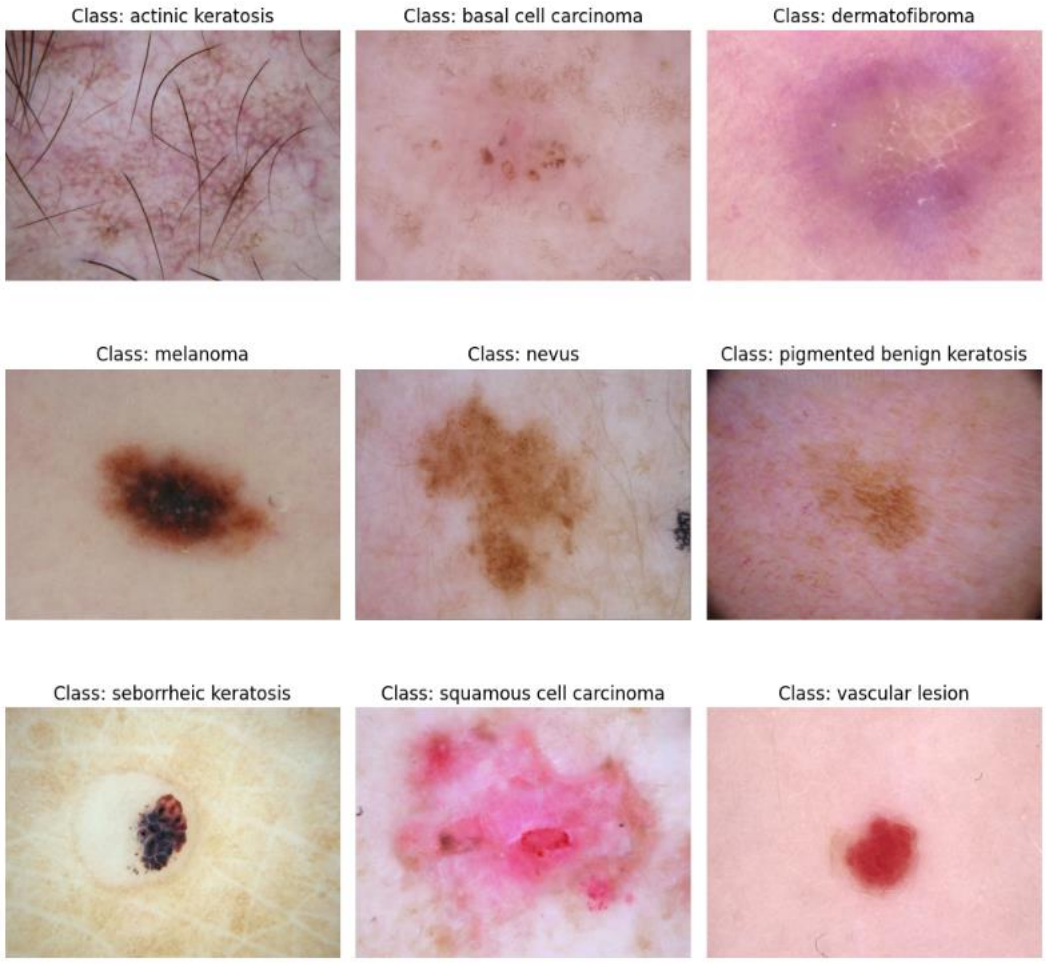
Melanoma Skin Cancer Detection

Problem statement:

To build a CNN based model which can accurately detect melanoma.

Melanoma is a type of cancer that can be deadly if not detected early. It accounts for 75% of skin cancer deaths. A solution which can evaluate images and alert the dermatologists about the presence of melanoma has the potential to reduce a lot of manual effort needed in diagnosis.

Below are the different Types of skin cancer with one example



Model 1 – Basic model

Layer (type)	Output Shape	Param #
rescaling (Rescaling)	(None, 180, 180, 3)	0
conv2d (Conv2D)	(None, 180, 180, 32)	896
conv2d_1 (Conv2D)	(None, 180, 180, 32)	9248
max_pooling2d (MaxPooling2D)	(None, 90, 90, 32)	0
conv2d_2 (Conv2D)	(None, 90, 90, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 45, 45, 64)	0
conv2d_3 (Conv2D)	(None, 45, 45, 128)	73856
max_pooling2d_2 (MaxPooling2D)	(None, 22, 22, 128)	0
flatten (Flatten)	(None, 61952)	0
dense (Dense)	(None, 256)	15859968
dense_1 (Dense)	(None, 9)	2313

Total params: 15964777 (60.90 MB)
Trainable params: 15964777 (60.90 MB)
Non-trainable params: 0 (0.00 Byte)

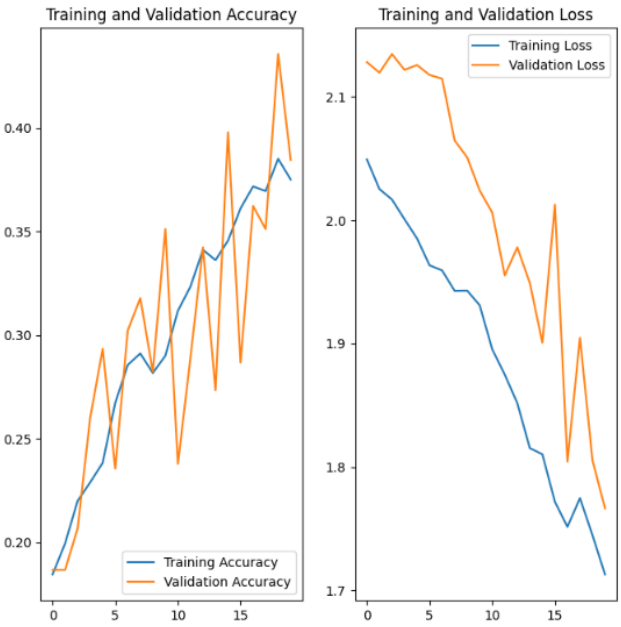
Observation :-

- It is observed that validation accuracy is fluctuating alot indicating that the model is highly overfitted

Model 2 – Data augumentation strategy and drop out layer to handle overfitting

Layer (type)	Output Shape	Param #
sequential_1 (Sequential)	(None, 180, 180, 3)	0
rescaling_4 (Rescaling)	(None, 180, 180, 3)	0
conv2d_18 (Conv2D)	(None, 180, 180, 32)	896
dropout_17 (Dropout)	(None, 180, 180, 32)	0
conv2d_19 (Conv2D)	(None, 180, 180, 32)	9248
max_pooling2d_14 (MaxPooling2D)	(None, 90, 90, 32)	0
dropout_18 (Dropout)	(None, 90, 90, 32)	0
conv2d_20 (Conv2D)	(None, 90, 90, 64)	18496
max_pooling2d_15 (MaxPooling2D)	(None, 45, 45, 64)	0
dropout_19 (Dropout)	(None, 45, 45, 64)	0
conv2d_21 (Conv2D)	(None, 45, 45, 128)	73856
max_pooling2d_16 (MaxPooling2D)	(None, 22, 22, 128)	0
dropout_20 (Dropout)	(None, 22, 22, 128)	0
conv2d_22 (Conv2D)	(None, 22, 22, 256)	295168
max_pooling2d_17 (MaxPooling2D)	(None, 11, 11, 256)	0
dropout_21 (Dropout)	(None, 11, 11, 256)	0
flatten_4 (Flatten)	(None, 30976)	0
dense_8 (Dense)	(None, 512)	15860224
dropout_22 (Dropout)	(None, 512)	0
dense_9 (Dense)	(None, 9)	4617

Total params: 16262505 (62.04 MB)
Trainable params: 16262505 (62.04 MB)
Non-trainable params: 0 (0.00 Byte)



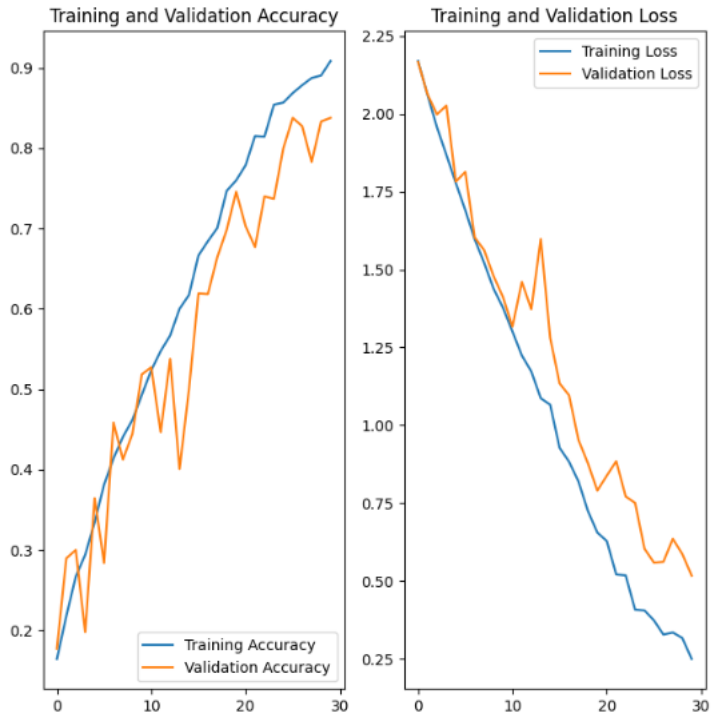
Observation :-

- Still there is some overfitting
- Accuracy need to be improved

Model 3 – Handle class imbalance

Layer (type)	Output Shape	Param #
rescaling_5 (Rescaling)	(None, 180, 180, 3)	0
conv2d_23 (Conv2D)	(None, 180, 180, 32)	896
dropout_23 (Dropout)	(None, 180, 180, 32)	0
conv2d_24 (Conv2D)	(None, 180, 180, 32)	9248
max_pooling2d_18 (MaxPooling2D)	(None, 90, 90, 32)	0
dropout_24 (Dropout)	(None, 90, 90, 32)	0
conv2d_25 (Conv2D)	(None, 90, 90, 64)	18496
max_pooling2d_19 (MaxPooling2D)	(None, 45, 45, 64)	0
dropout_25 (Dropout)	(None, 45, 45, 64)	0
conv2d_26 (Conv2D)	(None, 45, 45, 128)	73856
max_pooling2d_20 (MaxPooling2D)	(None, 22, 22, 128)	0
dropout_26 (Dropout)	(None, 22, 22, 128)	0
flatten_5 (Flatten)	(None, 61952)	0
dropout_27 (Dropout)	(None, 61952)	0
dense_10 (Dense)	(None, 256)	15859968
dense_11 (Dense)	(None, 9)	2313

Total params: 15964777 (60.90 MB)
Trainable params: 15964777 (60.90 MB)
Non-trainable params: 0 (0.00 Byte)



Observation –

- Overfitting has been reduced to great extend
- Accuracy of train data is around 90% and accuracy of validation data is 83%

Questions

1. Which class has the least number of samples?

   Seborrheic keratosis has the least number of samples – 77

2. Which classes dominate the data in terms proportionate number of samples?

   Pigmented benign keratosis and melanoma dominate the data in terms proportionate number of samples