CS512-
Segmentation and Classification of Skin Lesions

Authors,
Akshay R (A20442409)
Sourav Yadav(A20450418)

Abstract:

We are going to work on a project to identify the type of skin cancer by segmentation and classification which will save a lot of time and effort taken to review each and every report manually. In our opinion, transformation driven by technology has changed the paradigm of human interaction with data. Usually, training of neural networks for automated diagnosis of pigmented skin lesions is hampered by the small size and lack of diversity of available dataset of dermatoscopic images. We are going to tackle this problem by taking reference to the HAM10000 "Human Against Machine with 10000 training images" Kaggle dataset where data is collected as dermatoscopic images from different populations, acquired and stored by different modalities. In light of this objective, with the analysis of a dataset of skin lesions, we seek to derive a general analysis and modeling approach to segment and predict the diagnostic carcinoma category of skin lesions.

Overview:

Skin cancer is a common disease that affects a large number of people. Some worrying facts about skin cancer are as follows:
Every year there are more new cases of skin cancer than the combined incidence of cancers of the breast, prostate, lung, and colon.
An estimated 87,110 new cases of invasive melanoma will be diagnosed in the U.S. in 2017.
The estimated 5-year survival rate for patients whose melanoma is detected early is about 98 percent in the U.S. The survival rate falls to 62 percent when the disease reaches the lymph nodes, and 18 percent when the disease metastasizes to distant organs.
Early detection is critical!

The dataset we have contains images that belong to the following seven categories. They are:

Nv - Melanocytic nevi [6705 images]:
Melanocytic nevi are benign neoplasms of melanocytes and appear in a myriad of variants, which are all included in our series. The variants may differ significantly from a dermatoscopic point of view

Mel - Melanoma [1113 images]:
Melanoma is a malignant neoplasm derived from melanocytes that may appear in different variants. If excised at an early stage it can be cured by simple surgical excision. Melanomas can be invasive or non-invasive (in situ). We included all variants of melanoma including melanoma in situ but did exclude non-pigmented, subungual, ocular or mucosal melanoma.

Bkl - Benign keratosis [1099 images]:
"Benign keratosis" is a generic class that includes seborrheic ker- atoses ("senile wart"), solar lentigo - which can be regarded a flat variant of seborrheic keratosis - and lichen-planus like keratoses (LPLK), which corresponds to a seborrheic keratosis or a solar lentigo with inflammation and regression.The three subgroups may look different dermatoscopically, but we grouped them together because they are similar biologically and often reported under the same generic term histopathologically. From a dermatoscopic view, lichen planus-like keratosis is especially challenging because they can show morphologic features mimicking melanoma and are often biopsied or excised for diagnostic reasons.

Bcc - Basal cell carcinoma [514 images]:
Basal cell carcinoma is a common variant of epithelial skin cancer that rarely metastasizes but grows destructively if untreated. It appears in different morphologic variants (flat, nodular, pigmented, cystic, etc), which are all included in this set.

Akiec - Actinic Keratoses and intraepithelial Carcinoma [327 images]:
Actinic Keratoses (Solar Keratoses) and intraepithelial Carcinoma (Bowen's disease) are common non-invasive variants of squamous cell carcinoma that can be treated locally without surgery. Some authors regard them as precursors of squamous cell carcinomas and not as actual carci- nomas. There is, however, agreement that these lesions may progress to invasive squamous cell carcinoma - which is usually not pigmented. Both neoplasms commonly show surface scaling and commonly are devoid of pigment. Actinic keratoses are more common on the face and Bowen's disease is more common on other body sites. Because both types are induced by UV-light the surrounding skin is usually typified by severe sun damage except in cases of Bowen's disease that are caused by human papillomavirus infection and not by UV. Pigmented variants exist for Bowen's disease and actinic keratoses. Both are included in this set.

Vasc - Vascular [142 images]:
Vascular skin lesions in the dataset range from cherry angiomas to angiokeratomas [25] and pyogenic granulomas [26]. Hemorrhage is also included in this category.

Df - Dermatofibroma [115 images]:
Dermatofibroma is a benign skin lesion regarded as either a benign proliferation or an inflammatory reaction to minimal trauma. It is brown, often showing a central zone of fibrosis dermatoscopically.

The explanation on type of the skin cancer are referred from this link - https://www.nature.com/articles/sdata2018161

Exploratory Data Analysis:

The data that we have consisted of two types. Namely,

Flat file:
lesion_id – ID for individual person's image data
image_id – ID for individual image
dx – Type of cancer

- Melanocytic nevi - nv
- Melanoma - mel
- Benign keratosis-like lesions - Bkl
- Basal cell carcinoma - Bcc
- Actinic keratoses - Akiec
- Vascular lesions - Vasc
- Dermatofibroma - Df

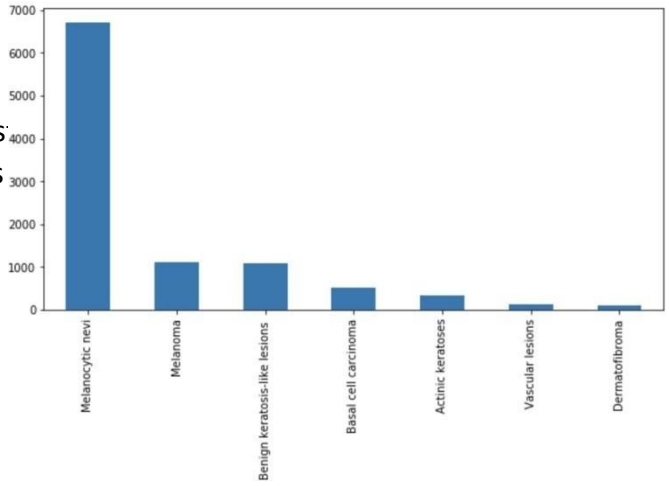
dx_type – Diagnosis type

- Histopathology - histo
- Confocal - confocal
- Follow up – follow_up
- Consensus - consensus

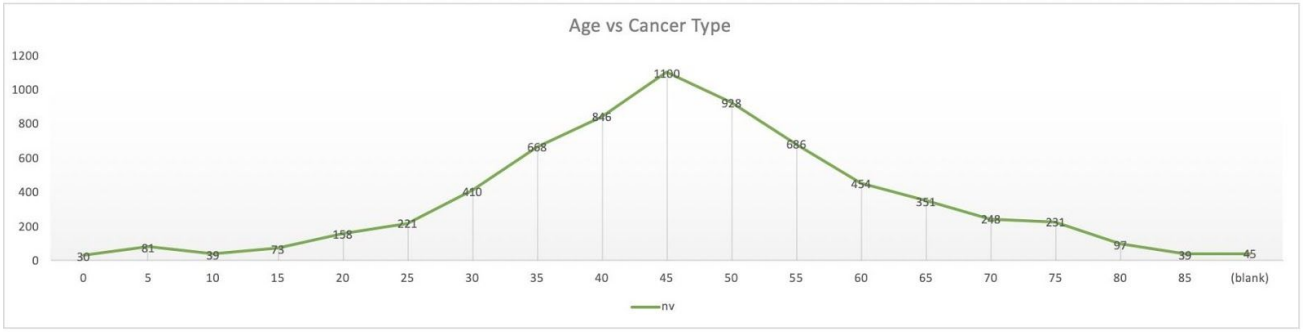
age – Age of the patient
sex – Gender of the patient (M/F)
localization – Body part of the patient, where the cancer is diagnosed.
.

Images:
Count - 10,015
Resolution – 600x450
Unique – 5,514
Duplicated – 4,501

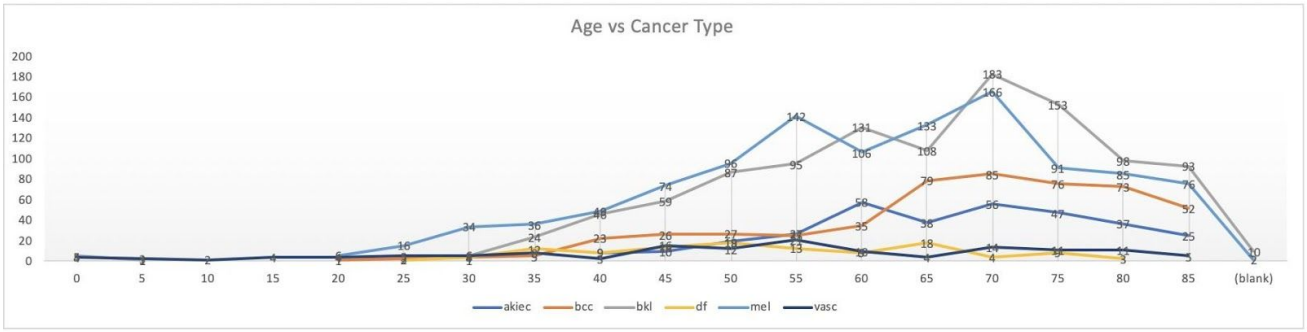
In order to understand the class (lesion type) dis
It is clearly evident that 70% of the data belongs
dealing with an imbalance dataset.

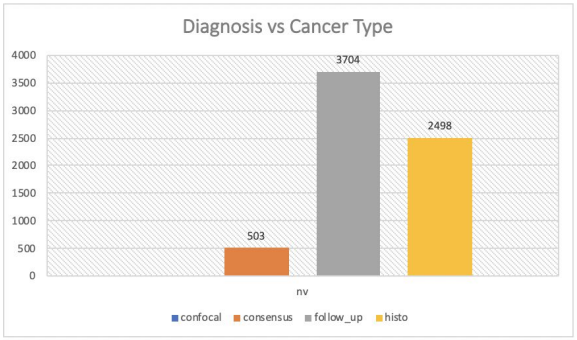


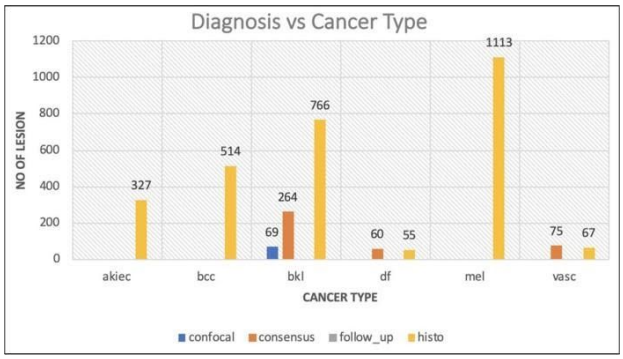
Age Distribution for Cancer type nv:

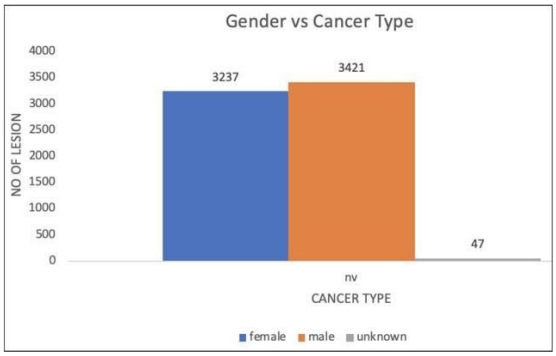


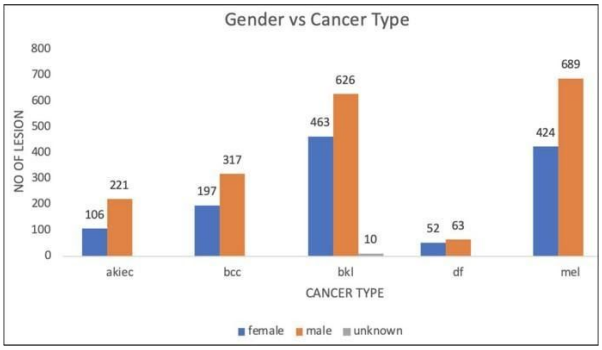
Age Distribution for all other Cancer Types:



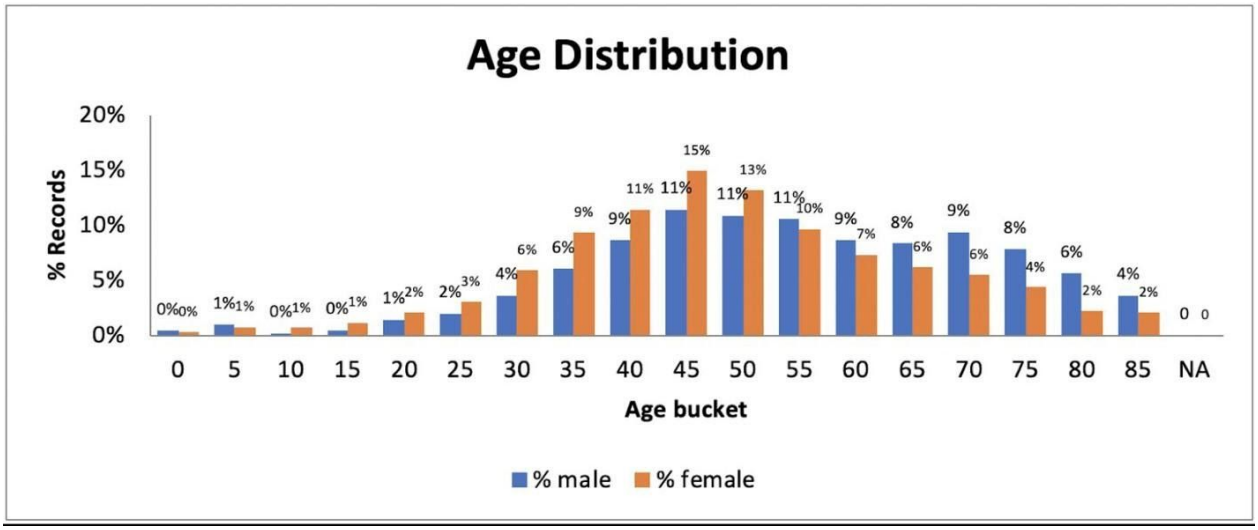






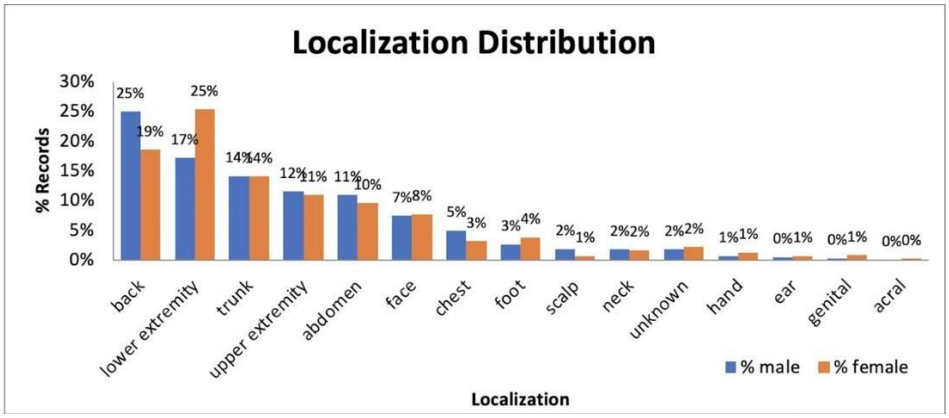


Age distribution of the observation in the dataset is plotted below:



From the age distribution plot, we can infer that 90% of the observations lie between the age range 30 to 80.

Localization distribution of the observation in the dataset is plotted below:



Proposed solution:

The image resolution of raw data is equal to 600x450. We reduce the dimension of the images to 100x75 for classification. During exploratory data analysis, it is evident that we are dealing with a large class imbalance. We try to address this issue by duplicating the train images for the classes that have less data and then performing data augmentation on these duplicated images in order to generate a balanced distribution of classes.

The image tensor is standardized using a mean-standard deviation approach. We calculate the mean and standard deviation for each of the 3 channels and standardize them separately.

First, we train a basic CNN model with several convolution, MaxPooling, BatchNormalization and Dropout layers. We evaluate the results on RMSprop with the help of a confusion matrix and AUC-ROC curves.

We extend our existing architecture with multiple InseptionResnet blocks to the previous model and evaluate the results in a similar way. To compare the performance of the problem specific architecture from the previous steps, we train an inceptionv3 classifier with pre-trained weights and connect it with a dense layer on top of it. We have three different versions of the same model here. First, we freeze all the weights and train the model. Then we unfreeze the top 19 layers for better feature extraction. and finally, we unfreeze all layers.

We then train a DenseNet121 model with all layers unfrozen and assess the performance of all the classifiers.

We then train a Unet model to predict the segmentation boundaries for our images. The model is trained using training data with their corresponding segmentation masks (200 images and corresponding masks). The model weights are saved.

We then run our pre-trained classifier along with the segmentation model to generate segmented classification results.

Methodology:

Data preprocessing:

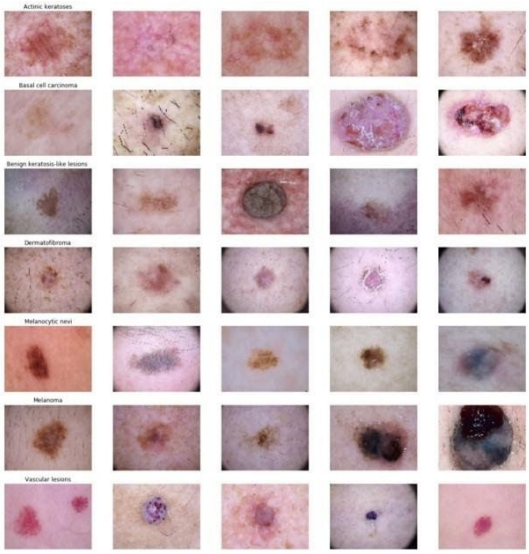
The initial step is to figure out the actual class of the images provided. Hence, we create a new column which maps the image id to the type of skin lesion.

During the exploratory data analysis, we observed that there are 57 observations with missing age value. Hence, we impute these values by their mean.

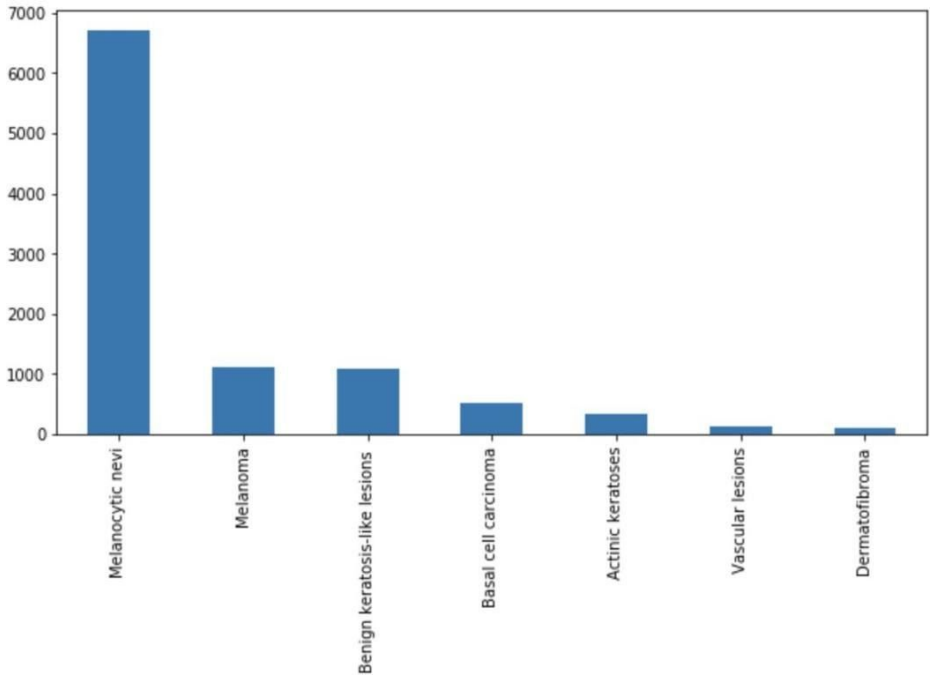
Loading and resizing of images:

- In this step, the images will be loaded into the column named image. We also resize the images as the original dimension of images are 450 x 600 x 3 which would increase the computational time to process the data. Hence, we choose a considerably scaled image of the dimension 100 x 75 x 3

Rescaled images are shown below:



Upsampling:



As we can clearly see from the image, there is a huge class imbalance in the dataset. We should be careful about 2 things in this scenario

1. Accuracy is not the right metric to access the models. We should compare the models based on precision, recall and f1 score. This is because the classifier might predict correctly only for the class which has maximum data
2. To make the classifier work well even for classes with less data points, we have to balance our train dataset. This can be done by adding duplicate versions of the same images and then alter them using various data augmentation techniques

Model architectures:

Model 1 – Basic Convolution Neural Network:

The architecture of this block is as follows:

Layer (type)	Output Shape	Param #
input_5 (InputLayer)	(None, 75, 100, 3)	0
conv2d_73 (Conv2D)	(None, 75, 100, 32)	896
conv2d_74 (Conv2D)	(None, 75, 100, 32)	9248
max_pooling2d_25 (MaxPooling	(None, 37, 50, 32)	0
batch_normalization_21 (Batc	(None, 37, 50, 32)	128
dropout_13 (Dropout)	(None, 37, 50, 32)	0
conv2d_75 (Conv2D)	(None, 37, 50, 64)	18496
conv2d_76 (Conv2D)	(None, 37, 50, 64)	36928
max_pooling2d_26 (MaxPooling	(None, 18, 25, 64)	0
batch_normalization_22 (Batc	(None, 18, 25, 64)	256
dropout_14 (Dropout)	(None, 18, 25, 64)	0
flatten_5 (Flatten)	(None, 28800)	0
dense_9 (Dense)	(None, 128)	3686528
batch_normalization_23 (Batc	(None, 128)	512
dropout_15 (Dropout)	(None, 128)	0
dense_10 (Dense)	(None, 7)	903

Total params: 3,753,895
Trainable params: 3,753,447
Non-trainable params: 448

This model is trained on the balanced train data and is first ran on RMSprop optimizer with 'categorical cross entropy' as the loss function as this is a multiclass classification problem. Confusion matrix is generated plotting true label vs predicted label to assess the performance of the classifier. Since there are 7 classes, we get a 7x7 confusion matrix.

From the confusion matrix we calculate precision, recall and f1-score for this model using the following formulas:

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive} = \frac{True\ Positive}{Total\ Predicted\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative} = \frac{True\ Positive}{Total\ Actual\ Positive}$$

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

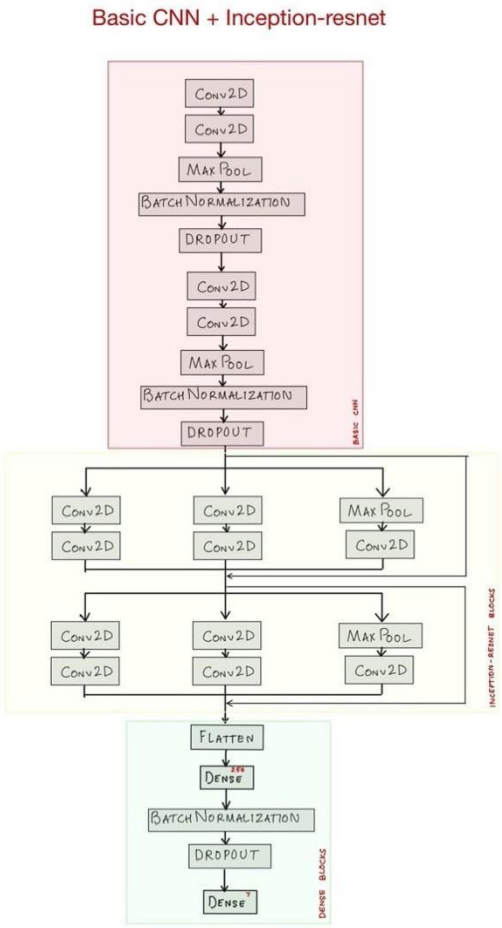
The results are then used to plot the AUC-ROC curve using true positive (TPR) and false positive rate (FPR).

$$\text{TPR /Recall / Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{FPR} = 1 - \text{Specificity} = \frac{FP}{TN + FP}$$

.

Model 2 a - CNN +InceptionResnet (RMSprop):



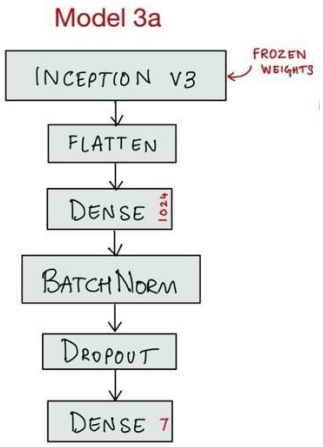
We decided to try to build our own architecture using the previous convolution model as base and passing it as input to an InceptionResnet block. The diagram shows the architecture of the new model. Two blocks of InceptionResnet are added to the Basic CNN model and are trained on a balanced dataset. The optimizer used here is RMSprop and the loss function used here is categorical crossentropy.

Model 2 b - CNN +InceptionResnet (SGD):

The same model is tested using an SGD optimizer to understand the impact of different optimizers on test data when trained on the same environment with similar learning rates

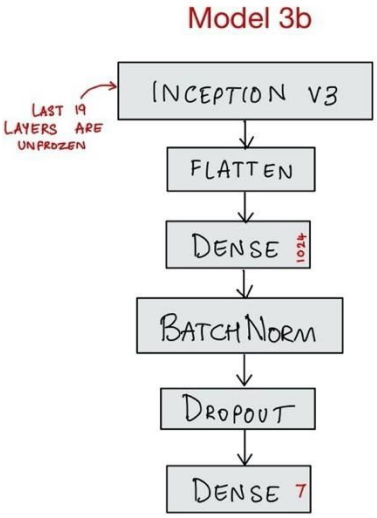
Model 3a - InceptionV3 (frozen weights):

We used an InceptionV3 classifier trained on ImageNet dataset as the base model and developed our architecture with 1024 neurons as the middle dense layer and an output layer with 7 neurons, one for each class. With RMSprop as the optimizer and categorical crossentropy as the loss function, the model performed poorly on the HAM10000 dataset. This is because the features extracted from the convolution layers of InceptionV3 does not match with our skin cancer images



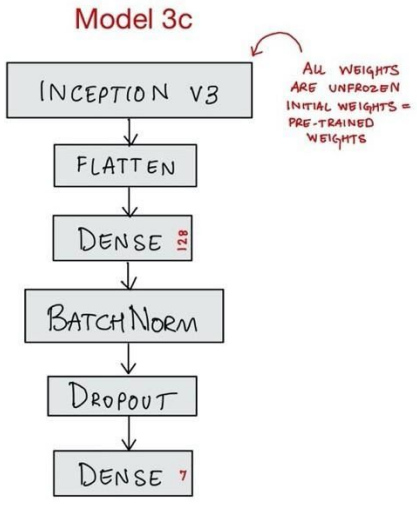
Model 3b - InceptionV3 (unfrozen top 19 layers):

To fine tune the above model by unfreezing the top 19 layers and retrain the weights. The classifier performed as expected.



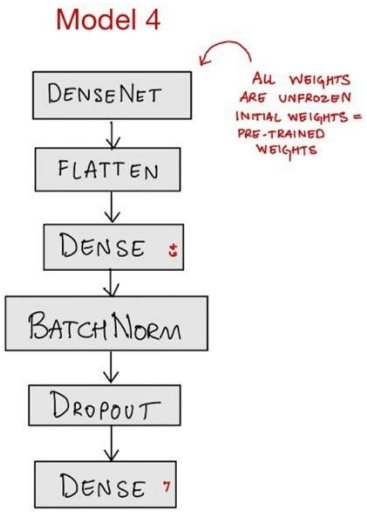
Model 3c - InceptionV3 (all unfrozen weights) (RMSprop):

Though there was an increase in the performance by unfreezing the top layers, it was evident that our dataset is very different to the ImageNet dataset as the classifier was still underperforming when compared to our base CNN. Hence, using a model pretrained on the ImageNet dataset didn't make sense. We unfroze all the weights of the inceptionV3 model and trained this with a dense architecture. However, to achieve faster convergence, we initialized the model with the pre-trained ImageNet weights rather than random weights.

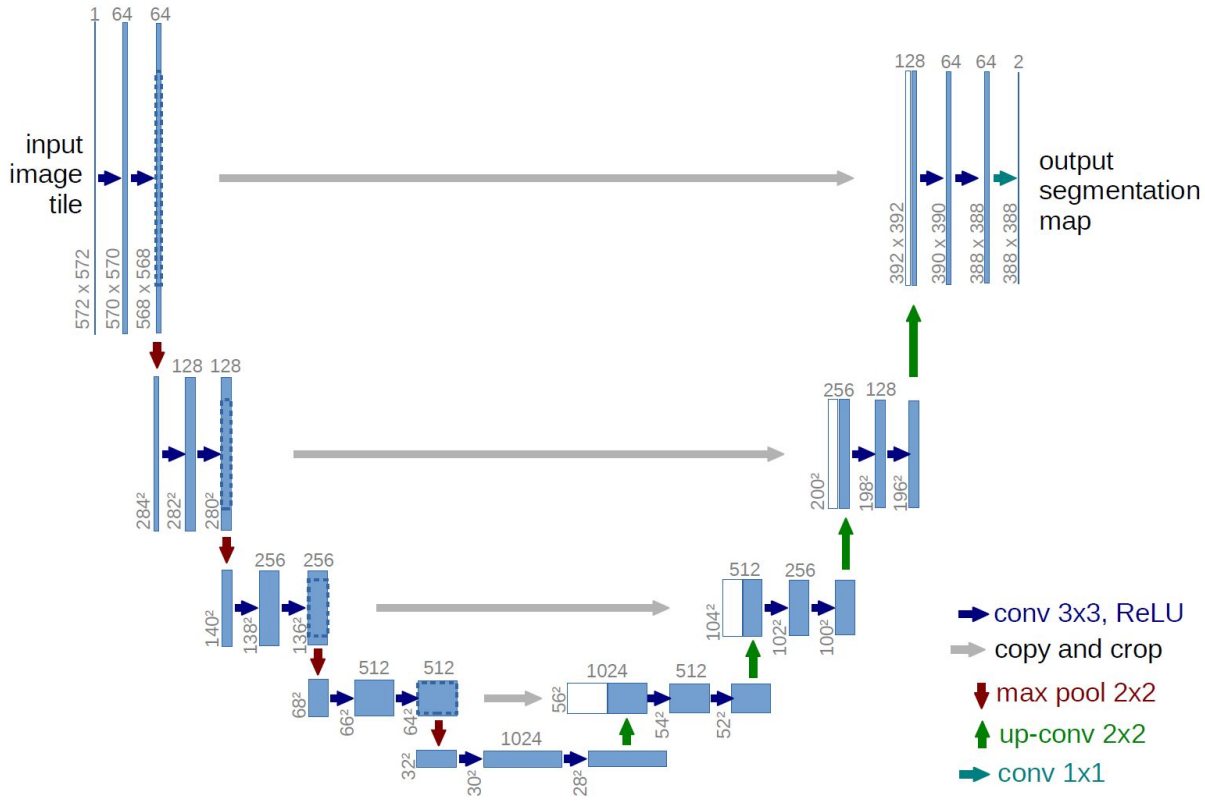


Model 4 – DenseNet (RMSprop):

As our final model we unfroze all the weights of the DenseNet121 model and trained this with a dense architecture with the pre trained weights as the initial weights for faster convergence.



Unet Model:

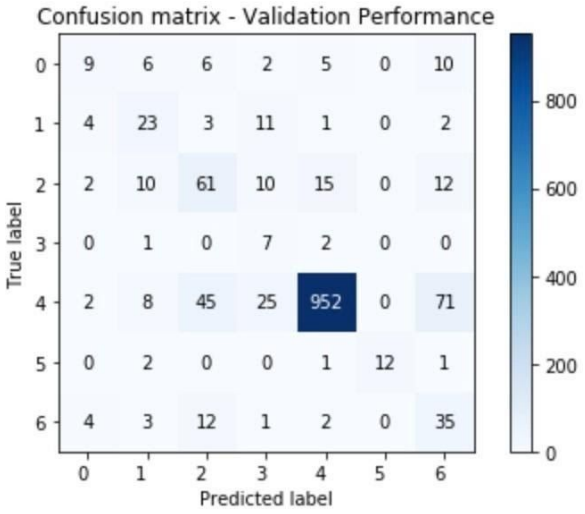


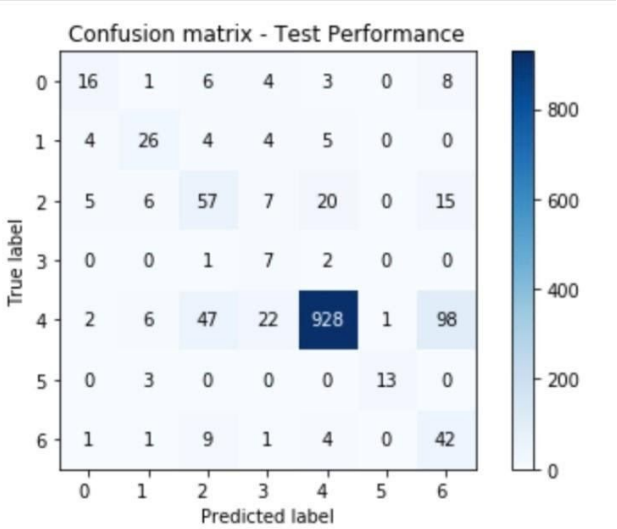
We train the unet model with 200 images to segment the image by generating a binary mask which we later use in combination with the original image for classification.

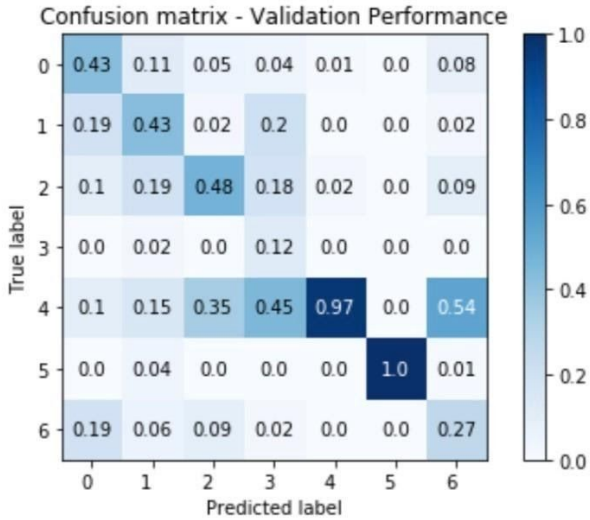
Results:

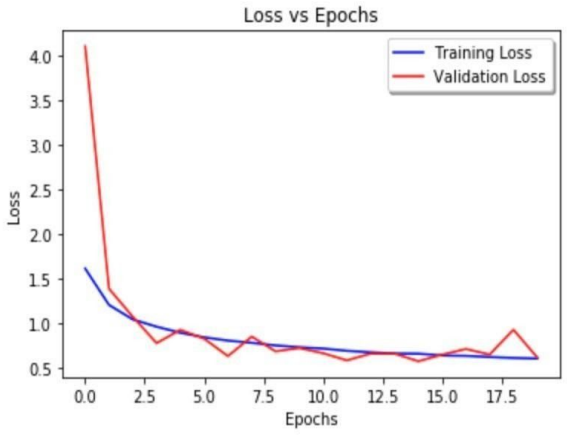
Here are the results of our best performing models:

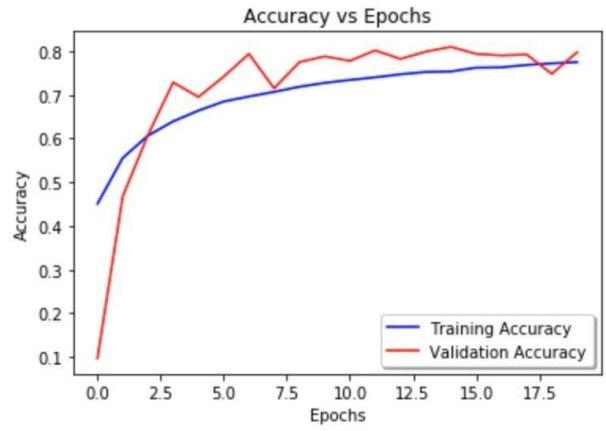
Model1 : Basic convolution model:



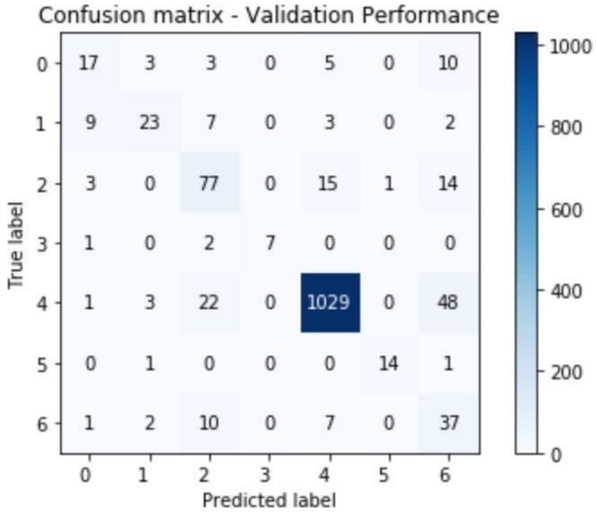


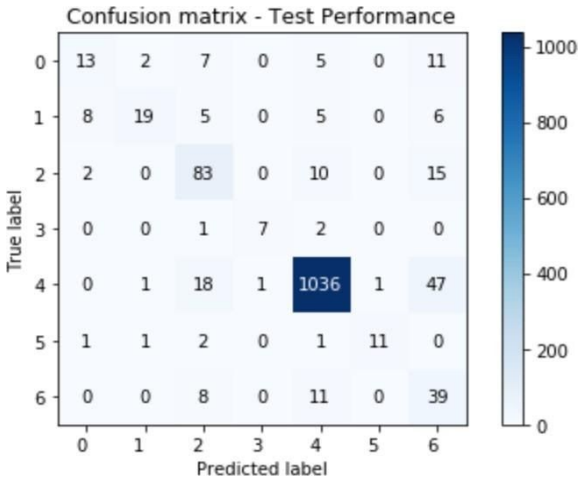


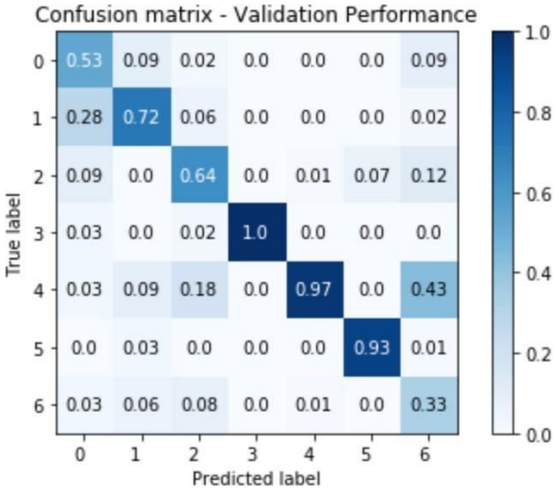


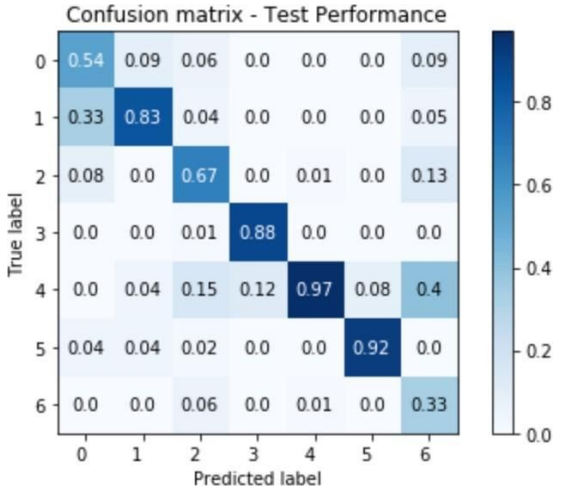


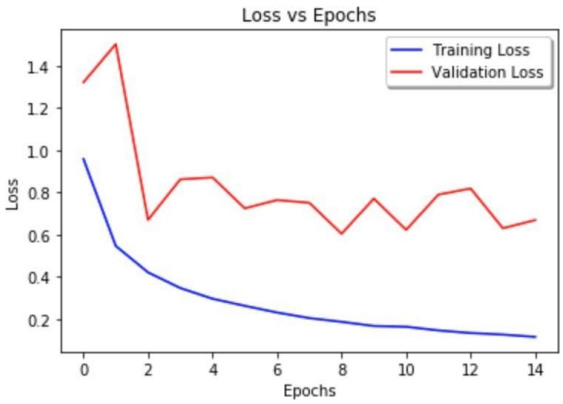
Model 4 – DenseNet (RMSprop):

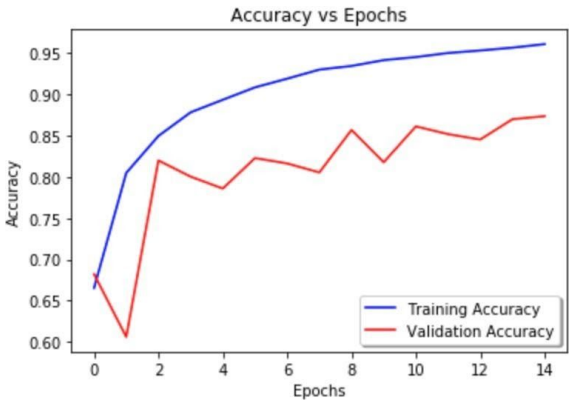






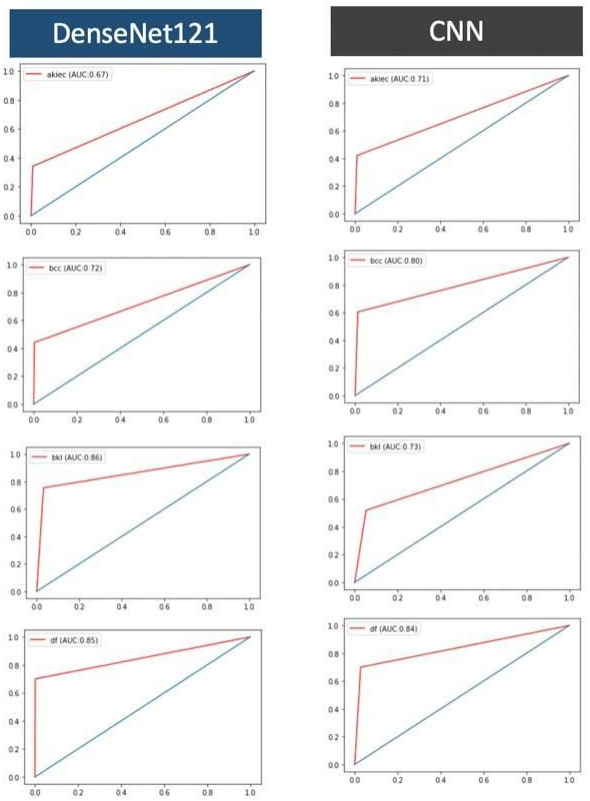


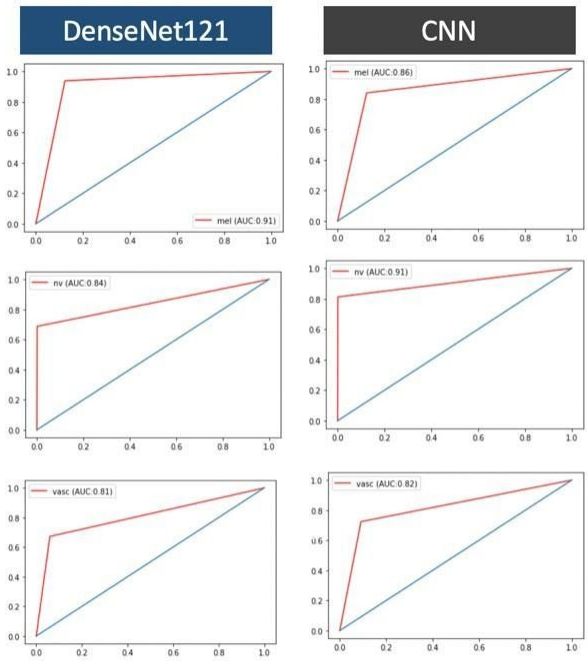




Model Name	Total Parameters	Non-Trainable Parameters	Optimizer	Learning Rate	Learning Rate decay	Batch Size	Average Time per epoch (seconds)	Validation Accuracy	Test Accuracy	AUC-ROC score
CNN	3,753,895	448	RMSprop	0.0005	N	128	340	80.99%	78.97%	0.80735
CNN + InceptionResnet (RMSprop)	1,607,207	1,216	RMSprop	0.001	Y	128	430	86.50%	83.60%	0.77622
CNN + InceptionResnet (SGD)	1,607,207	1,216	SGD	0.0005	Y	256	580	74.53%	73.89%	0.77502
InceptionV3 (Frozen Weights)	21,802,784	21,802,784	RMSprop	0.001	N	128	115	20.90%	19.87%	NA
InceptionV3 (Unfroze top 19 layers)	21,802,784		RMSprop	0.0001	N	128	240	77.43%	76.36%	0.71128
InceptionV3 (RMSprop)	22,66,471	34,688	RMSprop	0.0001	N	128	590	88.75%	87.88%	0.79243
InceptionV3 (SGD)	22,66,471	34,688	SGD	0.0001	Y	256	290	89.04%	87.96%	0.77261
DenseNet121	7,431,495	83,776	RMSprop	0.001	Y	128	1219	87.37%	87.59%	0.80773

The above table compares the performance of different model architectures on test data. From the table one might argue that CNN and DenseNet121 are the best performing models. But it is really so? Let's examine the AUC-ROC curves for both the above mentioned architectures

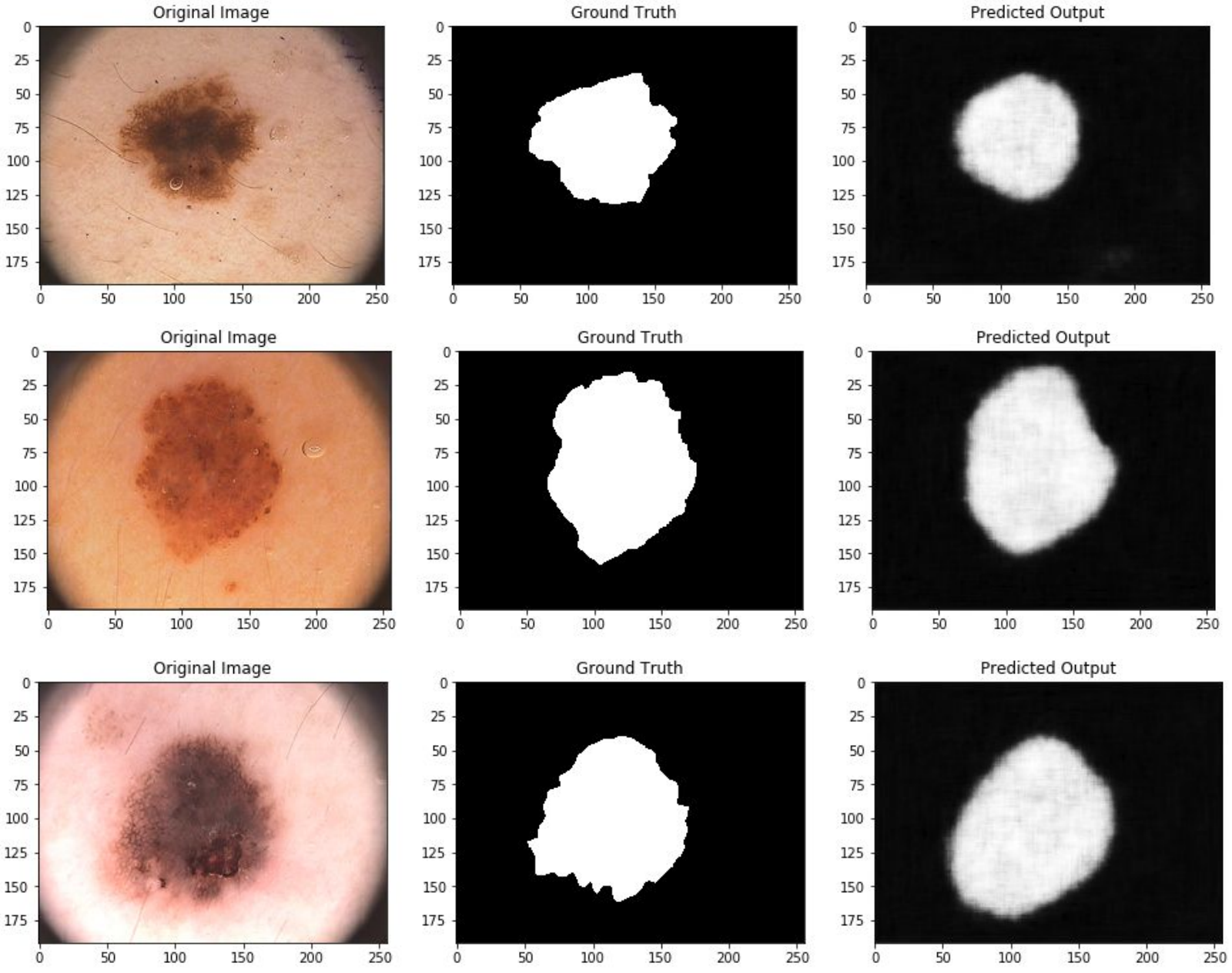




Label	Skin Lesion Type	Support
0	Akiec	38
1	Bcc	43
2	Bkl	110
3	Df	10
4	nv	1104
5	vasc	16
6	mel	58

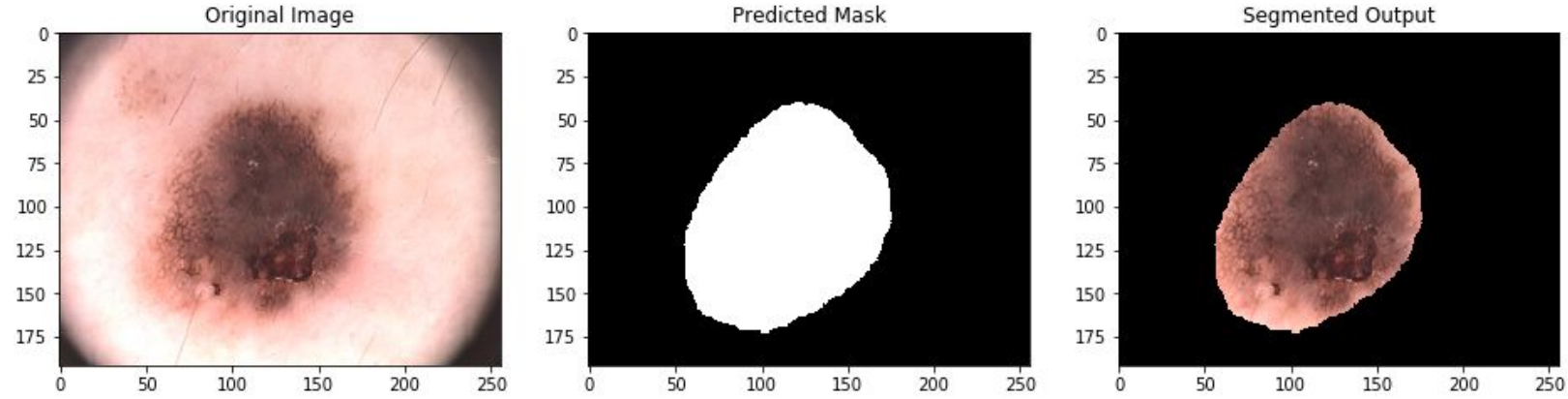
From the above AUC-ROC graphs, we observe that the CNN model has a high AUC_ROC score for skin lesion types nv and bkl. But it is gravely underperforming for the other minority classes. In a practical scenario, we would like our model classifier to have high precision and recall. It is better to have false positives than false negatives. In account of these points, and based on the performance, the DenseNet121 is the best performing classifier.

Segmentation results:

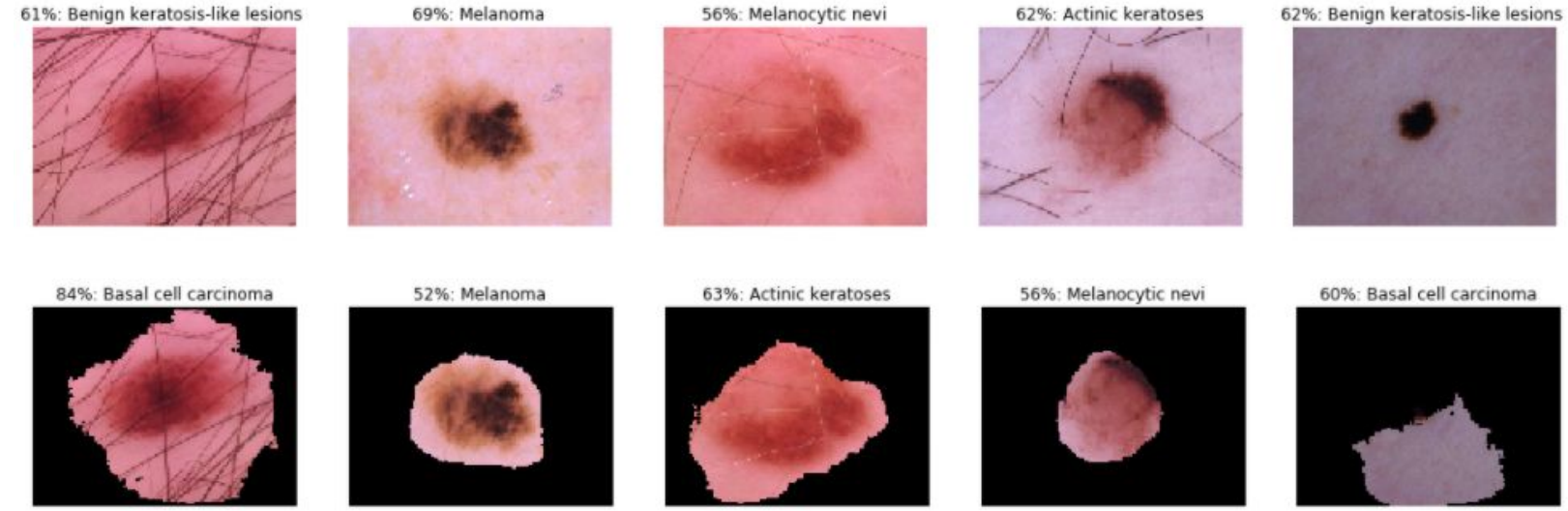


Unet was trained for 100 epochs and the predicted masks are as shown in the above diagrams.

Applying segmented masks the output can be visualized as shown in the figure below:



These segmented images are used to predict the class using ensemble models (VGG16 + ResNet50) trained on the HAM10000 RGB images. The results are shown in the diagram below.



Reference:

1. https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9
2. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5
3. https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000
4. https://www.researchgate.net/profile/Li_Jia_Li/publication/221361415_ImageNet_a_Large-Scale_Hierarchical_Image_Database/links/00b495388120dbc339000000/ImageNet-a-Large-Scale-Hierarchical-Image-Database.pdf
5.https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neuralnetworks-the-eli5-way-3bd2b1164a53
6. https://ml-cheatsheet.readthedocs.io/en/latest/backpropagation.html
7.https://github.com/fredguth/skin-cancer/blob/68ea3866a1ba0579c4b68380025c179fdc8363ff/ISIC-Task1-note5-Grupo_de_Estudo.ipynb
8.https://github.com/domingomery/imagenes/tree/7226b36731b87d0eb550843ebd4228d5251ca78d/proyecto