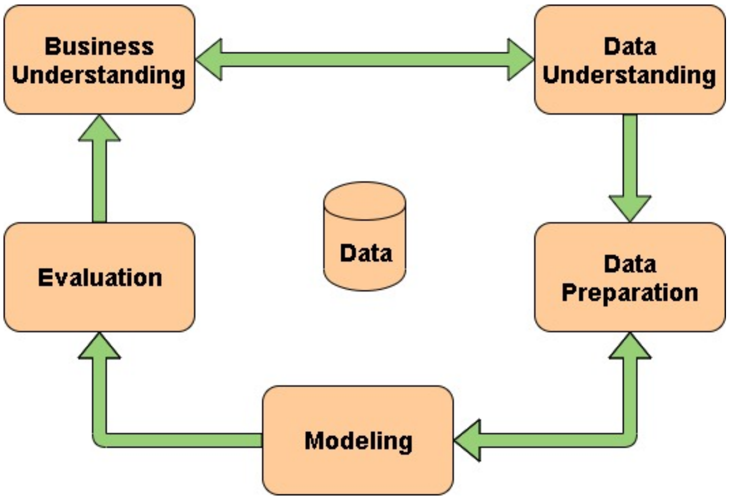

Figure 1: CRISP-DM Methodology

3.1 Data Understanding

The dataset chosen for this research is the ISIC 2018 challenge hosted by the International Skin Imaging Collaboration [1]. Tschandl et al. (2018) and Codella et al. (2018) have made significant contributions for the creation of the dataset. This dataset is also known as the HAM10000 dataset, it consists of total 10,015 images of seven types of skin lesions with dimensions 600 x 450 pixels. The dataset is provided with a metadata CSV file which consists information about each image, the lesion id, age, gender of the patient also how the ground truth is obtained that is either by histopathology or by a follow-up examination or by agreement of experts called consensus or by conforming in-vivo confocal microscopy. This dataset has been used by many authors in their work mentioned in section 2 like Tschandl et al. (2019), Sae-lim et al. (2019), Hekler et al. (2019) and Brinker et al. (2019). This dataset will be used to build a seven class skin lesion classification model.

Feature Name	Details
lesion_id	Unique lesion identification number
image_id	Unique Image identification number
dx	acronym for the skin lesion cell type
dx_type	Type of ground truth Validation
age	Age of the patient
sex	Gender of the patient
localization	Affected area of the body

Table 1: Features in the Metadata file

[1] https://challenge2018.isic-archive.com

3.2 Data Preparation

3.2.1 Exploratory Data Analysis

Exploratory data analysis is performed on the metadata file provided with the images. The file was loaded into the python environment and the sample distribution of the different classes was checked.

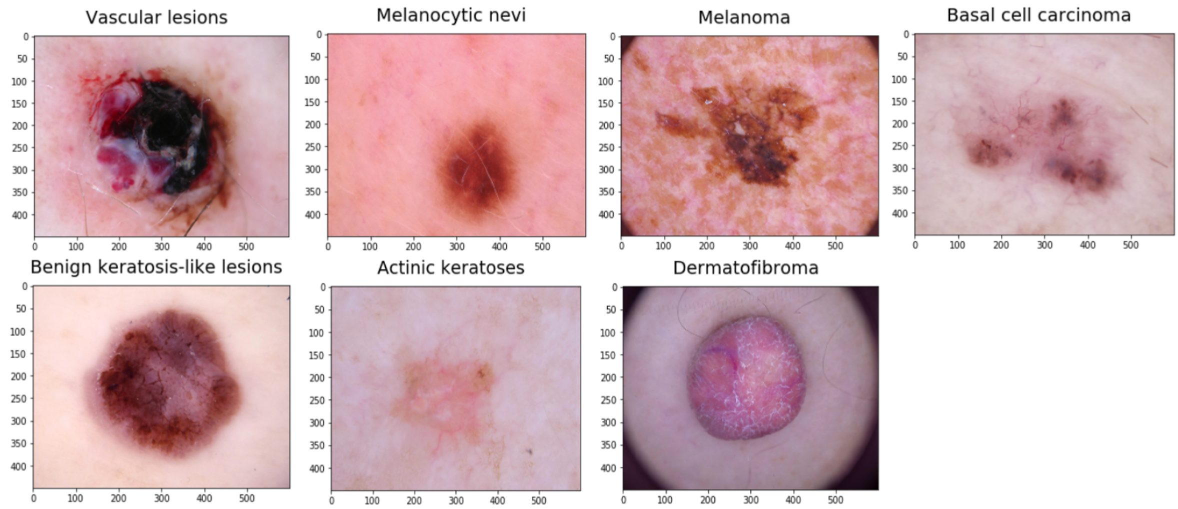

Figure 2: samples of each class

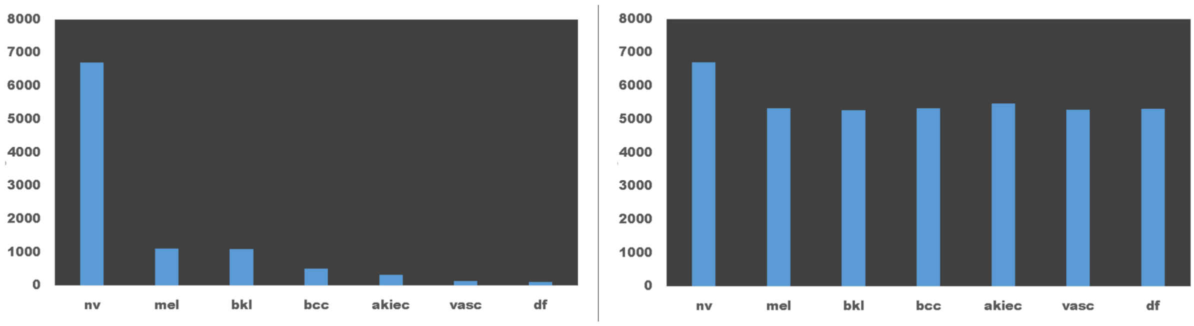

Figure 3: Samplewise class distribution before (left) and after (right) Over sampling

From the bar chart 3 , we can see that the dataset is highly imbalanced and majority samples are from the class Melanocytic nevi. Over sampling of the minority cases has been performed using, Image Augmentation techniques in order to avoid overfitting.

3.2.2 Morphological Operations

Morphology is the study of shapes and sizes of objects or structures Haralick et al. (1987). This is used in image processing to enhance the quality of an image. In morphological operations, a structuring element is used which checks an image pixel by pixel and processes the image according to the neighbouring pixel values. Basic Morphological operations include Erosion, Dilation, opening and closing. In our dataset, the dermoscopic images consist of noise elements like body hair which if not treated will impact the output of our model. Pathan et al. (2018) puts emphasis on treating noise distortion due to body hair in the images as we referred it in the section 2. As Fa et al. (2019) advocated the use of Erosion and Dilation in their data preparation, we will be performing Erosion and Dilation to get rid of the noise and distortions in our data. Erosion is the process of decomposing the image or shrinking the image. In erosion, the structuring element is compared with the pixels in the image and if a section of the image matches the element then it is replaced with the origin of the element. Dilation is the process of expanding the image by adding pixels to the image boundaries. This is done by vector addition or substitution of the structuring element to the image.

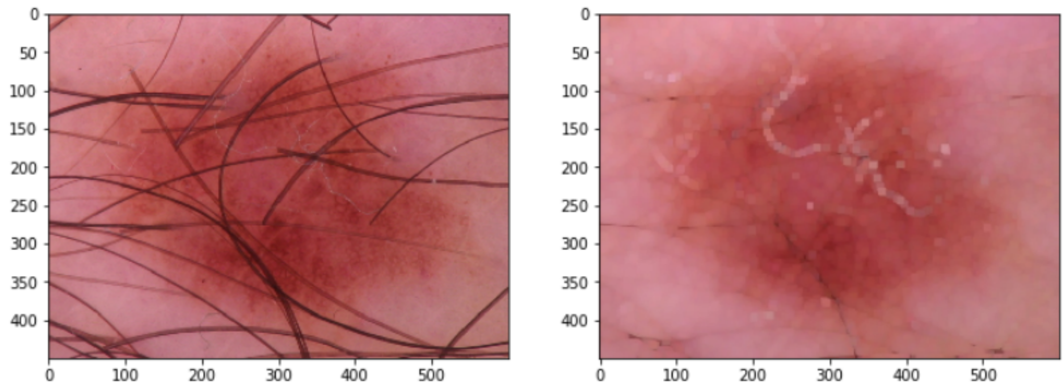

Figure 4: Image before (left) and after (right) Morphological Operations

3.2.3 Histogram Equalisation

A histogram can be explained as a graph of the pixel intensities in an image. Histogram equalisation is a technique used to enhance the quality of an image by improving their contrast Cheng and Shi (2004). In this method, the highest pixel intensities are distributed in the entire image thus making the image-rich in colour. In medical images, this technique helps to improve the contrast of the affected area also called the region of interest. This technique has been used by Rahman et al. (2017) and Yan et al. (2019) in their works and have achieved great results.

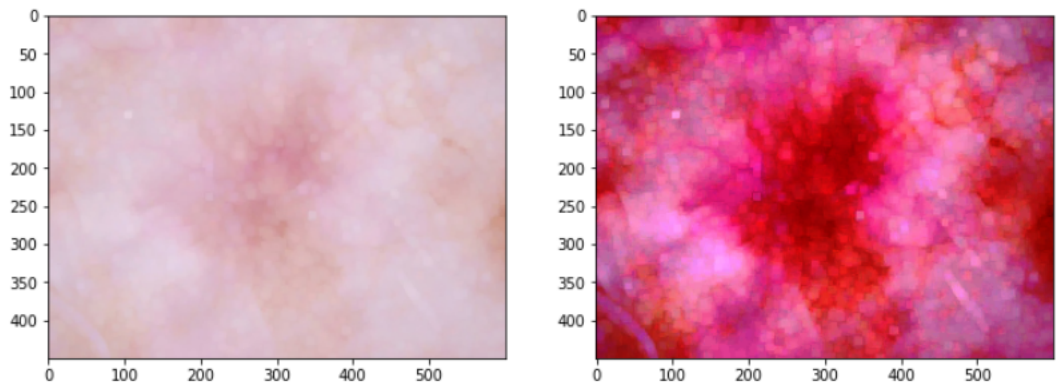

Figure 5: Image before (left) and after (right) Histogram Equalisation

4 Modeling

- **Convolutional Neural Networks:**
  Neural networks were first introduced in the 1950s but weren't used due to the vanishing gradient descent problem. In 1994 Yann LeCun introduced the world to the Convolutional Neural Network Yosinski et al. (2015) which is an efficient form of the original neural network and works well due to the property of backpropagation. Neural network architecture is inspired by the human brain and consists of various layers which extract information from even the most complex form of data like images.

- **Convolution Layer:**
  This layer consists of multiple neurons and for image processing, they consist of a 2x2 matrix. Hence these layers in Keras package are called Conv2D layers. This layer consists of filters specified by the user, this filter is placed on the three-channel image and dot product is found. These extracted features are passed to the next layer.

- **Max Pooling Layer:**
  As discussed in the previous section convolutional layers have a large number of weights which are obtained from the dot product of the input image and filters. A max pooling layer sub-samples the layer and extracts the maximum values by sliding a 2x2 window over each feature map of previous layer. Due to this, the computation costs are reduced and the high intensity pixels of the features are selected.

- **Global Average Pooling Layer:**
  It calculates the average output of the feature maps in the previous layers and passes it on to the next layer. These layers reduce the tendency of our model to overfit.

- **Dropout Layer:**
  Since CNN layers contain a larger number of parameter it becomes a powerful tool for prediction and classification problems. Due to its complex structure, it is prone to overfitting. The dropout layer is used to bypass this problem Srivastava et al. (2014). The dropout layer randomly deactivates some of the neurons in the layers, so that the model does not learn from the training data completely. Due to this, the model is able to generalize.

- **Loss Function:**
  A loss function is calculated by comparing the output of a convolutional neural network and the ground truth. The square of the difference between the predicted value and the actual value is called the loss function. This value is used to optimize the weights of the parameters in a convolutional neural network to improve its output through back propagation. Categorical crossentropy is used in this research as this is a classification problem, since the final dense layer outputs one of the seven skin lesion types.

- **Activation Function:**
  Activation functions decided the activation of a neuron. A neuron calculates the

weighted sum of the inputs and the bias. This is then passed through an activation function which decided to activate or deactivate the neuron. In this study, Rectified Linear Unit also knowns as **ReLU** is used in the convolutional layers and in one dense layer. It sets all the negative weight values to zero and all the positive weight values to one. The softmax function is used in the last fully connected layer, as the output needs to be one of the seven types of skin lesion.

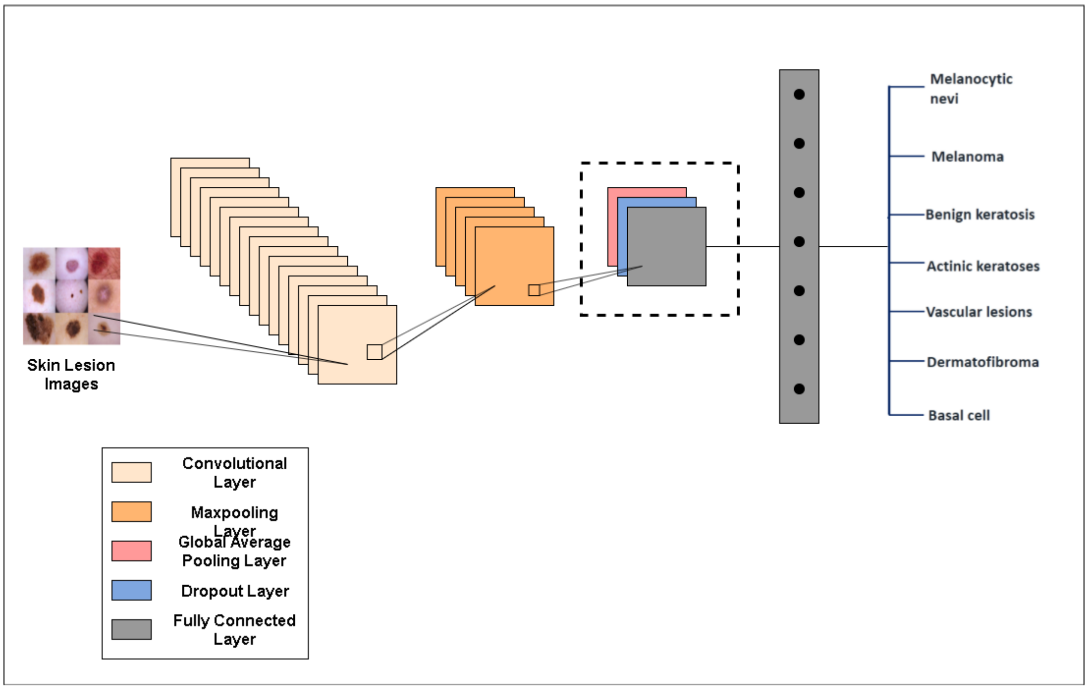

Figure 6: Modified VGG19 Architecture

- **VGG19 Architecture:**
  The VGG architecture was first introduced by Simonyan and Zisserman (2014) which won the ImageNet Large Scale Visual Recognition Challenge. Here VGG stands for the Visual Geometry Group in the Oxford University. It consists of sixteen convolutional layers with increasing depths. But each filter is of the size 3 x 3. It consists of five max-pooling layers. Due to this architecture, the model performs well and is able to generalise well on any data and extract features, so it has been chosen in this research as the model of choice. The original model consists of three fully connected dense layers which have been stripped and modified. In this research, a global average pooling layer, a dropout layer, a dense layer and a final softmax layer have been added to the VGG model, the architecture for which has been described in the diagram 6.

5 Design Specification

The research workflow 7 is inspired from the CRISP-DM methodology and by reviewing various works of literature in the skin lesion classification domain. According to the design first, the dataset is explored with the help of visualizations to get a basic understanding, missing values in the data are then imputed using mean imputations. Then the image

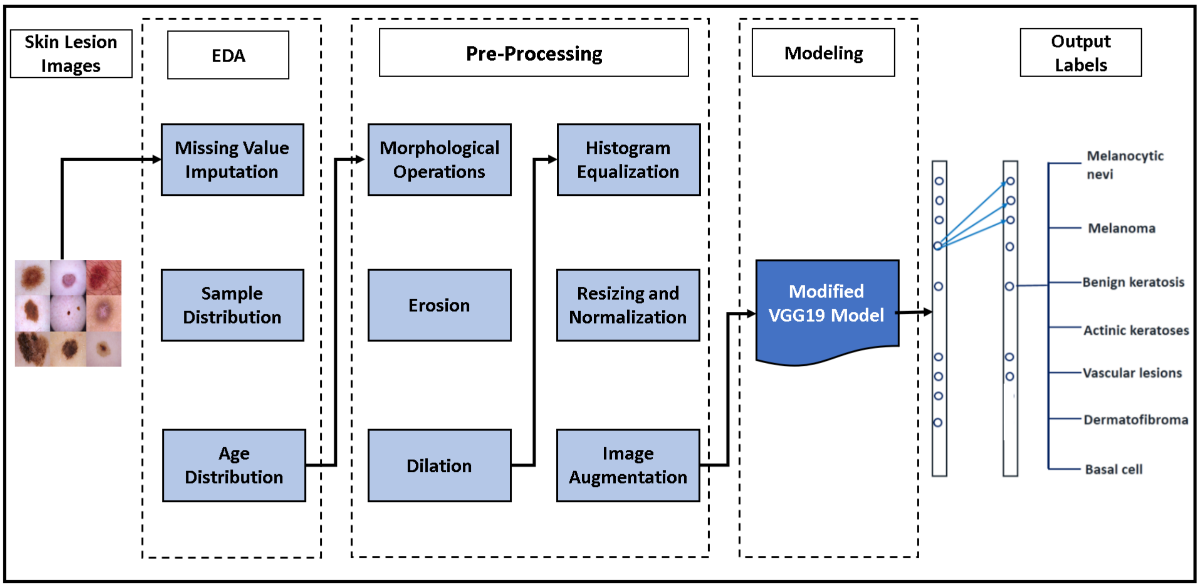

Figure 7: Research Workflow

data is pre-processed with morphological operations like erosion and dilation. Histogram equalisation is performed on the images to improve the contrast and enhance the overall brightness in the region of interest (ROI). The images are then resized and normalised to make them compatible according to the computing specifications. Image augmentation is performed to reduce the class imbalance, here the images are flipped and rotated. Then the appropriate model is chosen to carry out the multilabel classification. Finally, the results are evaluated with validation accuracy and test accuracy used for classification problems.

6 Implementation

- **Data Exploration (EDA):**
  The dataset was decompressed into two folders consisting of 5000 and 5015 images respectively. It also consisted of a metadata file. This CSV was loaded in the python Jupyter notebook environment using the Pandas library and read_csv function. From observing the first few rows of the dataset it was observed that the image id column was the same as the names of the images in the respective folders. In the Jupyter notebook environment a dictionary consisting of the image paths was created using the OS library and functions like os.path.join() and os.path.splitext(). Another dictionary was created with the skin lesion acronyms and their full medical terminologies as key-value pairs. The image paths and the skin lesion names were mapped to the data frame using the map() function. The skin lesion type column was categorically encoded using the categorical() function in the Pandas library. The data was then checked for missing values, the age column had 57 missing values. These values were imputed using the mean imputation method. Then the cell type class distribution was checked by visualizing the skin lesion type variable in the metadata CSV. Patient age, gender type and type of diagnoses was also visualized for observation and gaining a better understanding of the available data.

- **Pre-processing:**

Morphological operations have been performed using the Open CV package. The image is loaded in the python environment, then a structuring element that is a 5 x 5 identity matrix is generated and then the erosion operation is performed on the dataset to shrink the image. The erosion function is iterated only once. Then dilation is performed on the image by using the same structural element on the image and the process has been iterated two times. This image is then saved. The entire erosion and dilation process is automated by looping for all the 10,015 images. Histogram equalisation is performed on these images for enhancing the features in the image. The image is first loaded in the Jupyter notebook and saved like an array in a variable. Then the image is flattened using the flatten() function. This flattened array of the image is plotted to check the pixel distribution. Then a histogram function is defined which plots the histogram of the image. A cumulative sum function is defined which calculates the cumulative sum of the histogram created. The obtained cumulative sum values are then normalised between 0-255 pixels. The flattened image is reshaped. Thus spreading the pixel distribution and enhancing the contrast of the image.

- **Image Normalization, Augmentation and Stratified Sampling:**
  The original dimensions of the images are 450 x 600 x 3, which cannot be handled by the computer given its specifications, so they were resized to 100 x 75 pixels and were mapped to the data frame using the map() function and the lambda function. Before the modelling stage, the samples are split into train, test and validation sets. The ratio used for stratified sampling is 70 % for training the model, 10 % for validation of the model and 20 % for testing the model. So out of the 10015 images, 7210 images, 802 images and 2003 images are randomly split into train, validation and test sets respectively. Images are scaled with the Standard Normalization technique. Image augmentation is performed on the training set. ImageDataGenerator() function is used for performing the augmentation task. The images in the train set are randomly rotated by 10 degrees, zoomed by a ratio of 0.1. They are also shifted in width and height by a ratio of 0.1.

- **Modeling:**
  From the Keras applications the VGG19 model was loaded into the Jupyter notebook environment. Imagenet weights were also loaded. The last three fully connected layers were dropped. A global average pooling layer was added, a dropout layer with a dropout range of 0.5 was added and a final dense layer with softmax activation function was added as final layers of the model.TensorFlow GPU was activated to obtain better computing power. The model was run for 40 epochs with a batch size of 10. The ReLu activation function was used for the first fully connected layer and Softmax function was used for the final Fully connected layer because the output needs to be one of the seven types of skin lesion cells. The model was run for four different cases and the validation and test accuracy was noted. The four cases were as follows

  1. On Raw Data without Pre-Processing
  2. With Morphological Operations
  3. With Histogram Equalisation
  4. With Over Sampling

7 Results

Case No.	Case Study Name	Val Acc (%)	Test Acc (%)
1	Raw Data	74.43	69.97
2	Morphological Operations	**78.68**	**74.08**
3	Histogram Equalization	70.94	70.19
4	Over sampling	63.82	59

Table 2: Results of the case studies

The table above is created to give a brief overview of the experiments performed and the results achieved from them. An in-depth discussion of each case with graphical representations has been provided in the sections below.

7.1 Model Training on Raw Data:

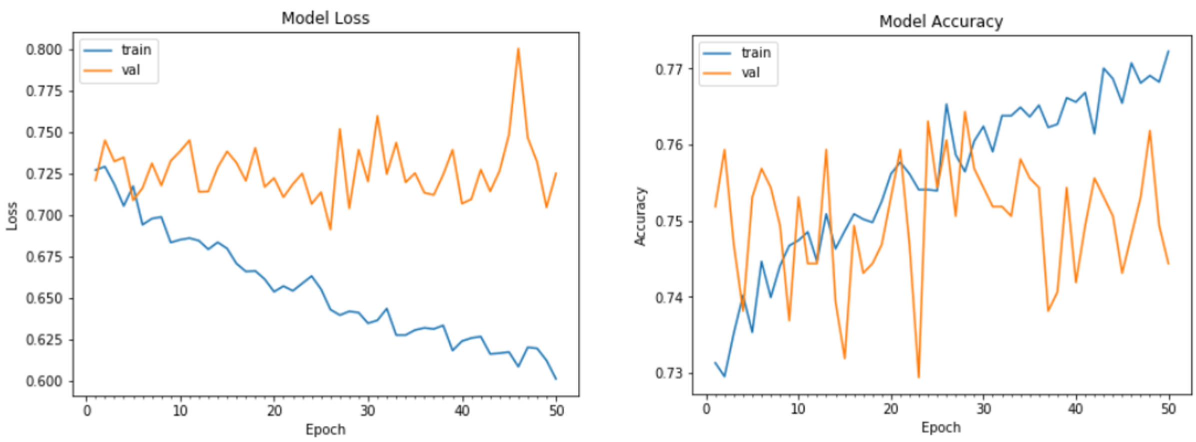

Figure 8: Loss and Accuracy Plot for Case 1

This case was performed to obtain a benchmark score for the comparison of all the cases. The images were directly loaded into the data frame without performing any pre-processing on it. It was resized to 100 x 75 pixels. As mentioned in section 6 the data was split into 70:10:20 ratio for train, validation and test respectively. From the Loss versus Epoch plot, it can be seen that the training loss reduces with increasing epoch but the validation loss is very unstable and keeps oscillating between 0.72 to 0.75 and spikes up to 0.80. This means that the validation loss doesn't reduce with increase in epoch and the model is not able to perform better on unseen data. This can also be seen from the accuracy versus epoch plot. Here the training accuracy increases with an increase in the number of epochs but the validation accuracy randomly oscillates and doesn't increase constantly throughout the training period. This is a case of overfitting. The model is trained on noisy data and it learns the noise in the data and therefore does not perform well on the unseen data. Even though the test accuracy achieved from this model is 69.97 % it is not reliable because it won't perform well on unseen data.

7.2 Model Training on data pre-processed with Morphological Operations:

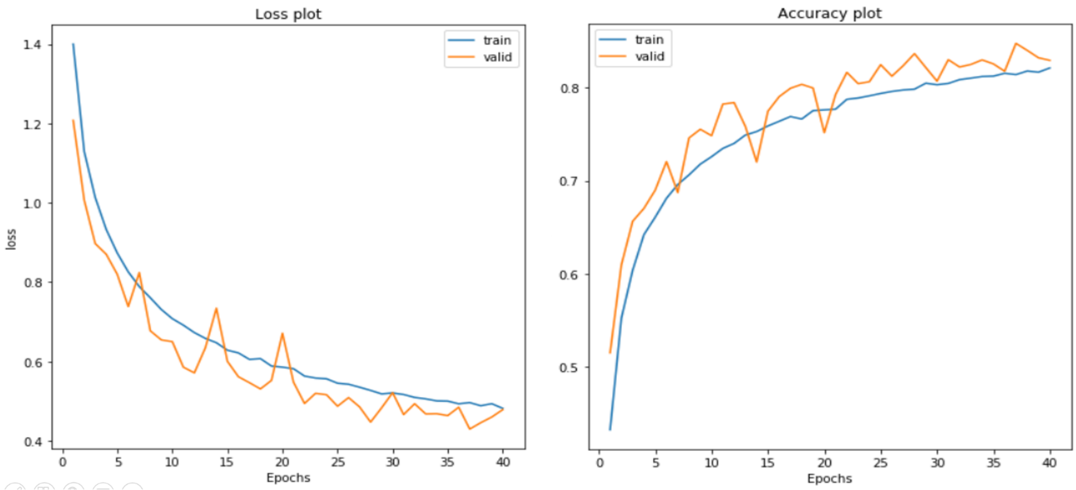

Figure 9: Loss and Accuracy Plot for Case 2

As discussed in sec 3.2.2 and 6 erosion and dilation was performed on the images to remove the noisy elements due to the body hair. This data was used to train the model and the results were recorded which can be seen in table 2. The loss versus epoch plot shows that training loss starts at 1.4 and keeps constantly reducing after every epoch till 0.4. The validation loss starts at 1.2 this is a sign that the model is actually learning from the training data and performing well on the validation data even on the first epoch, and steadily decreases with increasing epoch. In the accuracy plot we can see that the training, as well as the validation accuracy, grow constantly. This shows that the model is able to understand the patterns in the data and able to learn from it and perform well on validation data.

7.3 Model Training on Histogram Equalised data:

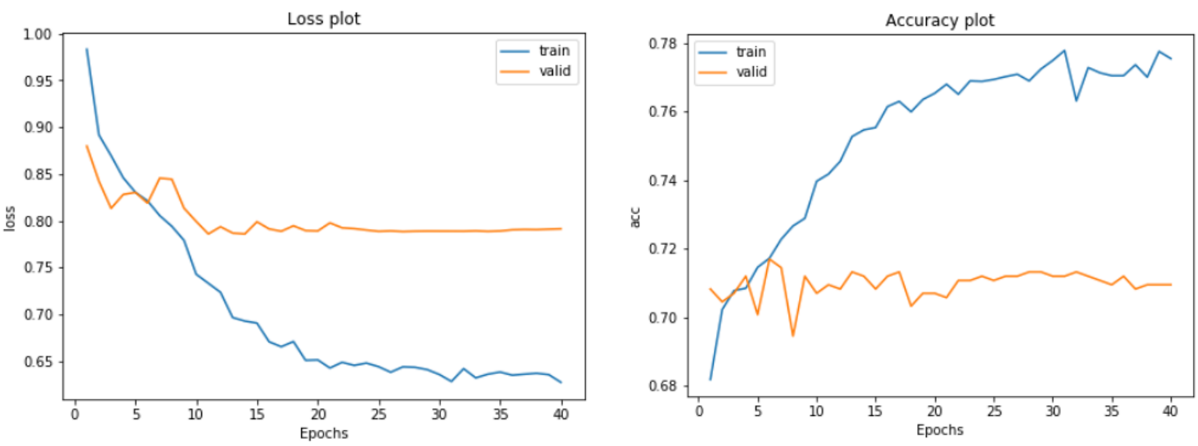

Figure 10: Loss and Accuracy Plot for Case 3

Performing Erosion and Dilation on the images improved the results, therefore histogram equalisation was performed with an attempt to enhance the results. The model was trained on these images. From the loss versus epoch plot, it can be seen that the results were not as expected. The training loss was reducing whereas the validation loss started reducing but remained constant at 0.80 after a few epochs. The accuracy plot was also similar in nature as the model learned even the irregularities from the training data but did not perform well on the validation data thus leading to overfitting. The reason for overfitting was image contrast enhancement of the non-region of interest area. Consider the figure 5, here the affected area is in the centre of the image but due to histogram equalisation the corners of the image have also become rich in colour. Thus the model learned from these patterns, therefore, was not able to perform well on test data.

7.4 Model Training on Up sampled data

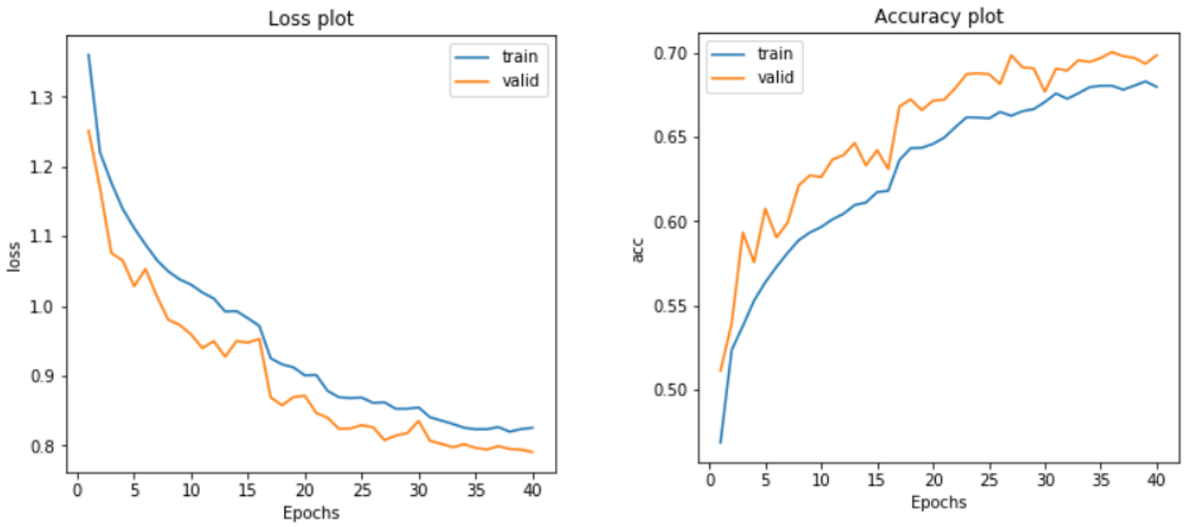

Figure 11: Loss and Accuracy Plot for Case 4

In order to treat the class imbalance in the data, the minority class data was over-sampled using image data generator. Erosion and Dilation were performed on this data and then the images were used to train the modified VGG19 model. From the loss plot, it can be observed that the model learned well from the training data and also performed well on the validation data so the loss reduced steadily with an increase in the number of epochs. Also in the accuracy plot, it can be observed that the model was able to achieve good accuracy on the validation data but was not able to perform well on test data and was able to achieve a score of 59 % test accuracy. The reason for this would be, the class imbalance in the test data. The training and the validation data were Oversampled but the test data cannot consist of duplicate or artificially generated data so the test set consisted of samples from the original dataset which was imbalanced thus the model was only able to correctly classify 59 % of the data.

The confusion matrix plot 12 plot is for case 2 7.2. The model is able to identify and classify the classes Melanocytic nevi, Melanoma, Benign keratosis-like lesions, Actinic keratoses and Basal cell carcinoma but does not do a good job at classifying the other two classes. Reason for which is elaborated in the discussion section 8.

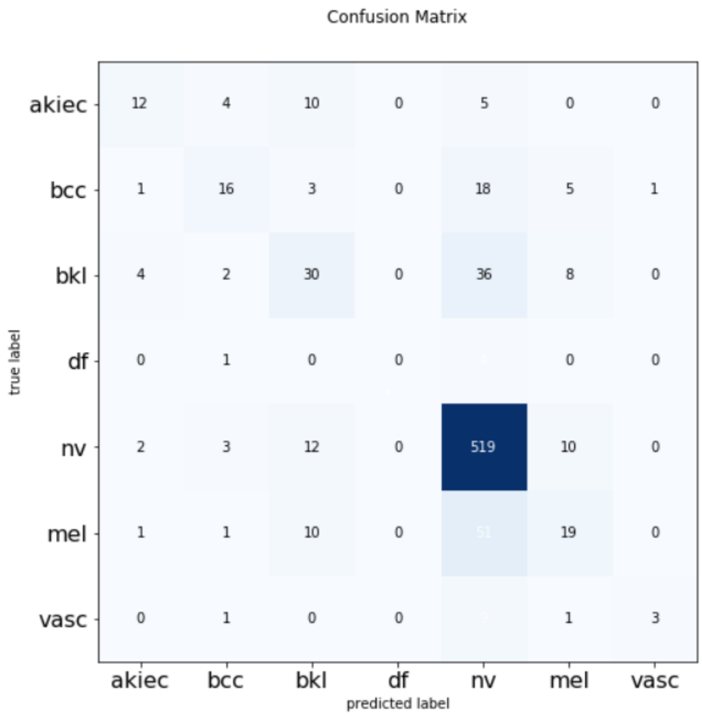

Figure 12: Confusion matrix for 9

8 Discussion

Multiple experiments have been performed and various approaches have been used to train the model to build a multi-class skin lesion classifier. In the previous section, the outcome and results of each experiment were stated. In this section, the reason and explanation for each chosen approach will be provided. The main objective of the research was to build a highly robust classifier. The metric chosen to evaluate the performance of the model was accuracy as classifying an image in the correct skin lesion type was of the highest importance. Case 1 7.1 was performed for benchmarking the scores. Case 2 7.2 was performed with an attempt to improve the results and this experiment gave the best results which are 78.68 % validation accuracy and 74.08 % test accuracy. This is because Morphological Operations were used to rid the images of noise present due to body hair. The use of this method was inspired by the works of Fa et al. (2019). With an attempt to improve the results further case 3 7.3 was performed on the images which were inspired from reading the works of Rahman et al. (2017) and Yan et al. (2019). The validation accuracy and the test accuracy did not improve and the reason for it can be observed in the images. Since the model was provided with the contrast-enhanced images the non-region of interest was also enhanced with high pixel values which in turn lead to overfitting. A better solution would be performing segmentation, contour extraction of the region of interest and masking of the non-region of interest in the images, and then training the model on these images. Various other attempts were made to improve the accuracy score by trying to build and train a CNN model from scratch with hyperparameter tuning. Other transfer learning models were trained to increase accuracy. One of the reasons why the performance could not be increased would be because of the presence of non-pigmented skin lesion cells like Dermatofibroma and Vascular lesions, on observing the samples of these cell types it can be understood that it is difficult to distinguish the region of interest and the unaffected area of the skin. This is because there is no variation in the colour of the skin. On the other hand cell types like Melanoma and Melanocytic nevi are pigmented skin lesions, which means the region of interest is distinguished from the unaffected region due to the brown and black pigments. This is also mentioned in the works of Tschandl et al. (2018).

9 Conclusion and Future Work

The objective of this study was to build a multi-class skin lesion classifier which detects seven different types of skin lesions. A modified VGG19 model was proposed to perform this task. The model was trained and tested on the HAM10000 dataset which was obtained from the ISIC 2018 challenge. The proposed model was run on raw data to get a baseline score for benchmarking purpose. Then the modified VGG19 model was also run on a data which was pre-processed with Erosion and Dilation techniques. The results from this experiment were better than all other experiments. Then the model was trained on histogram equalised images with an attempt to improve the validation and test accuracy. But performing histogram equalisation improved the contrast of the overall image which lead to the model learning the irregularities in the non-region of interest from the images and thus leading to model overfitting.

Oversampling technique was used to increase the samples in the minority classes. The test accuracy achieved for this experiment was the lowest as the imbalance in the training and validation set was handled but class imbalance did exist in the test set. The reason for that was that the model can learn via backpropagation on training and validation data, but on the test set, it cannot do the same. The best score achieved by the model was 74.08 % testing accuracy for case 2 7.2. Images of some skin lesion types like Dermatofibroma and Vascular lesions did not consist of high contrast pigmented cells in the region of interest, this is one of the reasons why the classifier accuracy could not be improved further. Pre-processing techniques like image segmentation, contour extraction and masking the non-region of interest would be suggested as future works.

10 Acknowledgement

I would like to thank Dr. Vladimir Milosavljevic for his mentoring, support and feedback. I would also like to thank my parents and friends for their constant support and motivation.

References

Alimboyong, C. R. and Hernandez, A. A. (2019). An Improved Deep Neural Network for Classification of Plant Seedling Images, *2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)* (March): 217–222.

Brinker, T. J., Hekler, A., Enk, A. H., Berking, C., Haferkamp, S., Hauschild, A., Weichenthal, M., Klode, J., Schadendorf, D., Holland-letz, T., Kalle, C. V. and Fro, S. (2019). ScienceDirect Deep neural networks are superior to dermatologists in melanoma image classification, **119**: 11–17.

Chatterjee, S., Dey, D. and Munshi, S. (2019). Computer Methods and Programs in Biomedicine Integration of morphological preprocessing and fractal based feature extraction with recursive feature elimination for skin lesion types classification, *Computer Methods and Programs in Biomedicine* **178**: 201–218.
URL: *https://doi.org/10.1016/j.cmpb.2019.06.018*

Cheng, H. D. and Shi, X. J. (2004). A simple and effective histogram equalization approach to image enhancement, *Digital Signal Processing: A Review Journal* **14**(2): 158–170.

Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Marchetti, M., Kittler, H. and Halpern, A. (2018). International Skin Imaging Collaboration (ISIC), pp. 1–12.

Fa, F., Peixoto, S. A. and Rebouc, P. P. (2019). ScienceDirect Automatic skin lesions segmentation based on a new morphological approach via geodesic active contour, **55**: 44–59.

Fan, X., Dai, M., Liu, C., Wu, F., Yan, X., Feng, Y. and Feng, Y. (2019). Effect of Image Noise on the Classification of Skin Lesions Using Deep Convolutional Neural Networks, *Tsinghua Science and Technology* **25**: 425–434.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). From data mining to knowledge discovery in databases, *AI Magazine* **17**(3): 37–53.

Gavrilov, D. A., Shchelkunov, N. N., Melerzanov, A. V. and Dolgoprudny, T. (2019). Deep learning based skin lesions diagnosis, **XLII**(May): 13–15.

Haralick, R. M., Sternberg, S. R. and Zhuang, X. (1987). Image Analysis Using Mathematical Morphology, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-9**(4): 532–550.

Harangi, B. (2018). Skin lesion classi fi cation with ensembles of deep convolutional neural networks, *Journal of Biomedical Informatics* **86**(June): 25–32.
URL: *https://doi.org/10.1016/j.jbi.2018.08.006*

Hekler, A., Utikal, J. S., Enk, A. H., Hauschild, A., Weichenthal, M., Maron, R. C., Berking, C., Haferkamp, S., Klode, J., Schadendorf, D., Schilling, B., Holland-letz, T., Izar, B., Kalle, C. V., Fro, S. and Brinker, T. J. (2019). ScienceDirect Superior skin cancer classification by the combination of human and artificial intelligence, **120**: 114–121.

Hosney, K. M., Id, M. A. K. and Foaud, M. M. (2019). Classification of skin lesions using transfer learning and augmentation with Alex-net, pp. 1–18.

Kalwa, U., Legner, C., Kong, T. and Pandey, S. (2019). Skin Cancer Diagnostics with an All-Inclusive Smartphone Application.

Ma, Z. and Tavares, J. M. R. (2016). A Novel Approach to Segment Skin Lesions in Dermoscopic Images Based on a Deformable Model, *IEEE Journal of Biomedical and Health Informatics* **20**(2): 615–623.

Mahbod, A., Schaefer, G., Ellinger, I., Ecker, R., Pitiot, A. and Wang, C. (2019). Computerized Medical Imaging and Graphics Fusing fine-tuned deep features for skin lesion classification, *Computerized Medical Imaging and Graphics* **71**: 19–29.
URL: *https://doi.org/10.1016/j.compmedimag.2018.10.007*

Marchetti, M. A., Codella, N. C., Dusza, S. W., Gutman, D. A., Helba, B., Kalloo, A., Mishra, N., Carrera, C., Celebi, M. E., DeFazio, J. L., Jaimes, N., Marghoob, A. A., Quigley, E., Scope, A., Yélamos, O. and Halpern, A. C. (2018). Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images, *Journal of the American Academy of Dermatology* **78**(2): 270–277.e1.

Maron, R. C., Weichenthal, M., Utikal, J. S., Hekler, A., Berking, C., Hauschild, A., Enk, A. H., Haferkamp, S., Klode, J., Schadendorf, D., Jansen, P., Holland-letz, T., Schilling, B., Kalle, C. V., Fro, S., Gaiser, M. R., Hartmann, D., Gesierich, A., Ka, K. C. and Brinker, T. J. (2019). ScienceDirect Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks, **119**: 57–65.

Navarro, F., Escudero-Viñolo, M. and Bescós, J. (2019). Accurate Segmentation and Registration of Skin Lesion Images to Evaluate Lesion Change, *IEEE Journal of Biomedical and Health Informatics* **23**(2): 501–508.

Pathan, S., Prabhu, K. G. and Siddalingaswamy, P. C. (2018). Hair detection and lesion segmentation in dermoscopic images using domain knowledge, pp. 2051–2065.

Rahman, M. A., Haque, M. T. and Fattah, S. A. (2017). Skin Lesions Classification Based on Color Plane-Histogram-Image Quality Analysis Features Extracted from Digital Images, pp. 1356–1359.

Riaz, F., Nawaz, R. and Coimbra, M. (2019). Active Contours Based Segmentation and Lesion Periphery Analysis for Characterization of Skin Lesions in Dermoscopy Images, *IEEE Journal of Biomedical and Health Informatics* **23**(2): 489–500.

Sae-lim, W., Wettayaprasit, W. and Aiyarak, P. (2019). Convolutional Neural Networks Using MobileNet for Skin Lesion Classification, *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)* pp. 242–247.

Seeja, R. D. and Suresh, A. (2019). Melanoma Segmentation and Classification using Deep Learning, (12).

Serte, S. and Demirel, H. (2019). Gabor wavelet-based deep learning for skin lesion classification, *Computers in Biology and Medicine* **113**(August): 103423.
URL: *https://doi.org/10.1016/j.compbiomed.2019.103423*

Simonyan, K. and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition, pp. 1–14.
URL: *http://arxiv.org/abs/1409.1556*

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.* **15**(1): 1929–1958.
URL: *http://dl.acm.org/citation.cfm?id=2627435.2670313*

Tschandl, P., Rosendahl, C. and Kittler, H. (2018). Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, *Scientific Data* **5**: 1–9.

Tschandl, P., Sinz, C. and Kittler, H. (2019). Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation, *Computers in Biology and Medicine* **104**(November 2018): 111–116.
URL: *https://doi.org/10.1016/j.compbiomed.2018.11.010*

Wirth, R. and Hipp, J. (1995). CRISP-DM : Towards a Standard Process Model for Data Mining, *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining* (24959): 29–39.

Yan, T., Zhang, L. and Peng, C. (2019). Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models, *Applied Soft Computing Journal* **84**: 105725.
URL: *https://doi.org/10.1016/j.asoc.2019.105725*

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. and Lipson, H. (2015). Understanding Neural Networks Through Deep Visualization.
URL: *http://arxiv.org/abs/1506.06579*

Zhang, N., Cai, Y.-x., Wang, Y.-y., Tian, Y.-t., Wang, X.-l. and Badami, B. (2019). Skin cancer diagnosis based on optimized convolutional neural network, *Artificial Intelligence In Medicine* p. 101756.
URL: *https://doi.org/10.1016/j.artmed.2019.101756*