Melanoma Detection Assignment

PREPARED BY

SANJEEV SURENDRAN

Assignment Objective

- Get dataset of 9 classes of skin cancer disease and prepare dataset for training, validation and testing.

Data Analysis

- Build CNN model and train it for 20 epochs. Observe the difference between adding data augmentation layer.

CNN Model

- Build and train CNN model for 30 epochs after rectifying class imbalance issue. Observe the performance and predict test data.

Business Solution

Business Understanding

To build a Convolutional Neural Network (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.

Task

Create, compile and train a CNN based model which can classify 9 types of skin cancer accurately.

Data Understanding

A zipped file containing 2 folders - Train and Test are given.

Train and Test folders are present for training-validation and predicting the model respectively.

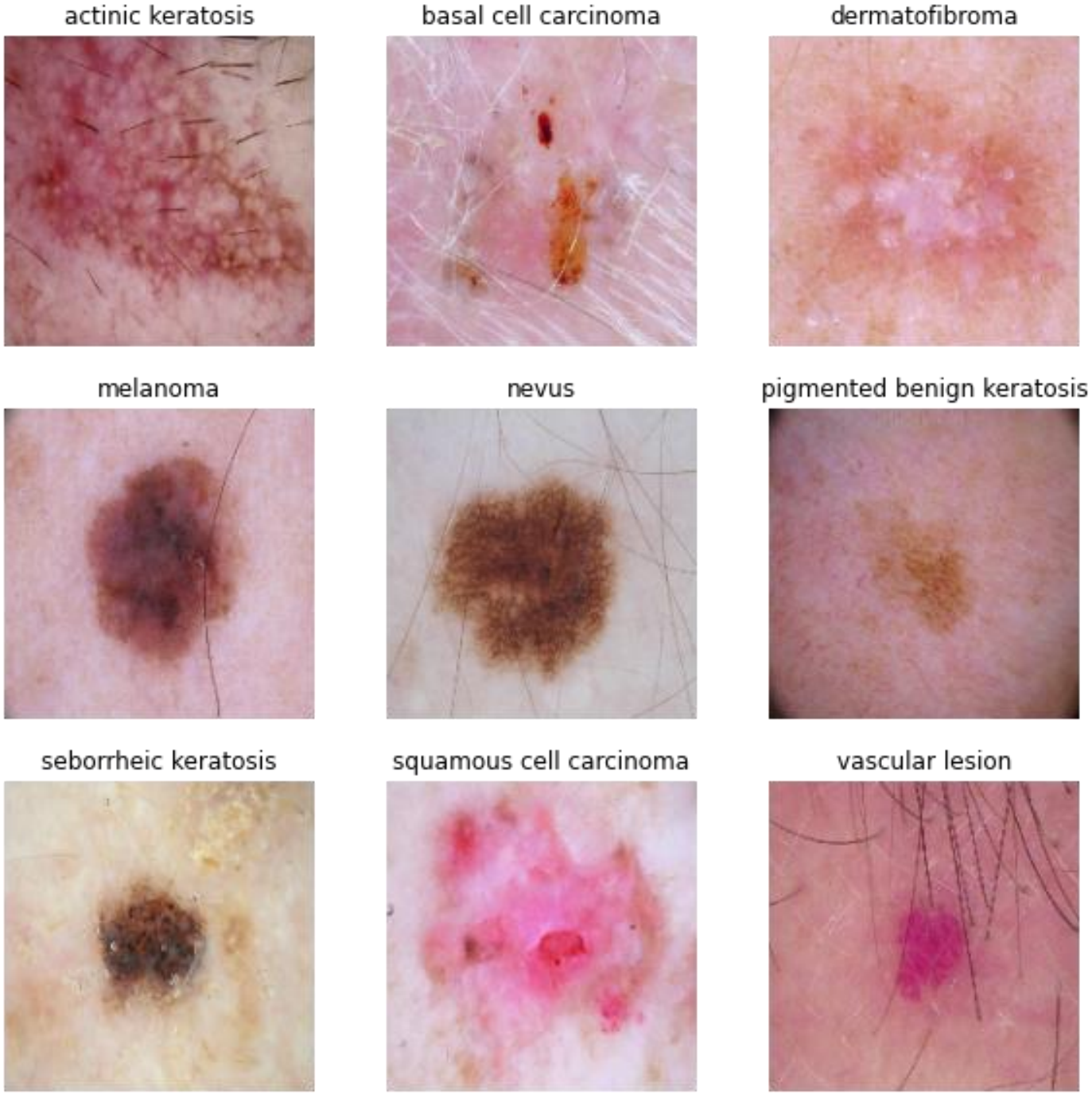
Each folder contains sub-folders with names of 9 types of skin cancer - actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma, vascular lesion.

These sub-folders contains respective skin cancer images.

Data Preparation and Visualization

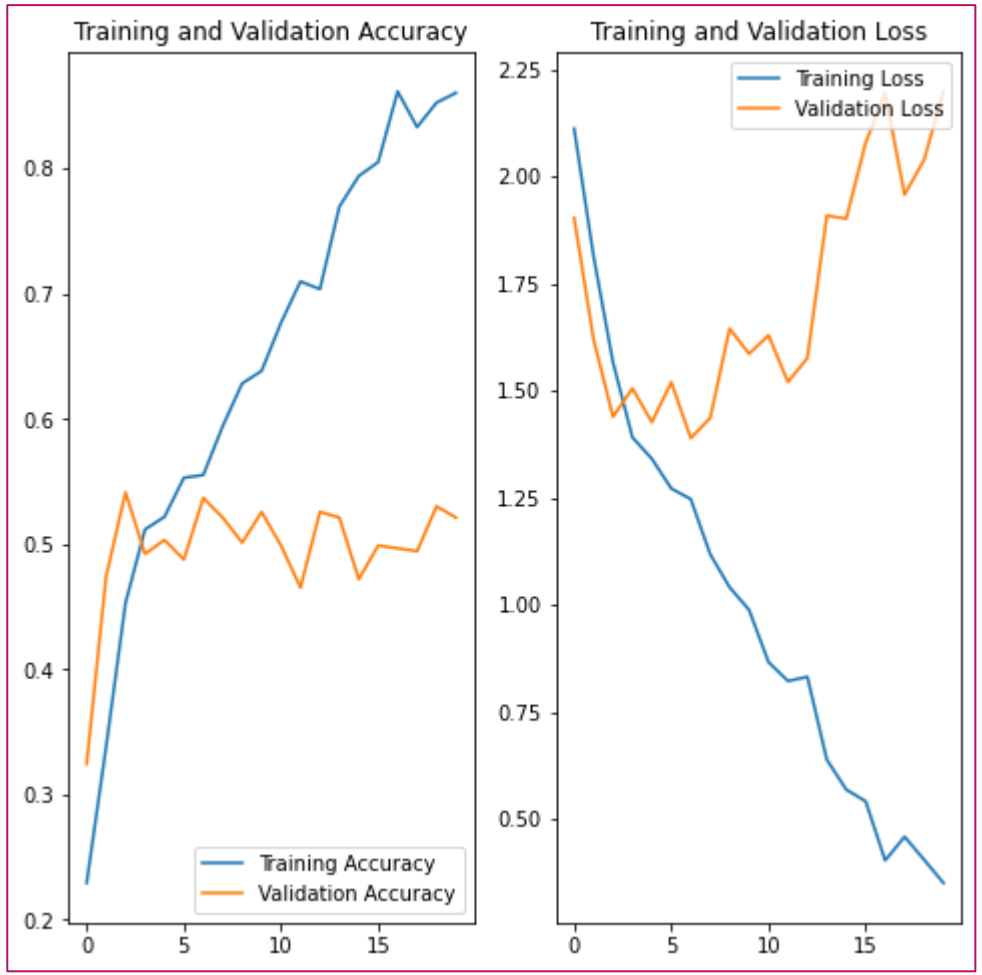
- **Data Reading/Data Understanding** → Defined the path for train and test images.
- **Dataset Creation** → Created train & validation dataset from the train directory with a batch size of 32. Also, resized images to 180x180.
- **Dataset Visualization** → Created a code to visualize one instance of all the nine classes present in the dataset.

Images from each classes from Train Dataset



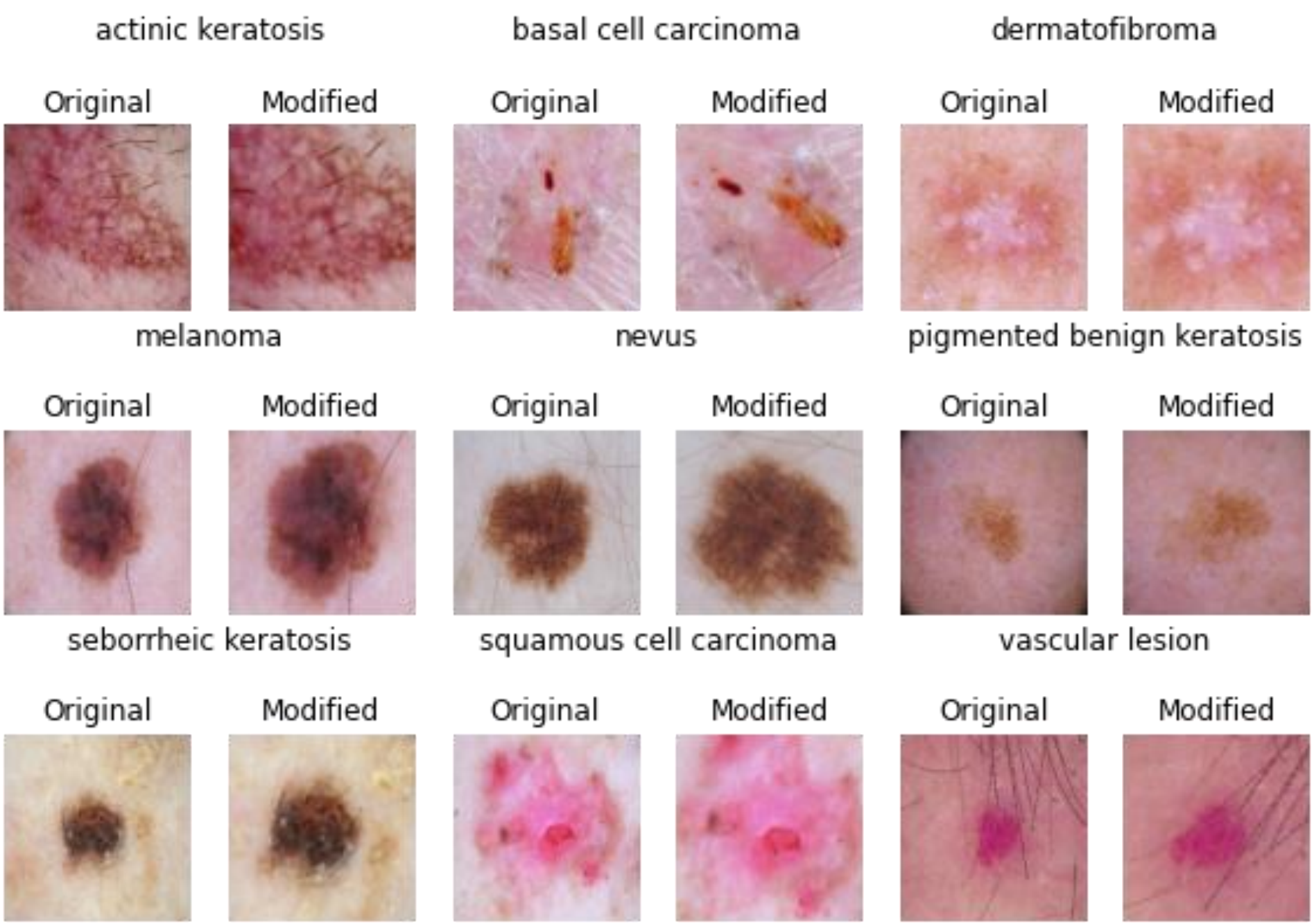
Base CNN Model Training

- **Model Building & training** → Created a CNN model to predict 9 types of skin cancer classes. Normalized pixel value to range [0,1]. Chose optimizer as Adam and Loss as Sparse Categorical Cross Entropy. Trained the model for 20 epochs. Observed the model performance - Overfitting of train dataset.
  - Training Dataset => loss: 0.35 - accuracy: 86.05 %
  - Validation Dataset => loss: 2.19 - accuracy: 52.13 %
- **Data Augmentation** → Created a augmentation layer to enhance/preprocess images for classification.
- **Model Building & training on the Augmented data** → Created a CNN model to predict 9 types of skin cancer classes. Preprocessed images using image augmentation layer. Normalized pixel value to rannge [0,1]. Added drop out layer to avoid overfitting. Chose optimizer as Adam and Loss as Sparse Categorical Cross Entropy. Trained the model for 20 epochs. Observed the model performance - Avoided overfitting but training dataset accuracy reduced and validation dataset accuracy is not improving.
  - Training Dataset => loss: 1.44 - accuracy: 48.38 %
  - Validation Dataset => loss: 1.43 - accuracy: 48.10 %



Base Model training - Accuracy and Loss of Model trained for 20 epochs for Training and Validation Dataset.

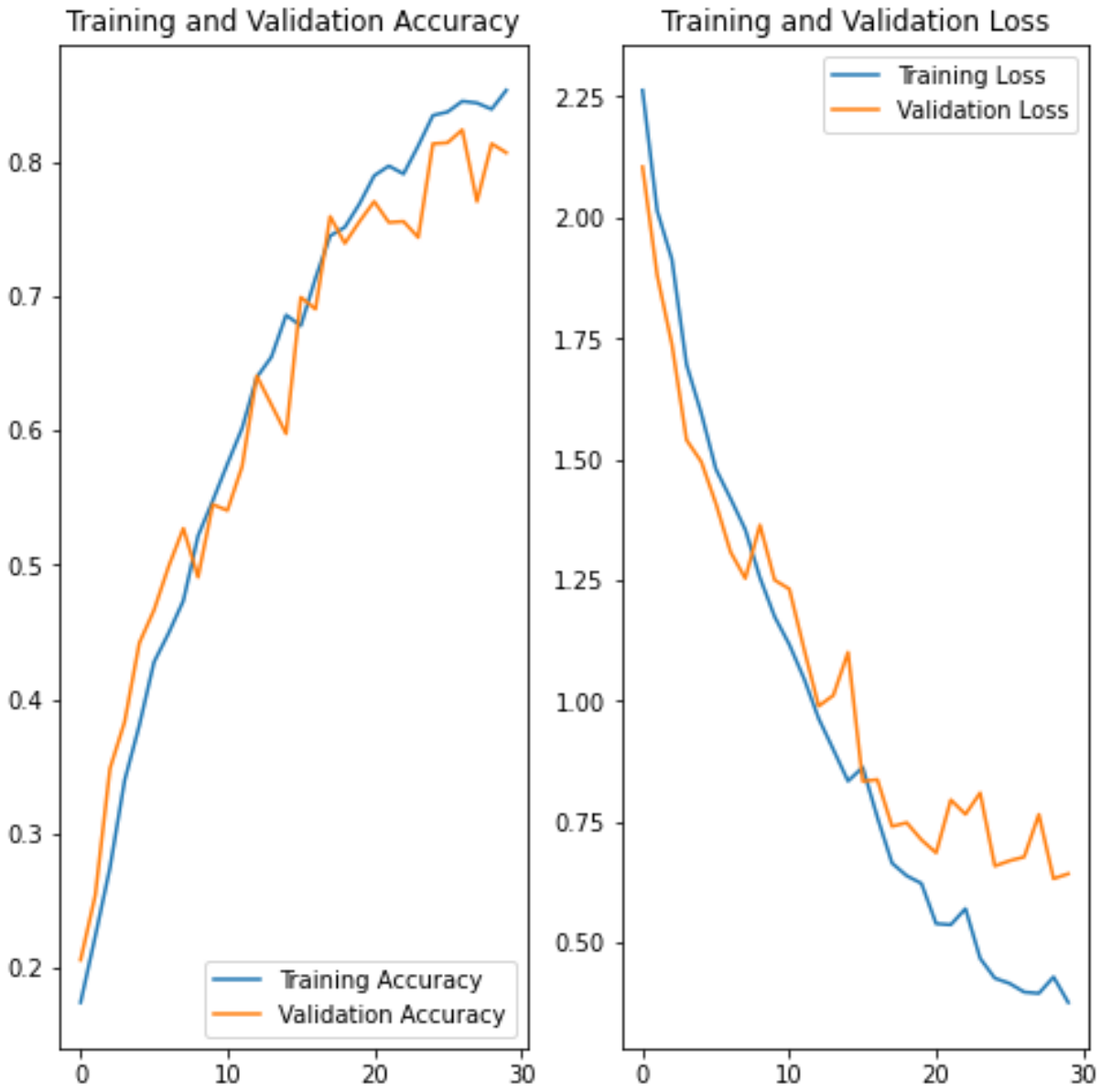
Base Model training on the Augmented data - Accuracy and Loss of Model trained for 20 epochs for Training and Validation Dataset.



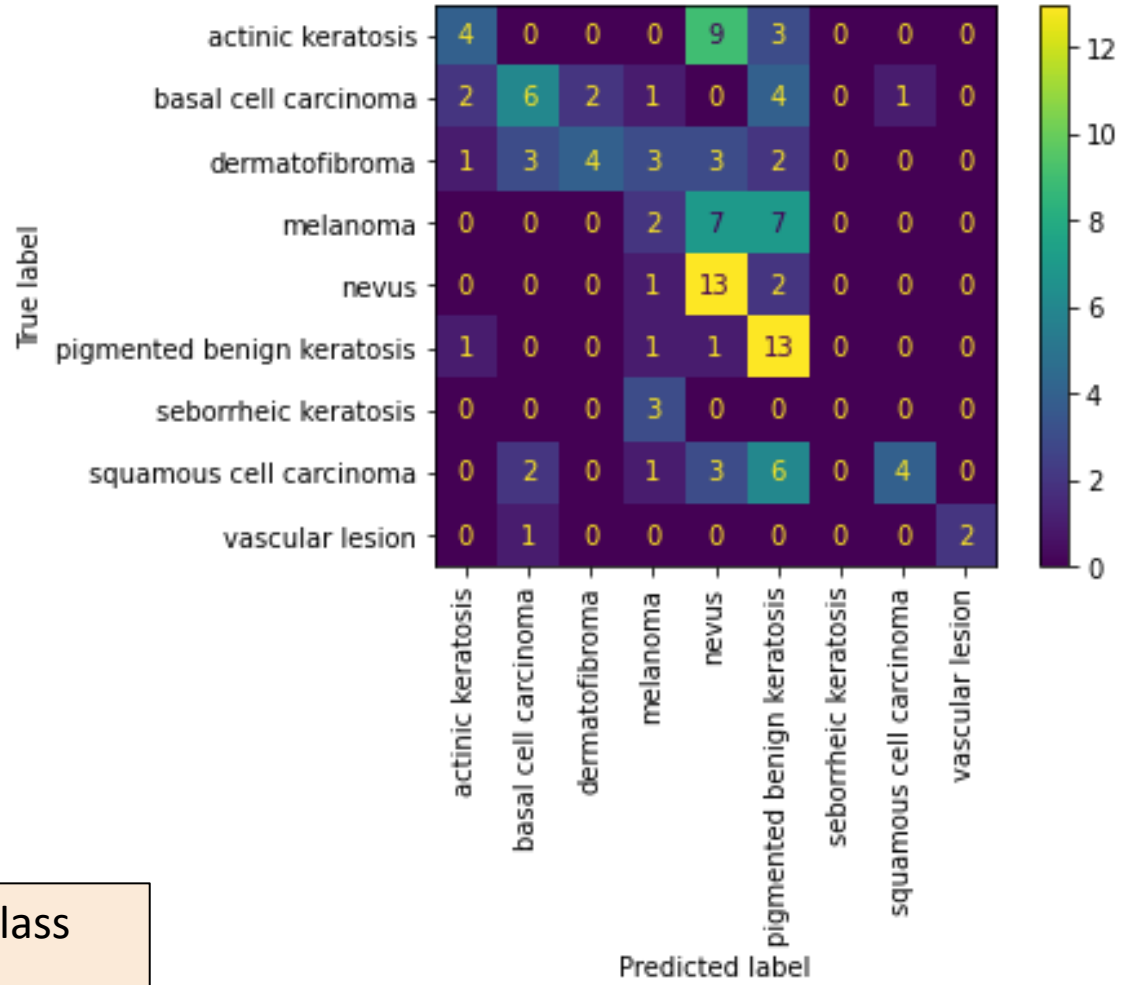
Images from each classes from Train Dataset – before and after applying data augmentation layer.

Final CNN Model Training

- **Class Distribution** → Examined class distribution.
  - seborrheic keratosis has the least number of samples only 77.
  - pigmented benign keratosis (462 Samples), melanoma (438 Samples), basal cell carcinoma (376 Samples), and nevus (357 Samples) classes dominates the data in terms proportionate number of samples .
- **Handling class imbalances** → Rectified class imbalances present in the training dataset with Augmentor library.
- **Model Building & training on the rectified class imbalance data** → Created a CNN model to predict 9 types of skin cancer classes. Normalized pixel value to range [0,1]. Added drop out layer to avoid overfitting. Chose optimizer as Adam and Loss as Sparse Categorical Cross Entropy. Trained the model for 20 and 30 epochs. Observed the model performance - Model trained for 30 epochs performs better in terms of accuracy and loss of training and validation dataset respectively. Although overfitting was avoided, the test data accuracy and loss is very less when compared to trained model.



Prediction of model on Test dataset – plotted as Confusion Matrix



Accuracy and Loss of Model trained for 30 epochs on the rectified class imbalance data for Training and Validation Dataset.

	precision	recall	f1-score	support
actinic keratosis	0.50	0.25	0.33	16
basal cell carcinoma	0.50	0.38	0.43	16
dermatofibroma	0.67	0.25	0.36	16
melanoma	0.17	0.12	0.14	16
nevus	0.36	0.81	0.50	16
pigmented benign keratosis	0.35	0.81	0.49	16
seborrheic keratosis	0.00	0.00	0.00	3
squamous cell carcinoma	0.80	0.25	0.38	16
vascular lesion	1.00	0.67	0.80	3
accuracy			0.41	118
macro avg	0.48	0.39	0.38	118
weighted avg	0.48	0.41	0.38	118

Model Classification Report

Image Prediction
Actual Class is nevus and
Predicted Class is nevus

Conclusion and Recommendation

Model trained on the rectified class imbalance data with 30 epochs is chosen to classify 9 types of skin cancer accurately.

- Training Dataset => loss: **0.37** - accuracy: **85.35** %
- Validation Dataset => loss: **0.64** - accuracy: **80.70** %
- Test Dataset => loss: **4.6** - accuracy: **39.83** %