Skin Cancer Detection

A Data Mining and Machine Learning Project

Roberto Di Lauro
D03000170

Roberto Tessitore
D03000122

Giovanni De Francesco
D03000161

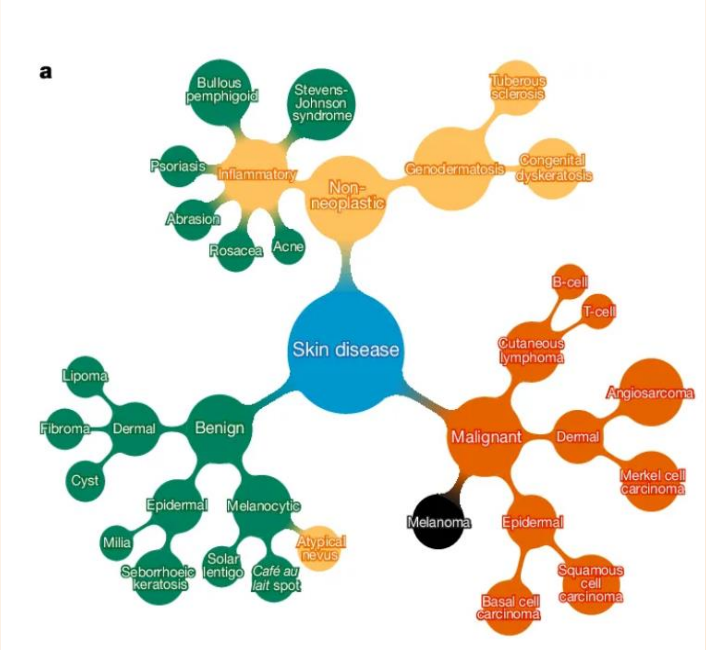
Table of contents

1 Problem Introduction and Domain understanding

- **Skin cancer** is one the most common malignancy worldwide, with **early detection** being critical for effective treatment.

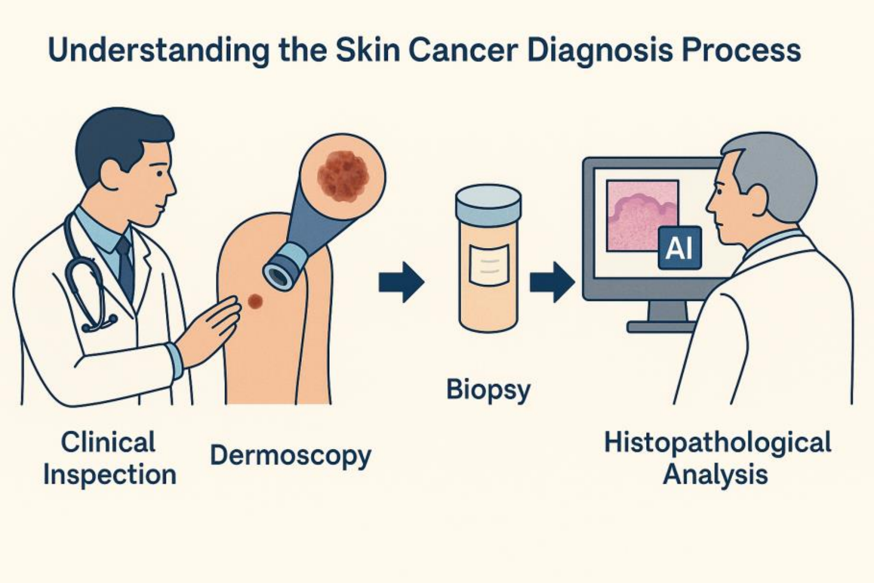
- Diagnosis typically begins with clinical and dermoscopic inspection, but access to expert dermatologists remains limited in many areas.

- **Computer-Aided Diagnosis** (CAD) systems offer a scalable solution, **assisting clinicians** by analyzing dermatoscopic images with machine and deep learning models

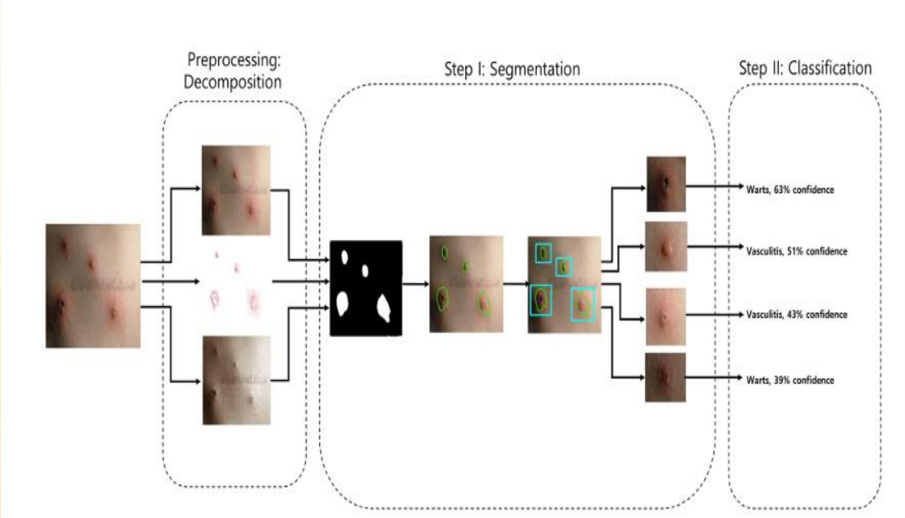


This project leverages the HAM10000 dataset and deep learning techniques **to develop a reproducible pipeline for multi-class skin lesion segmentation and classification**

1 Problem Introduction and Domain understanding



Segmentation techniques and deep learning models help to **extract meaningful patterns** and automate the diagnostic process, enabling faster and earlier results.



Traditional diagnosis relies on domain knowledge and **handcrafted features** (e. g. shape, color, texture) which **are limited in their ability** to capture subtle or complex patterns.

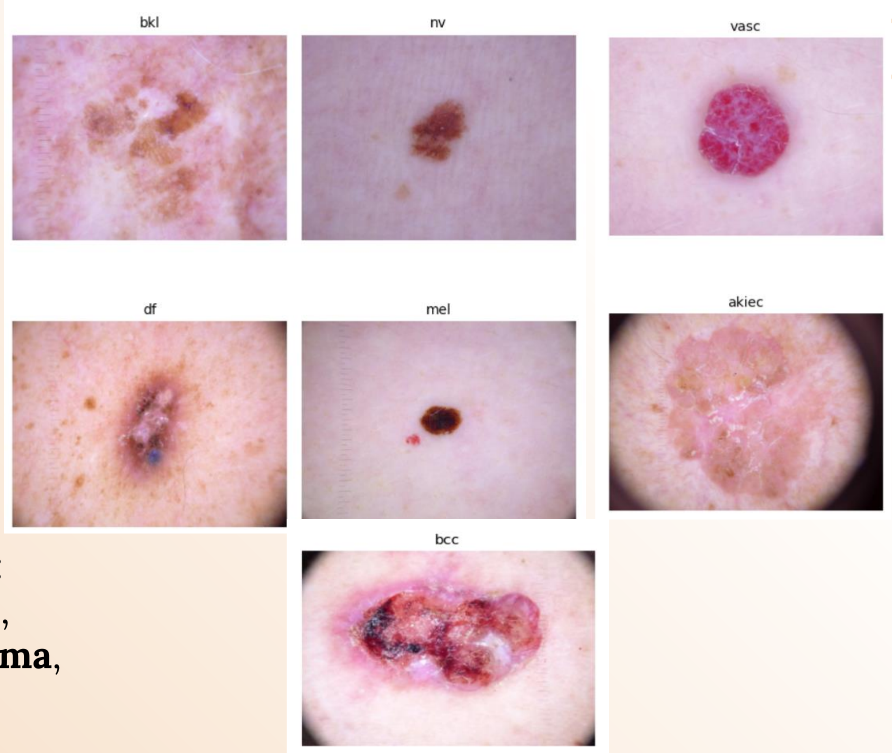
1.1

Data Understanding

- **HAM10000** (Human Against Machine with 10,000 training images) is a widely used dataset for **skin lesion analysis** in medical imaging.

- **10,015 dermatoscopic images (650x450 resolution)** collected from **diverse geographic locations** and acquired using **multiple imaging modalities**.

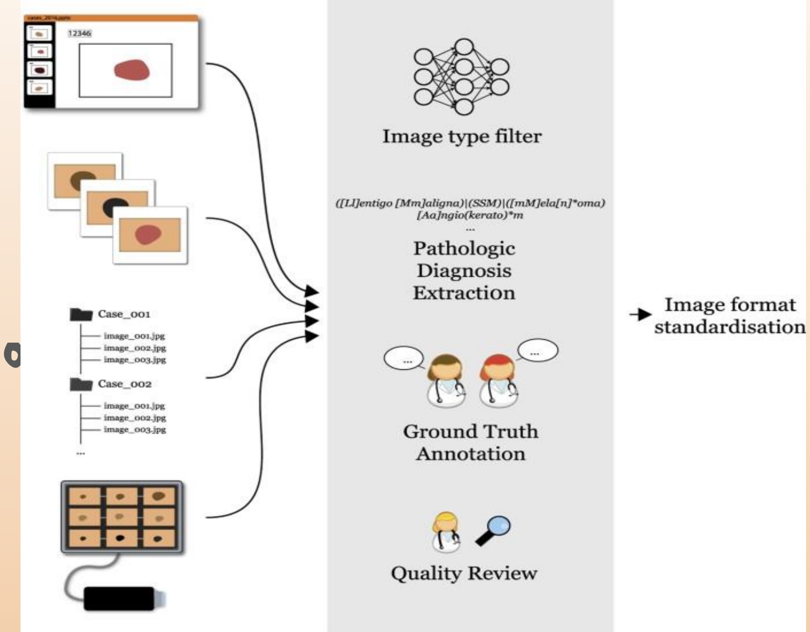
- **7 classes of lesions**, both benign and malignant:
  Melanocytic Nevi, **Melanoma**, Dermatofibroma,
  Actinic Keratosis and Intraepithelials carcinoma,
  Benign Keratosis-like carcinoma,
  Basal cell carcinoma, Vascular lesion



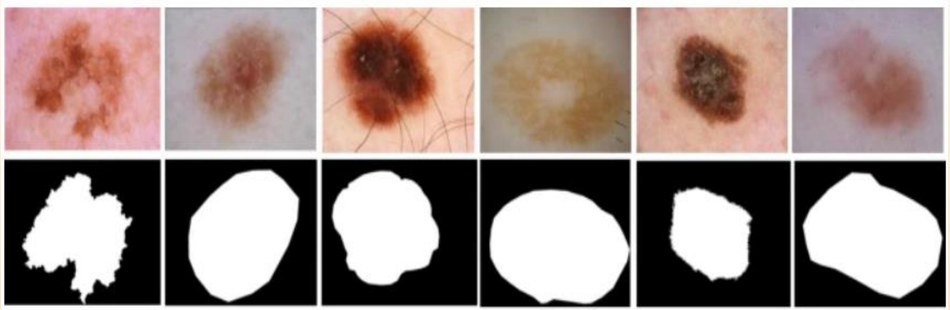
1.1

Data Understanding

Each image is accompanied by **rich metadata**, including: **Patient age, Anatomical site** of the lesion, **Type of ground truth** (e.g., histopathology, expert consensus)



The source also provides **segmentation masks**, which can be used for **supervised lesion segmentation tasks**.



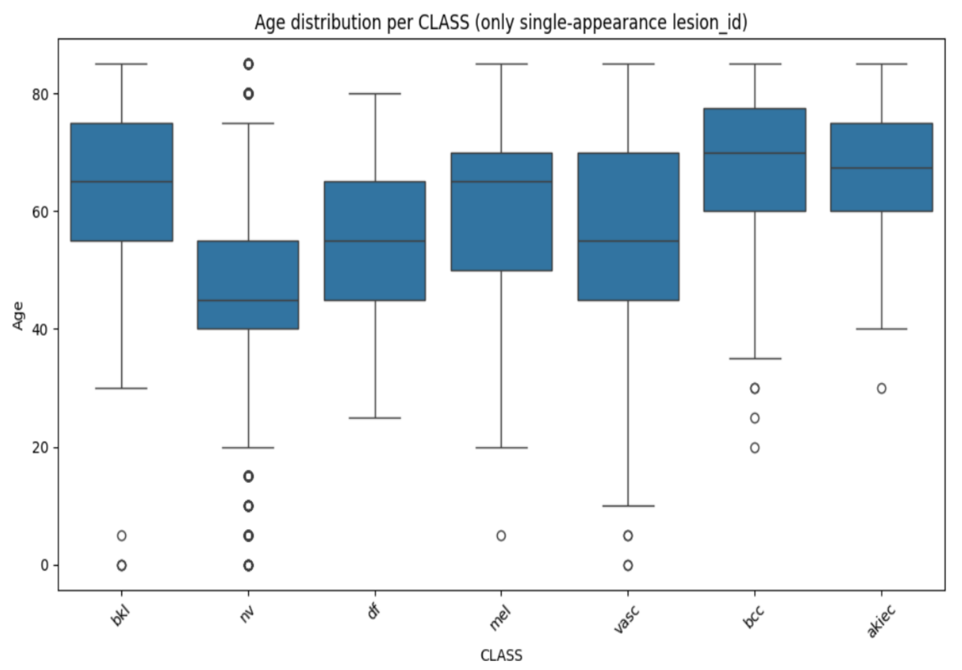
A **natural data augmentation** is already present: for some cases, a **zoomed-in version of the lesion** is included alongside the original image.

1.2

Data Cleaning

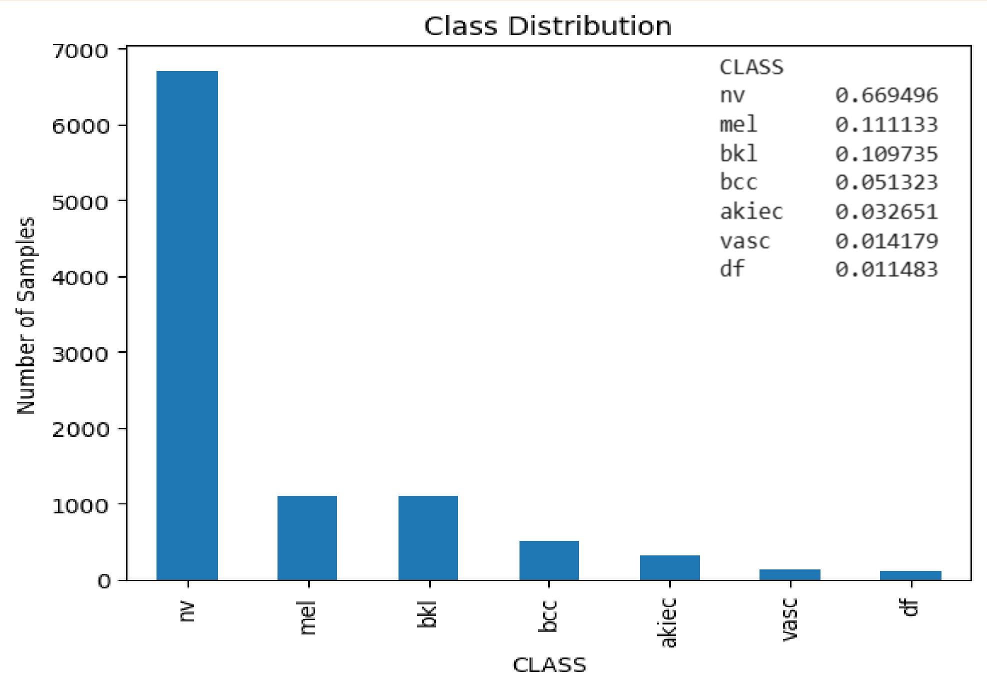
- Missing values : **57** for the column **«age»** for the classes **'nv'** , **'bkl'**, **'mel'**

- Since age was not crucial for classification, missing values were simply **imputed with the mean**, with no significant impact on performance.



2 Exploratory Data Analysis

- Univariate analysis : Class distribution



- Unbalanced dataset:

Melanocytic Nevi has **6705** samples;

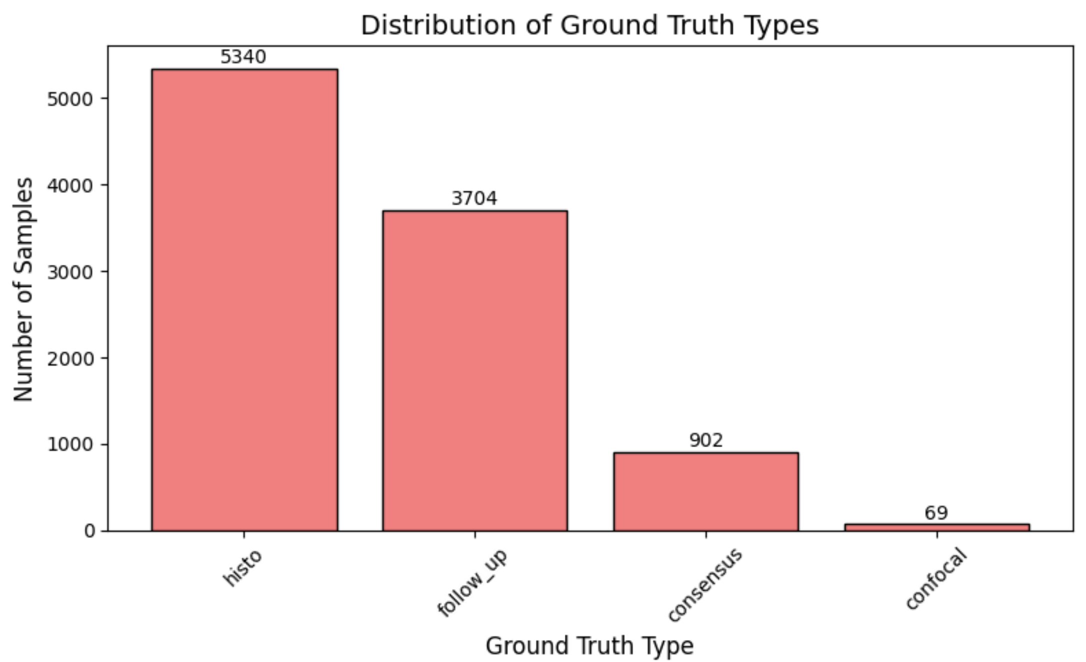
Dermatofibroma has only **115** samples;

Melanoma, despite being the most critical skin lesion, is **significantly underrepresented**, with only **~1,000 images**.

2

Exploratory Data Analysis

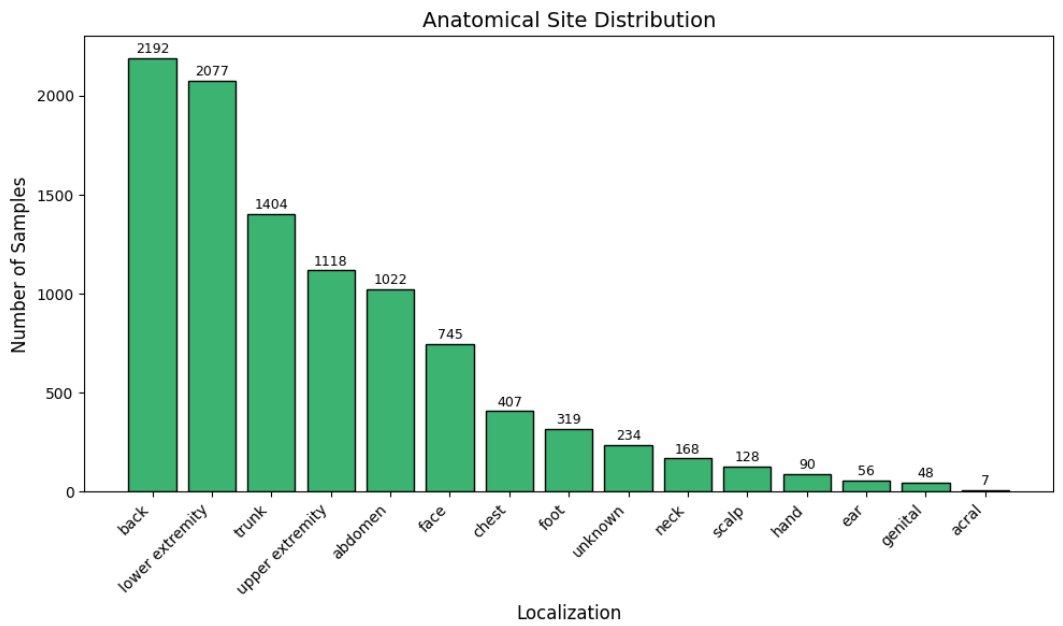
- Univariate analysis : Ground truth type distribution



- **Histopathology**: specialized dermatopathologists performed histopathologic diagnoses
- **Confocal Microscopy:** an in-vivo imaging technique at near-cellular resolution, used to verify benign facial keratoses
- **Follow-up:** Stability across 1.5 years was accepted as evidence of benignity, assessed by dermatologists
- **Expert consensus:** Diagnosis assigned via independent consensus by two experts, only if both agreed unequivocally(no follow-up info)

2 Exploratory Data Analysis

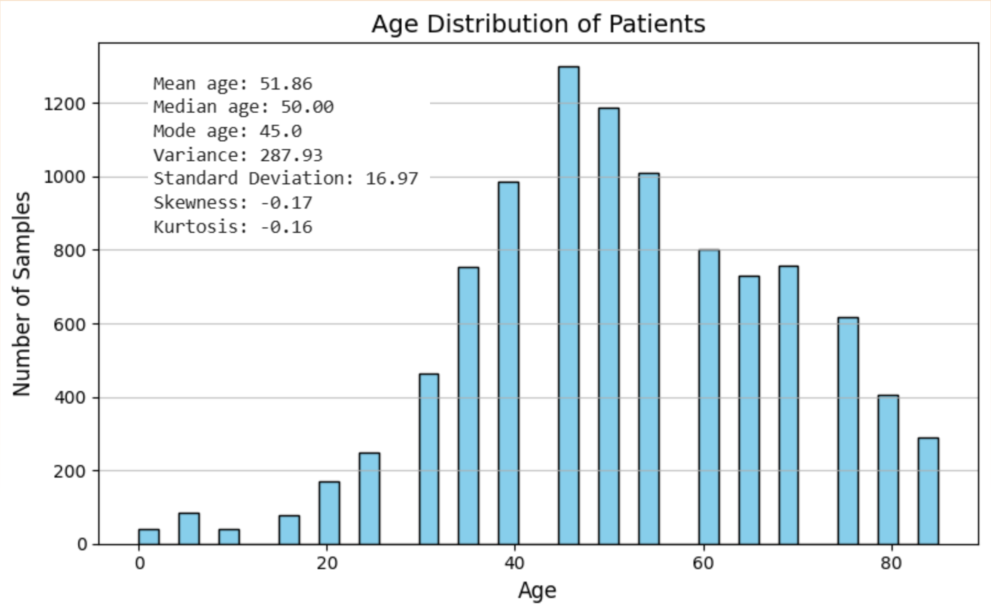
- Univariate analysis : Anatomical site distribution

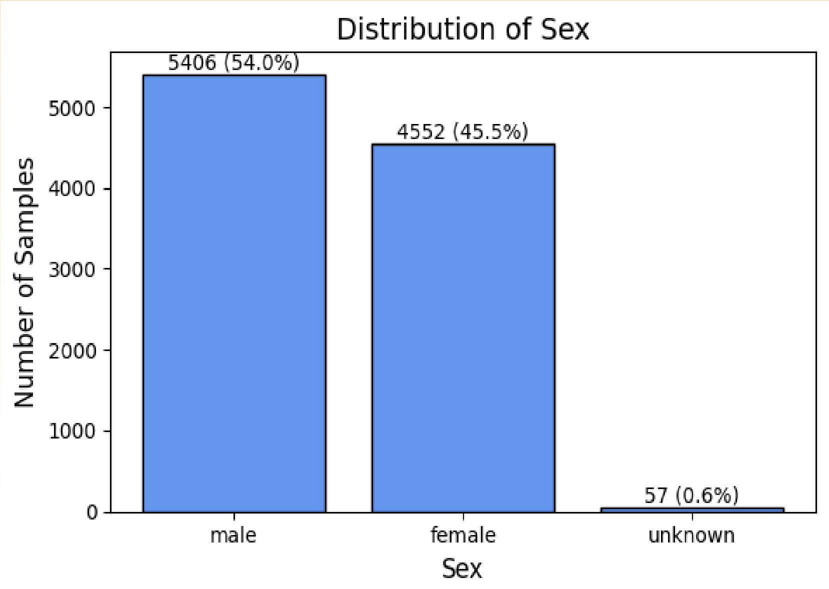


- Skin lesions can be influenced by **genetic factors** and **sun exposure**, which may cause mutations. Lesions are more commonly found on **sun-exposed areas** of the body.

2 Exploratory Data Analysis

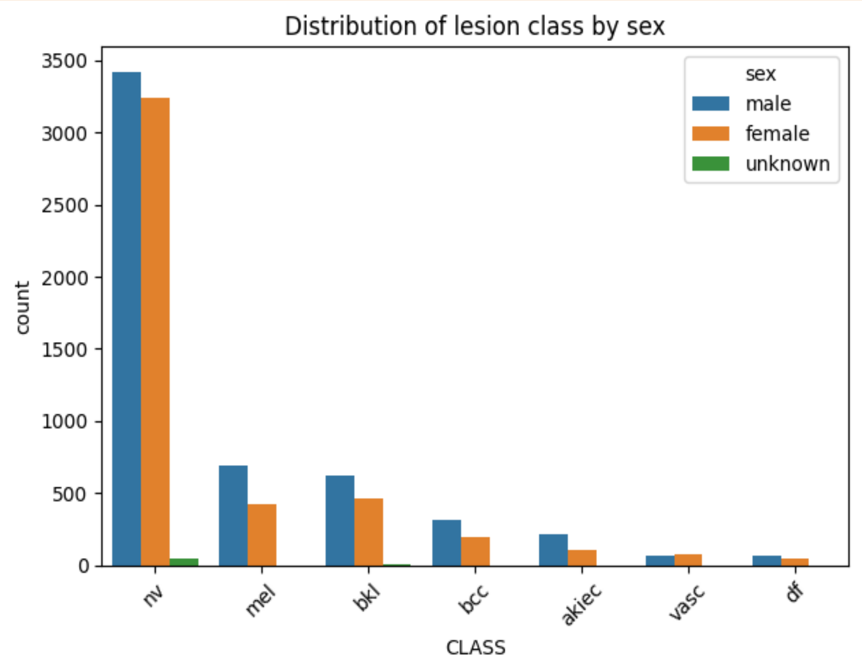
- Univariate analysis : Age and Gender distribution





2 Exploratory Data Analysis

- Bivariate analysis : Lesion class by sex (categorical)



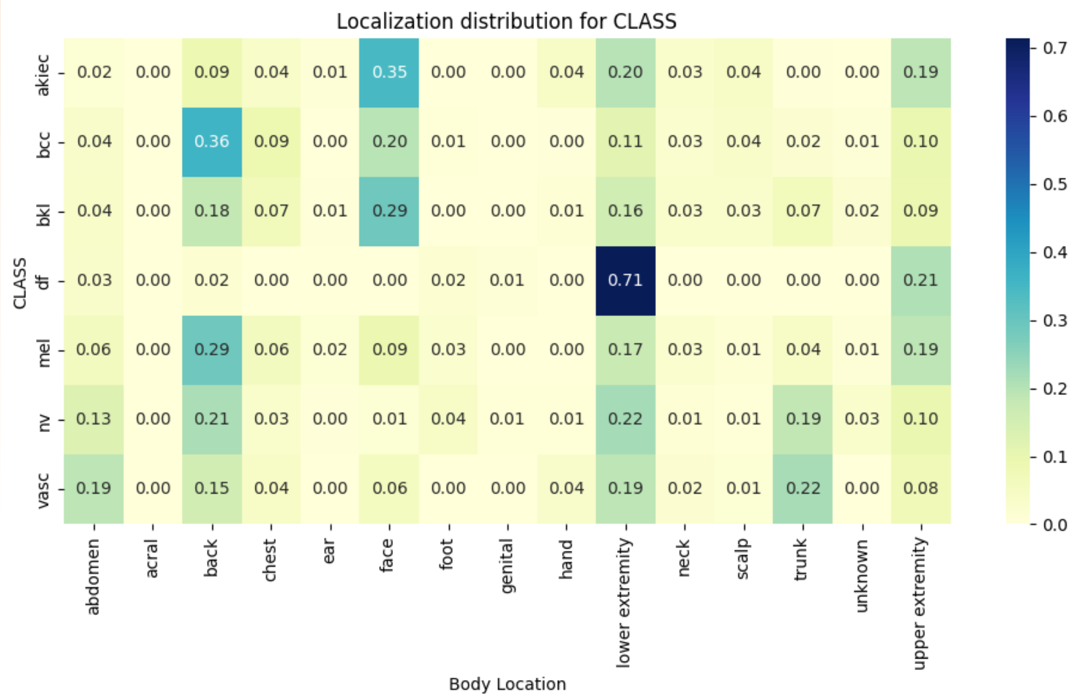
The Chi-square test has been applied to assess whether there is a **statistically significant association** between the lesion class and the patient's sex.

```
Chi2 statistic: 106.90381271641068
p-value: 2.4464388098587195e-17
```

Visual inspection of class distributions by sex shows very similar relative proportions. Despite statistical significance(due to the large dataset), the practical difference is minimal.

2 Exploratory Data Analysis

- Bivariate analysis : Localization distribution for Class



- Melanoma, Basal Cell Carcinoma (BCC), and Benign Keratosis-Like Lesions (BKL) mainly occur on the back.
- Dermatofibroma is predominantly found on the lower extremities, accounting for 71% of cases.
- Portions of AKIEC lesions (35%), as well as BCC and BKL, are located on the face

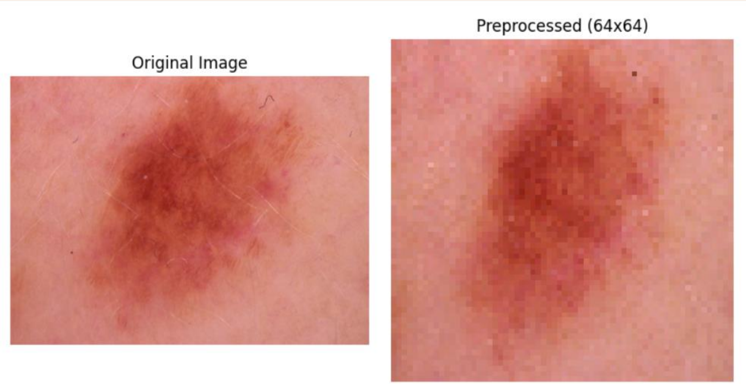
Chi-Square test results:

```
Chi2 statistic: 2821.9101978213816
p-value: 0.0
Degrees of freedom: 84
```

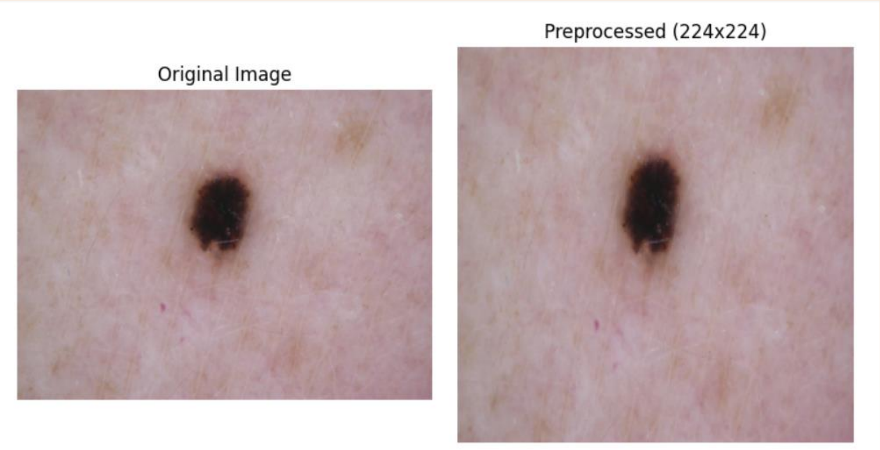
3

Image Pre-processing

- Original images sized 650x450 pixels need resizing
- Pixel values normalized from [0, 255] to [0, 1]



64x64 resolution is too low and causes loss of detail



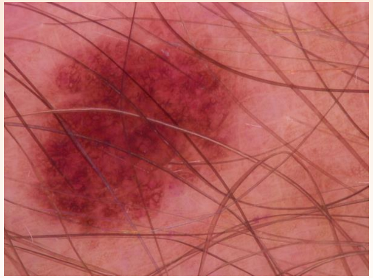
224x224 provides a good balance between computational cost and image quality improvement.

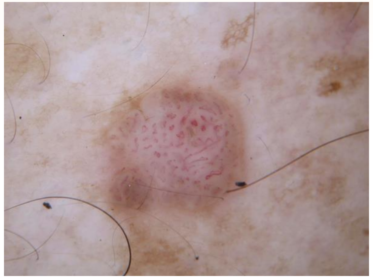
3 Image Pre-processing

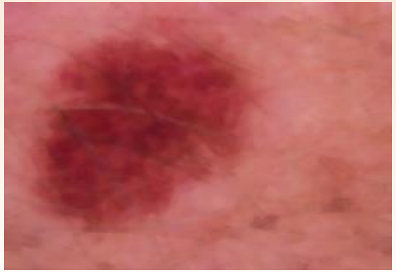
To ensure good quality skin images and enhance lesion visibility, additional preprocessing steps may be necessary, such as hair removal, denoising, or lighting enhancement.

Our approach: **Hair removal**, **Sharpening** and **CLAHE (Contrast Limited Adaptive Histogram Equalization)**

Original image







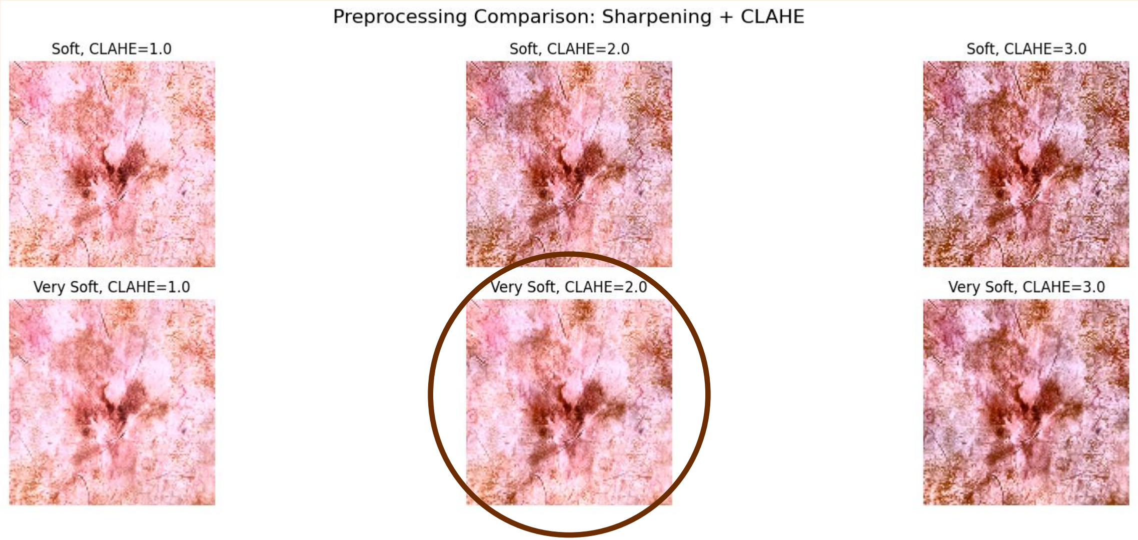
Without hair

3

Image Pre-processing

To ensure good quality skin images and enhance lesion visibility, additional preprocessing steps may be necessary, such as hair removal, denoising, or lighting enhancement.

Our approach: **Hair removal(Dull-Razor algorithm)**, **Sharpening** and **CLAHE (Contrast Limited Adaptive Histogram Equalization)**



3 Image Pre-processing

To ensure good quality skin images and enhance lesion visibility, additional preprocessing steps may be necessary, such as hair removal, denoising, or lighting enhancement.

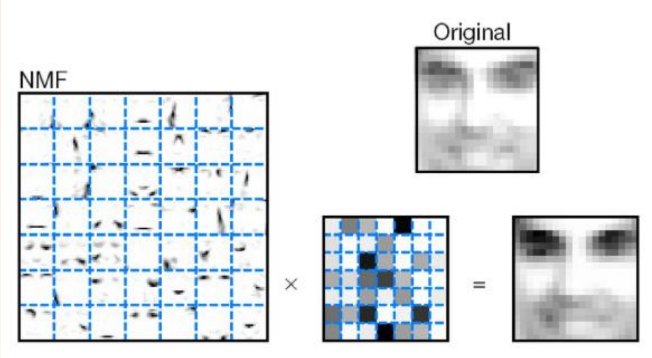
Our approach: **Hair removal**, **Sharpening** and **CLAHE (Contrast Limited Adaptive Histogram Equalization)**

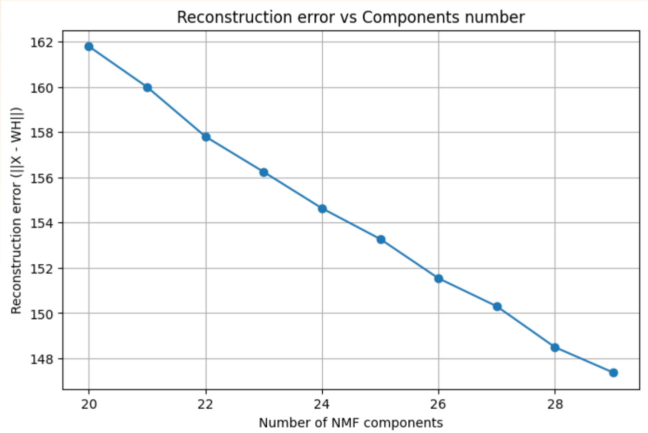
Final pre-processed images

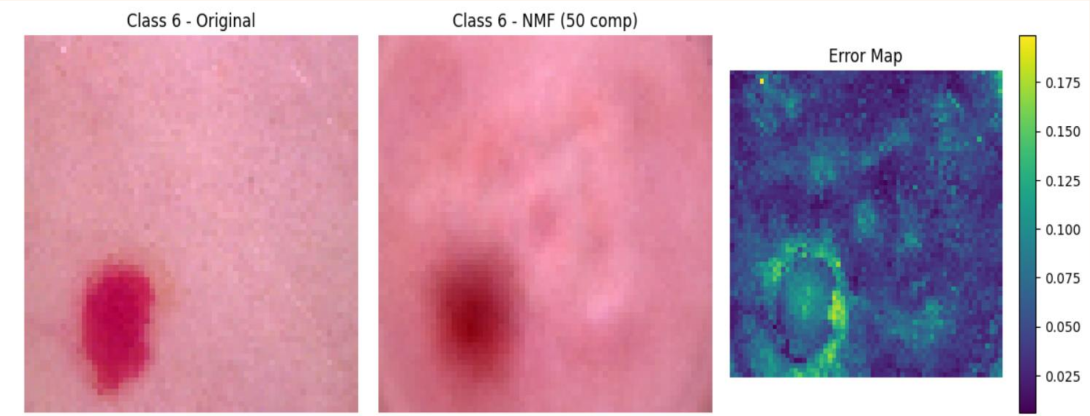
3.5 Feature extraction and dimensionality reduction

Applying NMF to decompose images into **parts-based**, interpretable features by factorizing the data into non-negative basis and activation matrices.

- Exploring the efficiency of the technique on a subset of 100 samples for each class: **64x64** case



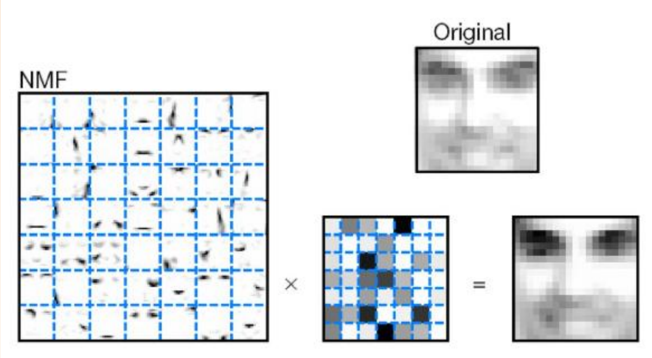




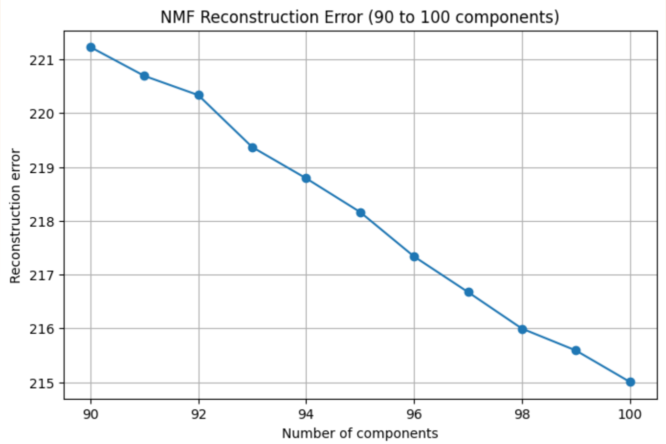
The reconstruction error remains high, the number of components is still suboptimal

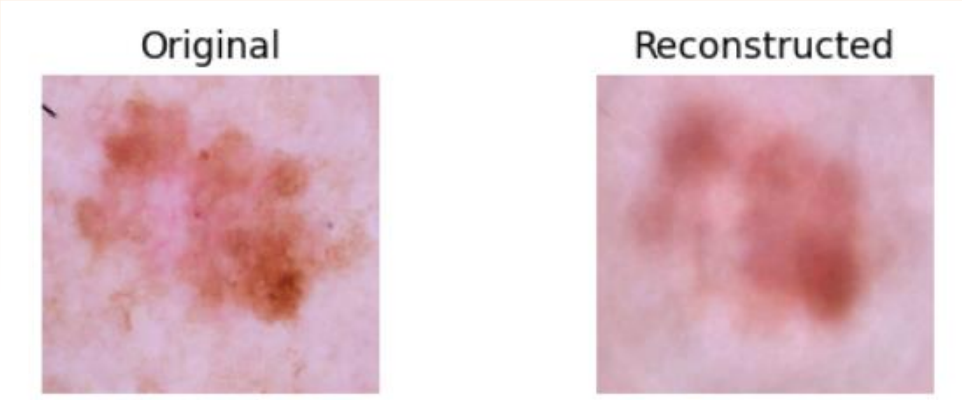
3.5 Feature extraction and dimensionality reduction

Applying NMF to decompose images into **parts-based**, interpretable features by factorizing the data into non-negative basis and activation matrices.



- Exploring the efficiency of the technique on a subset of 100 samples for each class: **128x128** case

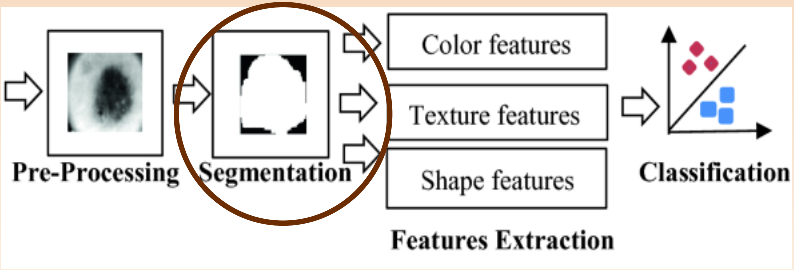




The reconstruction error remains high, the number of components is still suboptimal

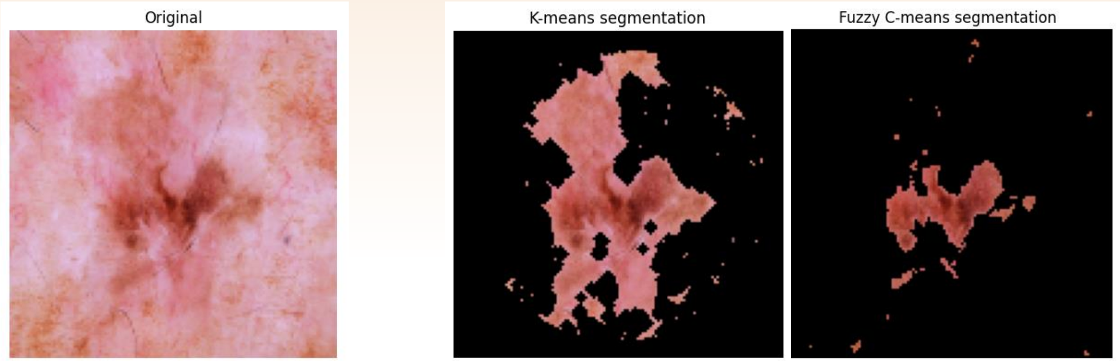
4

Image Segmentation



•Segmenting lesions helps to:
Focus the analysis on the **region of interest (ROI)** and Remove **irrelevant background** (basing on pixel intensity and texture)

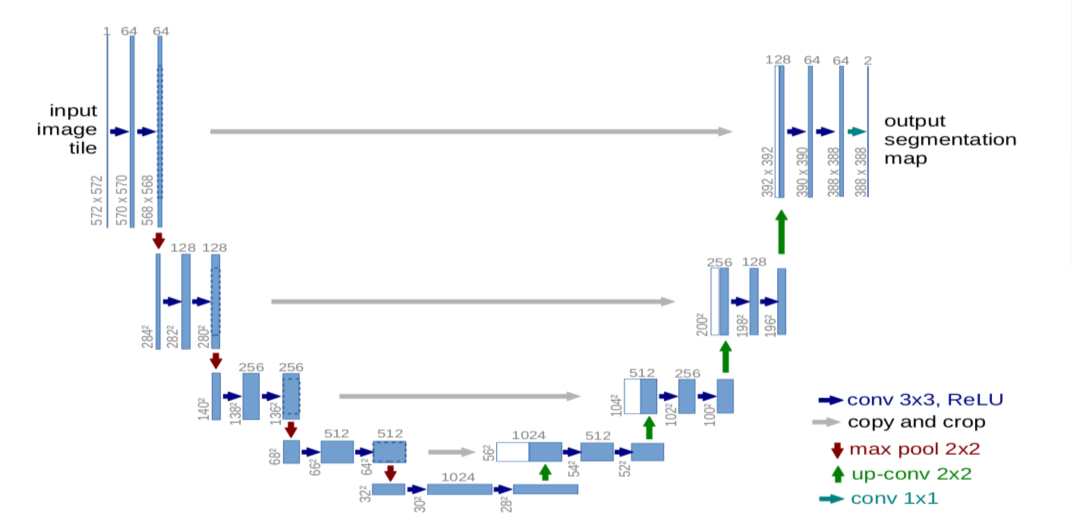
Clustering –based segmentatation techniques like **K-Means** (hard boundaries) **or Fuzzy C-Means**(better for blurry boundaries) may be implemented



4

Image Segmentation(U-Net CNN)

U-Net is a convolutional neural network designed for biomedical segmentation. It consists of a **contracting path** (encoder) to capture context and an **expanding path** (decoder) for precise localization. Well-suited for pixel-wise segmentation of lesions thanks to its **symmetric architecture** and **skip connections**



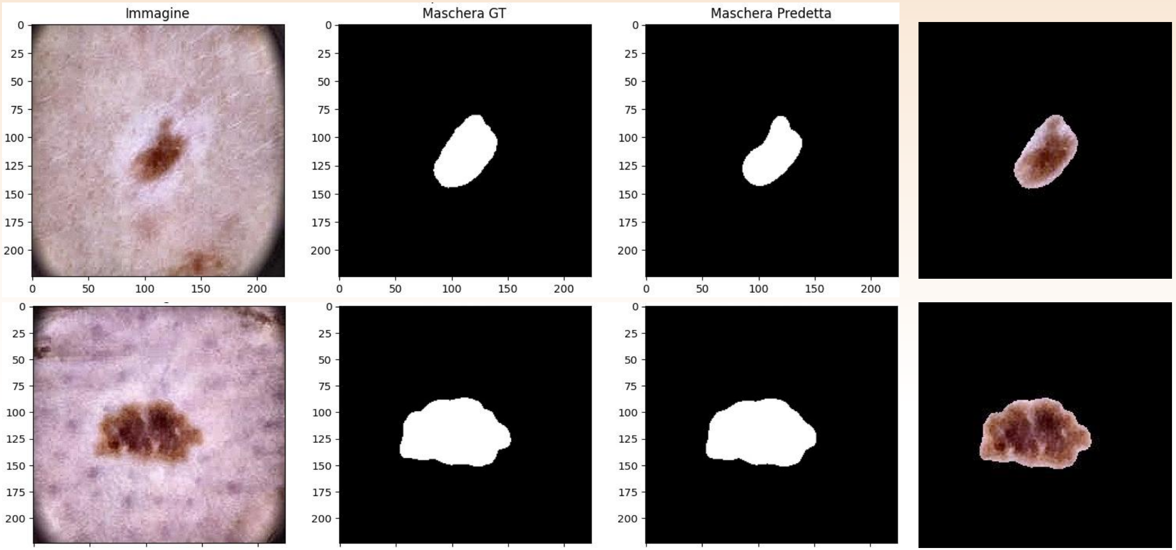
- U-Net **pre-trained** on **ImageNet** dataset
- **Encoder** structure of **ResNet-34**
- Fine-tuned after the first 5 epochs

The architecture in image is purely illustrative, meant to give a general idea of the model, and does not represent the actual architecture used

4

Image Segmentation(U-Net CNN)

- Applying resizing on the Segmentation maps
- Splitting the images **into train(80%)** ,**val(20% of train)** and **test(20%)** set
- Augmentatation on the train set: flipping, shifting, rotations to better generalize
- Normalizing on ImageNet statistics



- **Loss** function : a combined loss of **BCE** (focusing on per-pixel accuracy) and **Dice**(robust to evaluate mask overlap)
- Results on test set :

```
Dice medio: 0.9372
IoU medio: 0.8898
Sensitivity media: 0.9500
Specificity media: 0.9729
```

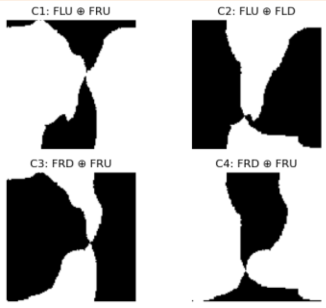
4.5

Feature extraction

The **ABCD method** is a clinical guideline to evaluate moles for malignancy risk:

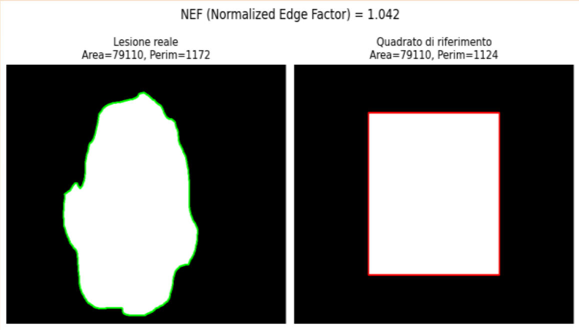
•**Asymmetry**: Irregular shape or structure when divided in quadrants:

$$dif_i = \frac{1}{N} \sum_{x,y} C_i(x,y)$$



- **Border**: Jagged or poorly defined edges

$$\text{NEF} = \frac{P_{\text{mole}}}{4\sqrt{n}}$$



- **Color**: Multiple tones or uneven pigmentation: Converting from RGB to HSV and extracting a weighted mean of the Hue channel making the image more robust to light variations

- **Diameter**: large lesions may pose greater risk (typically >6 mm). **Not used**

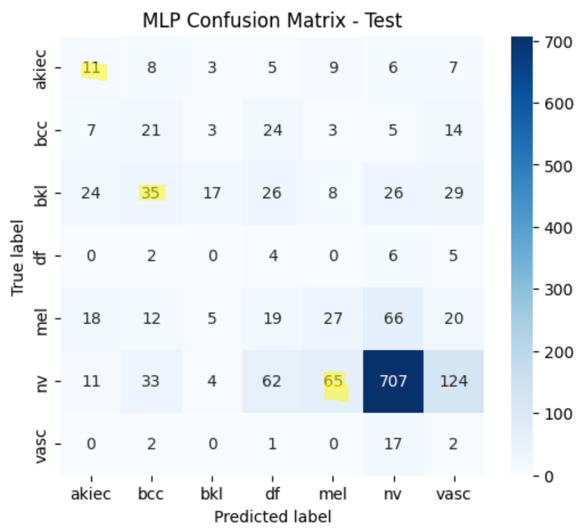
Classification models

5

SVM and MLP based classification using ABC features

- **MLP(1 hidden layer)**
- Used CrossEntropy loss, Adam optimizer
- Grid-Search over learning rate(0.001), N. of hidden neurons(100), activation function(ReLU)
- Test results:

	precision	recall	f1-score	support
0	0.15	0.22	0.18	49
1	0.19	0.27	0.22	77
2	0.53	0.10	0.17	165
3	0.03	0.24	0.05	17
4	0.24	0.16	0.19	167
5	0.85	0.70	0.77	1006
6	0.01	0.09	0.02	22



Classification models

5

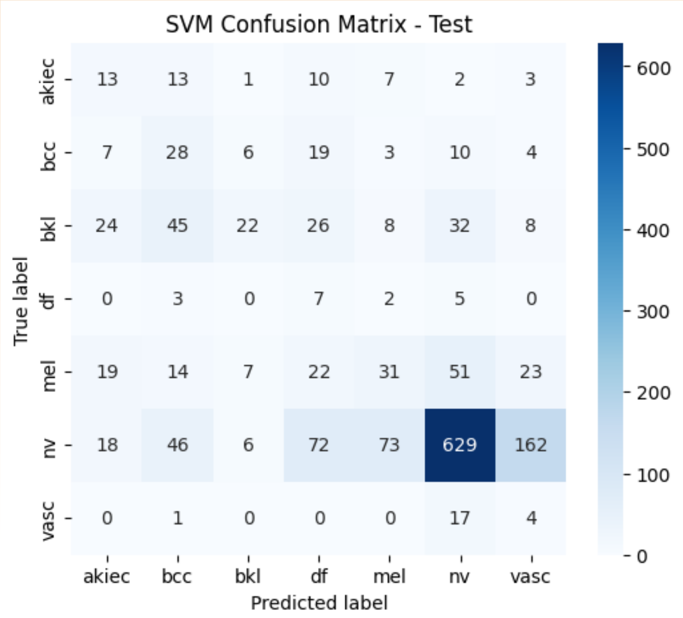
SVM and MLP based classification using ABC features

- **SVM**

- Tuning hyperparameters: **C (0.1) , Kernel(radial fixed), γ(scaled to data)** with a small Grid-Search

SVM Test Report:

	precision	recall	f1-score	support
0	0.16	0.27	0.20	49
1	0.19	0.36	0.25	77
2	0.52	0.13	0.21	165
3	0.04	0.41	0.08	17
4	0.25	0.19	0.21	167
5	0.84	0.63	0.72	1006
6	0.02	0.18	0.04	22
accuracy			0.49	1503



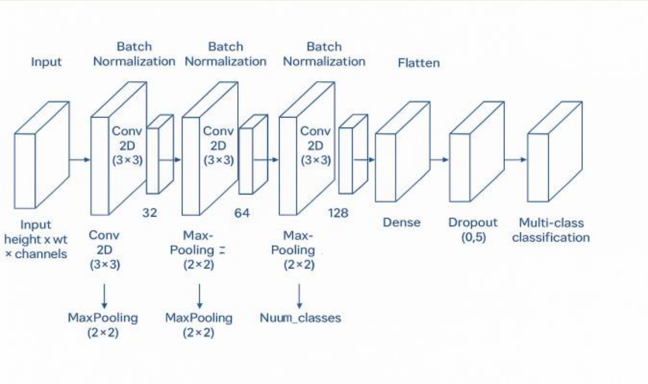
- Moderate overall accuracy(49%), but underperforming on malignant classes

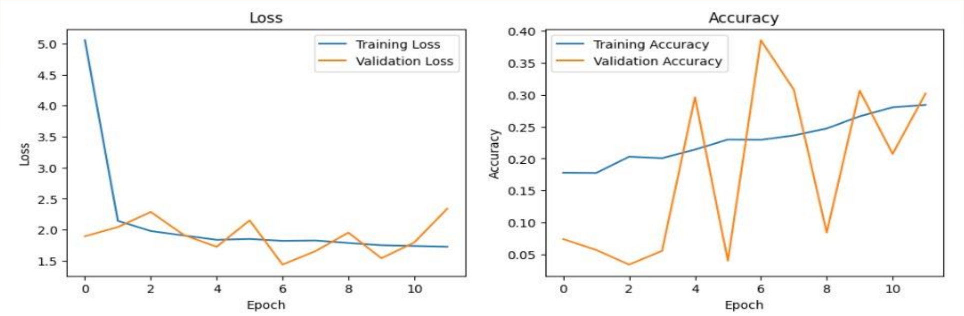
5

Classification models

Model 1: Training a small CNN from scratch on segmented images

- Splitting train e validation set **stratified(80/20 %)**
- **Hybrid sampling:** combines oversampling below-mean classes and undersampling above-mean classes
- Augmentation zooming, shifting and rotating
- Early stopping with patience 5





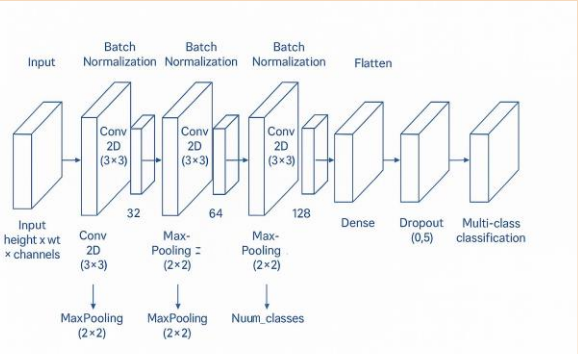
- **Validation loss** remains high and volatile, showing **no clear downward trend**

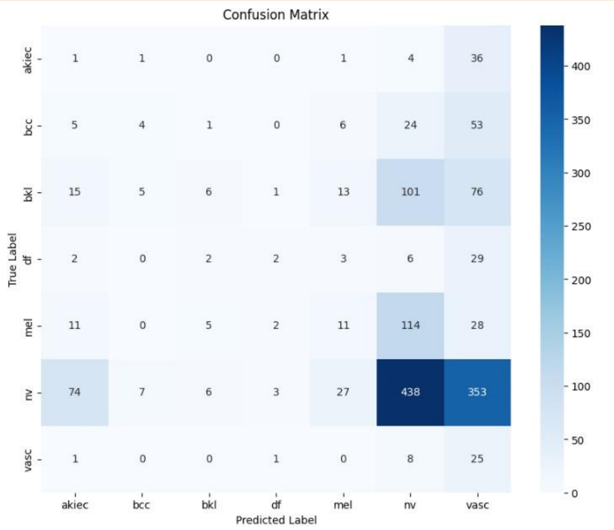
5

Classification models

Model 1: Training a small CNN from scratch on segmented images

- Test set: external **1511 images from ISIC dataset**





Classification report:

	precision	recall	f1-score	support
akiec	0.01	0.02	0.01	43
bcc	0.24	0.04	0.07	93
bkl	0.30	0.03	0.05	217
df	0.22	0.05	0.08	44
mel	0.18	0.06	0.09	171
nv	0.63	0.48	0.55	908
vasc	0.04	0.71	0.08	35
accuracy			0.32	1511
macro avg	0.23	0.20	0.13	1511
weighted avg	0.46	0.32	0.36	1511

- Melanoma and Basal-cell carcinoma have a very low recall
- Melanocytic Nevi is the most well predicted although recall is below 50%
- General accuracy 32% show that the model don't perform well as expected

5

Classification models

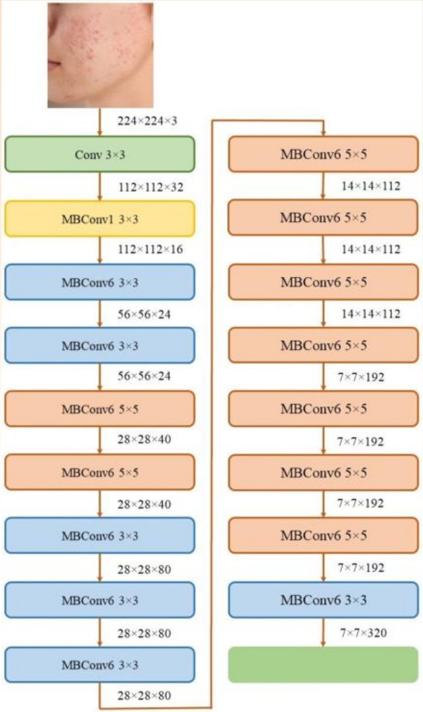
Implementing **EfficientNet-B0 (pre-trained** on **ImageNet)** as backbone on **original images without segmentation** to have a baseline model

- Splitting train e validation set **stratified**
- **Hybrid sampling:** combines oversampling below-mean classes and undersampling above-mean classes
- Augmentation zooming, shifting and rotating

```
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 224, 224, 3)]     0

efficientnetb0 (Functional   (None, 7, 7, 1280)        4049571
)

global_average_pooling2d (   (None, 1280)              0
GlobalAveragePooling2D)

dropout (Dropout)            (None, 1280)              0

dense (Dense)                (None, 7)                 8967

=================================================================
Total params: 4058538 (15.48 MB)
Trainable params: 8967 (35.03 KB)
Non-trainable params: 4049571 (15.45 MB)
```

Adding an head to the model:
- GlobalAveragePooling
- Dropout (p = 0.2)
- Final Dense Layer

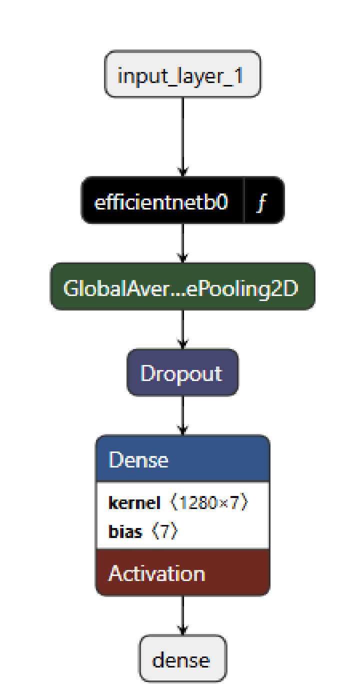
Optimizer: Adam
Loss function : CCE



5

Classification models

Implementing **EfficientNet-B0 (pre-trained** on **ImageNet)** as backbone on original images without segmentation to have a baseline model

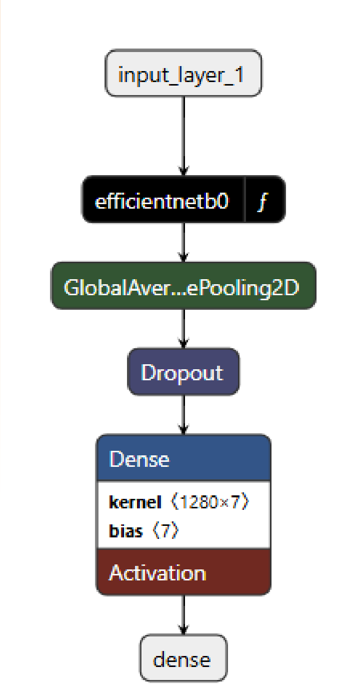


After 40 epochs training accuracy reaches a plateau at 77%, while validation loss and accuracy shows noise and accuracy is around 70%

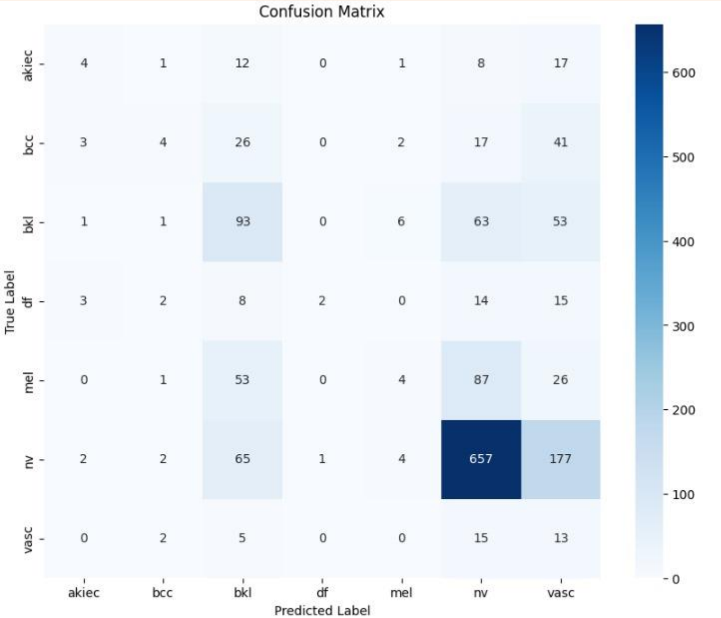
Classification models

5

Implementing **EfficientNet-B0 (pre-trained** on **ImageNet)** as backbone on original images without segmentation to have a baseline model



Test Results :



Classification report:

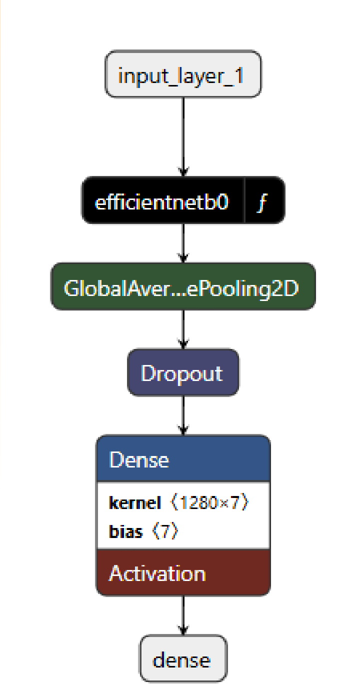
	precision	recall	f1-score	support
akiec	0.31	0.09	0.14	43
bcc	0.31	0.04	0.08	93
bkl	0.35	0.43	0.39	217
df	0.67	0.05	0.09	44
mel	0.24	0.02	0.04	171
nv	0.76	0.72	0.74	908
vasc	0.04	0.37	0.07	35
accuracy			0.51	1511
macro avg	0.38	0.25	0.22	1511
weighted avg	0.58	0.51	0.52	1511

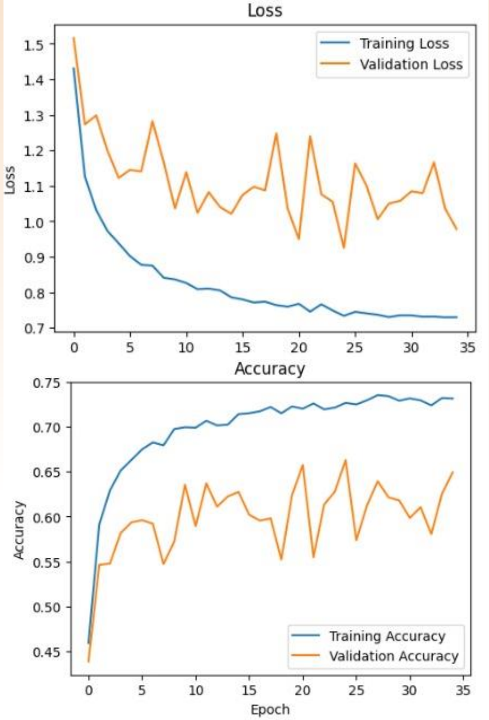
- Melanoma has 2% of recall, only 4 True Positives
- BKL has an higher recall (43%)

5

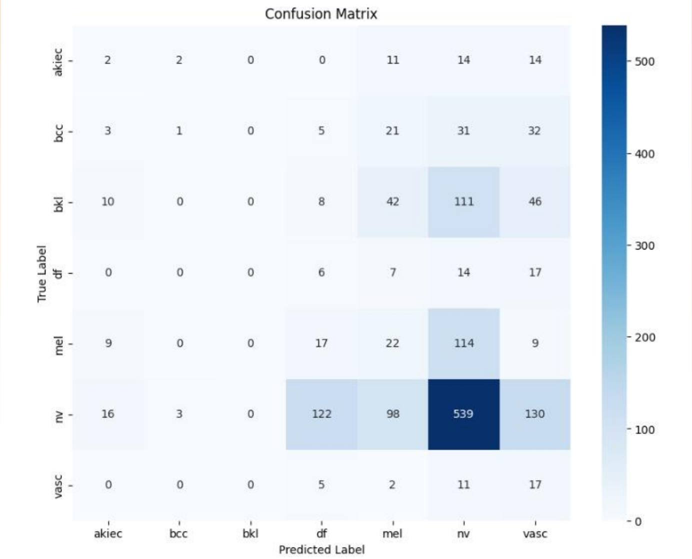
Classification models

Implementing **EfficientNet-B0 (pre-trained** on **ImageNet)** as backbone on **pre-processed** images **with segmentation**





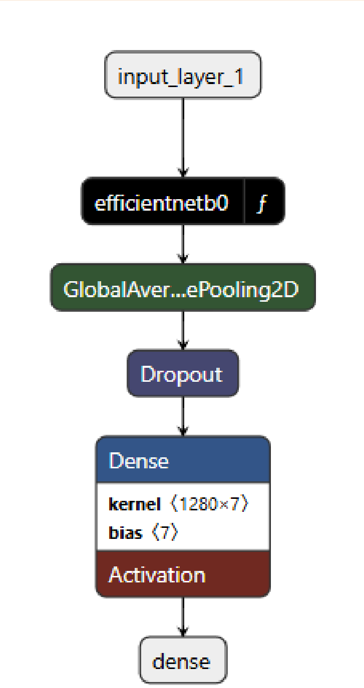
Test Results :



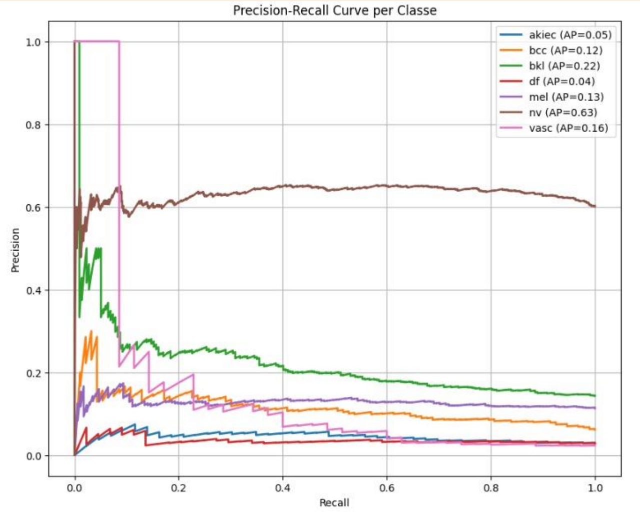
5

Classification models

Implementing **EfficientNet-B0 (pre-trained** on **ImageNet)** as backbone on **pre-processed** images **with segmentation**



Test Results :



Classification report:

	precision	recall	f1-score	support
akiec	0.05	0.05	0.05	43
bcc	0.17	0.01	0.02	93
bkl	0.00	0.00	0.00	217
df	0.04	0.14	0.06	44
mel	0.11	0.13	0.12	171
nv	0.65	0.59	0.62	908
vasc	0.06	0.49	0.11	35
accuracy			0.39	1511
macro avg	0.15	0.20	0.14	1511
weighted avg	0.41	0.39	0.39	1511

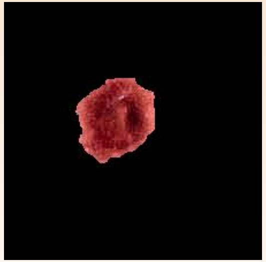
- Now Melanoma has a slightly higher recall but still very low

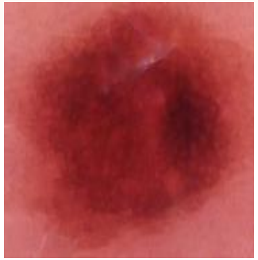
- Many images are misclassified as Melanocytic Nevi, this is a real problem

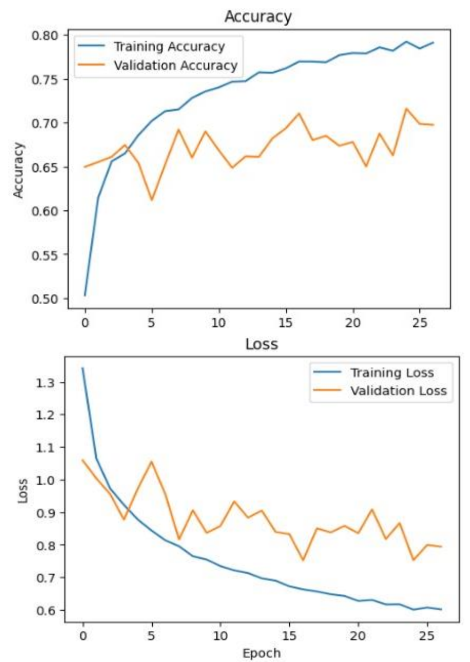
5

Classification models

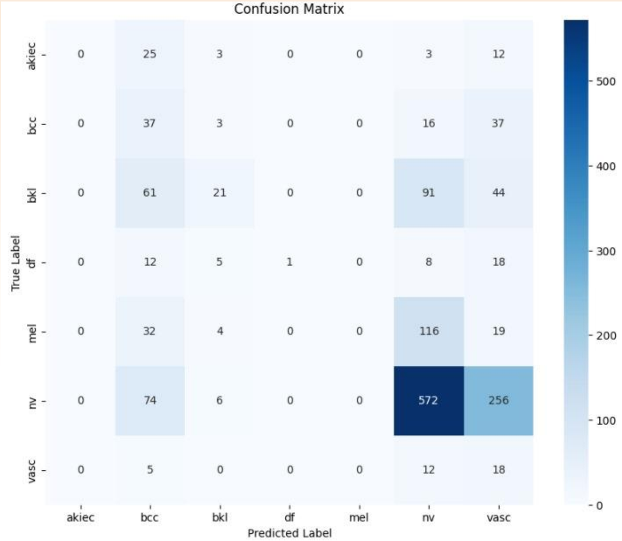
Implementing **EfficientNet-B0 (pre-trained** on **ImageNet)** as backbone on **pre-processed** images **with segmentation.** It has been cropped a rectangle as contour of the image.







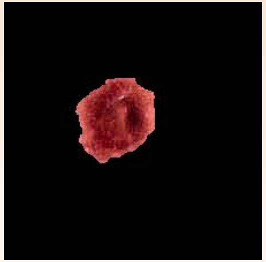
Test Results :

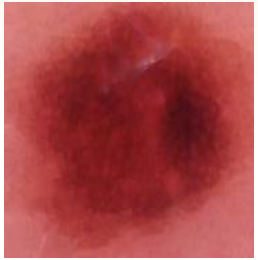


5

Classification models

Implementing **EfficientNet-B0 (pre-trained** on **ImageNet)** as backbone on **pre-processed** images **with segmentation.** It has been cropped a rectangle as contour of the image.

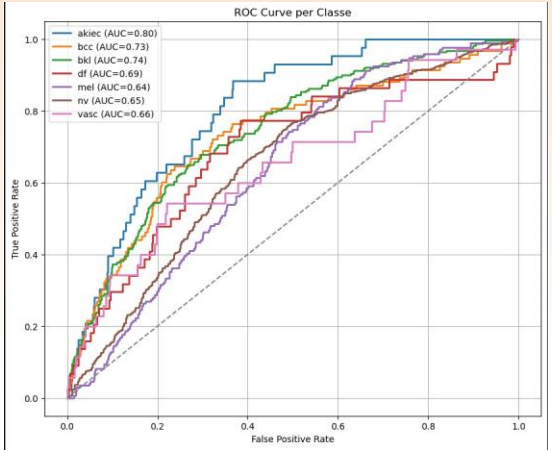




Test Results :

Classification report:

	precision	recall	f1-score	support
akiec	0.00	0.00	0.00	43
bcc	0.15	0.40	0.22	93
bkl	0.50	0.10	0.16	217
df	1.00	0.02	0.04	44
mel	0.00	0.00	0.00	171
nv	0.70	0.63	0.66	908
vasc	0.04	0.51	0.08	35
accuracy			0.43	1511
macro avg	0.34	0.24	0.17	1511
weighted avg	0.53	0.43	0.44	1511

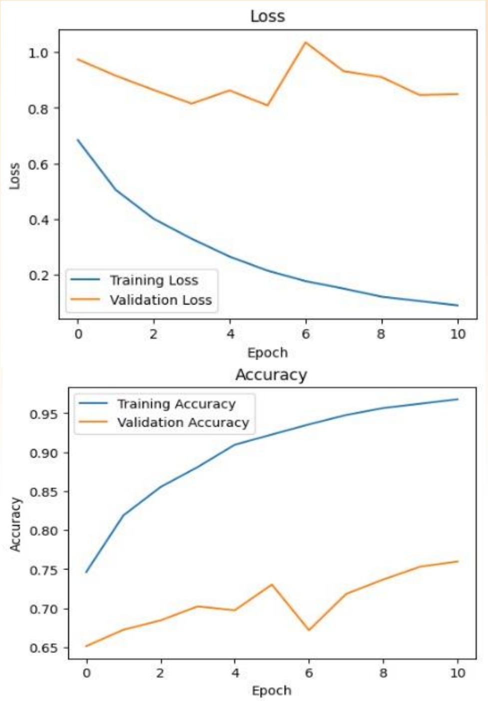


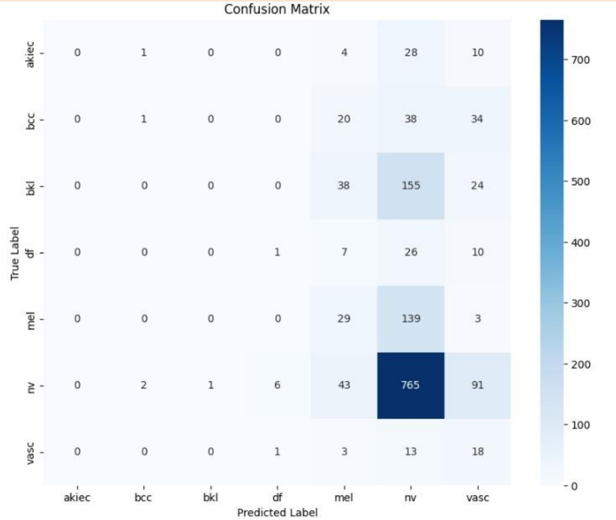
- Ignoring the predicted lesion borders, the model performs poorly on the test set, with no melanoma cases correctly predicted

5

Classification models

Implementing **EfficientNet-B0 (pre-trained** on **ImageNet)** as backbone on **pre-processed** images **with segmentation** and fine-tuning on our dataset sfreezing all the layers.





- Early-stopping with a small patience let stop the training

Test Results :

Classification report:

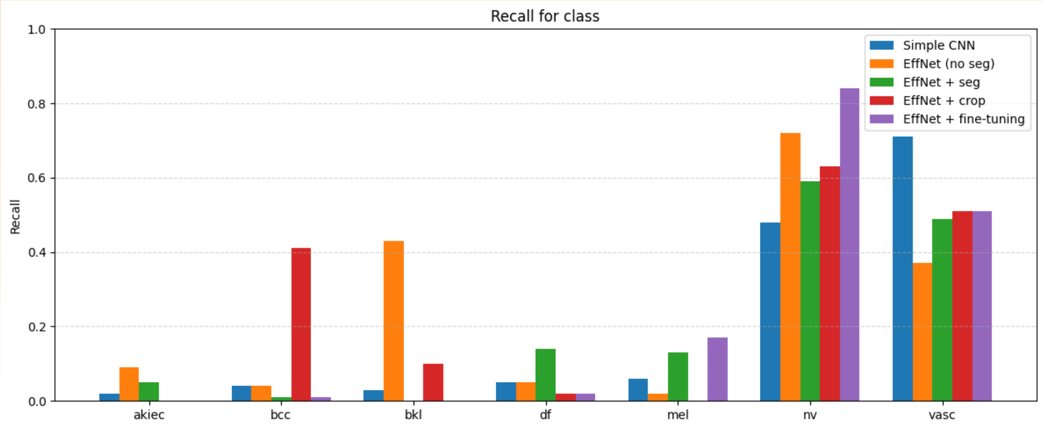
	precision	recall	f1-score	support
akiec	0.00	0.00	0.00	43
bcc	0.25	0.01	0.02	93
bkl	0.00	0.00	0.00	217
df	0.12	0.02	0.04	44
mel	0.20	0.17	0.18	171
nv	0.66	0.84	0.74	908
vasc	0.09	0.51	0.16	35
accuracy			0.54	1511
macro avg	0.19	0.22	0.16	1511
weighted avg	0.44	0.54	0.47	1511

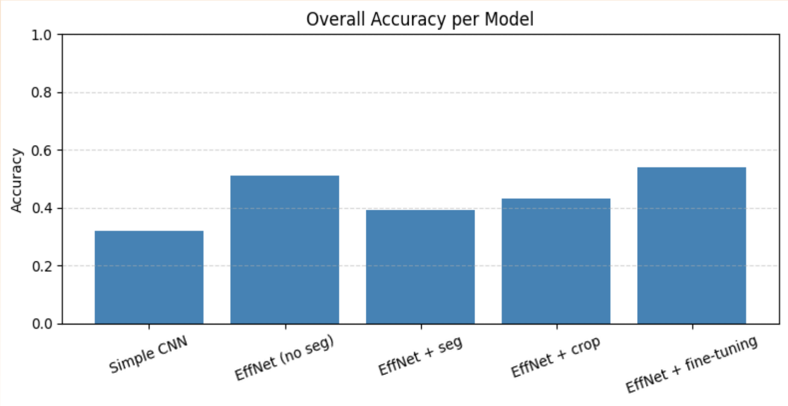
- After fine-tuning recall on Melanoma has increased to 17%
- Akiec and BKL are never predicted

4

Model Comparison

There have been purposed different model **with** and **without** applying **segmentation**. Here a comparison between them:





- Melanoma, Basal cell carcinoma and Akiec(pre-cancerous) have always **recall under 50%. Accuracy never reach 60%** neither fine-tuned model

- There is not a clear outperformance of models using segmented images. This dataset has dermatoscopic images well focused on the lesion

4 Conclusion and future work improvements

Overall, the models did not perform well on the test set, except for **melanocytic nevi**, which consistently showed better results. Malignant lesions like Melanoma and Basal Cell Carcinoma are generally not well classified. Segmentation on the dataset has not provided big improvements on performances.

One possible reason of the “ bad results” is the **high intra-class variability** of melanocytic nevi, which makes classification more complex and less reliable.

Improvements?

- Studying firstly the problem without considering Melanocytic Nevi or binarizing the scenario(Benign vs Malignant)

- Exploring the space of **hyperparameters**(with CV and GridSearch) and the **architecture** using **optimization algorithm** like Genetic Algorithm, Differential Evolution

- Improving the **domain knowledge** in order to better understand how to pre-process with more detail medical images