Skin Cancer Image Detection Report

The dataset consists of 2357 images of malignant and benign oncological diseases, which were formed from the International Skin Imaging Collaboration (ISIC). All images were sorted according to the classification taken with ISIC, and all subsets were divided into the same number of images, with the exception of melanomas and moles, whose images are slightly dominant. The data set contains the following diseases:

- Actinic keratosis
- Basal cell carcinoma
- Dermatofibroma
- Melanoma
- Nevus
- Pigmented benign keratosis
- Seborrheic keratosis
- Squamous cell carcinoma
- Vascular lesion

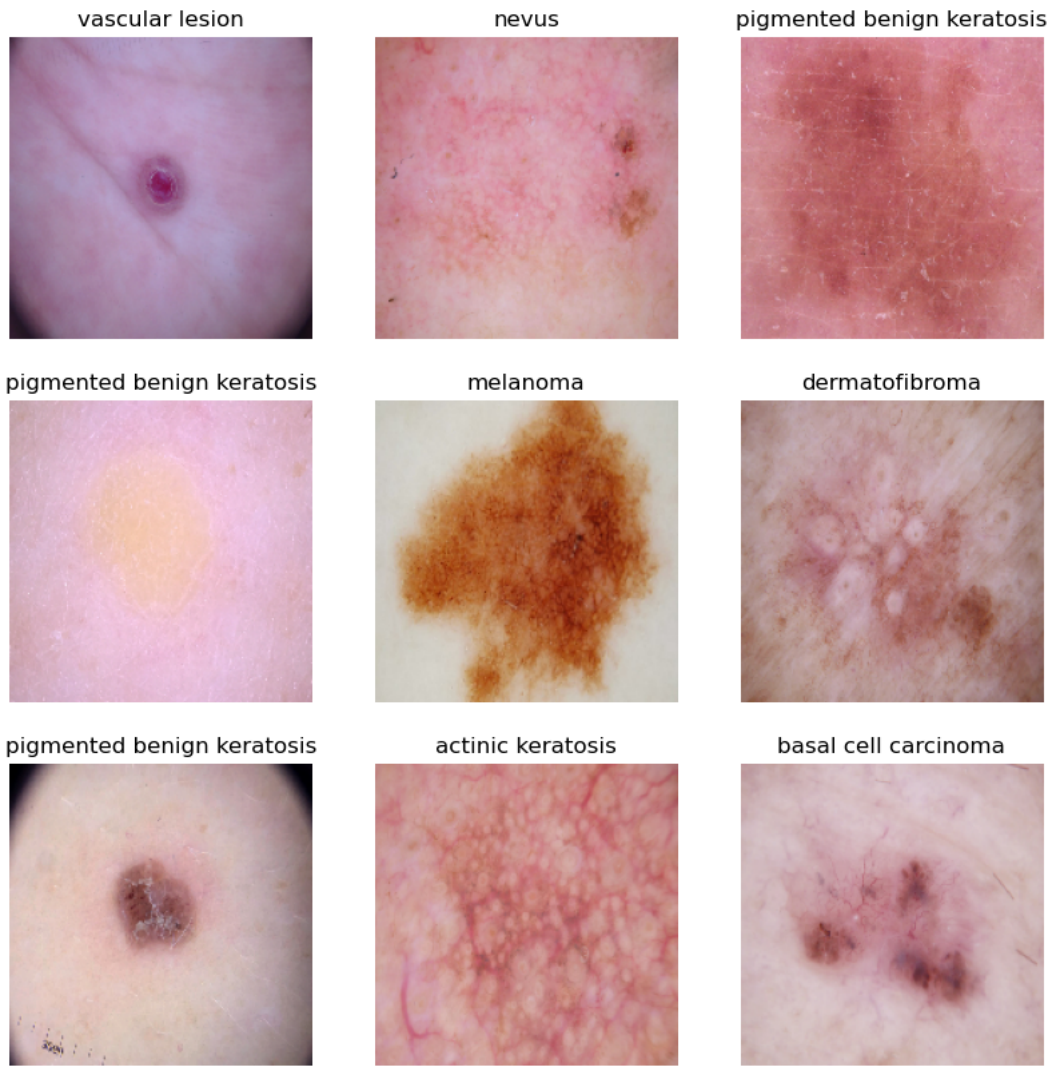
Problem statement: To build a CNN based model which can accurately detect melanoma. Melanoma is a type of cancer that can be deadly if not detected early. It accounts for 75% of skin cancer deaths. A solution that can evaluate images and alert dermatologists about the presence of melanoma has the potential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

The strategy for building the model comprised of the following steps:

- Data Reading and understanding
- Visualizing the one instance from each of the class
- Model Building
- Model Evaluation

Data reading and Understanding:

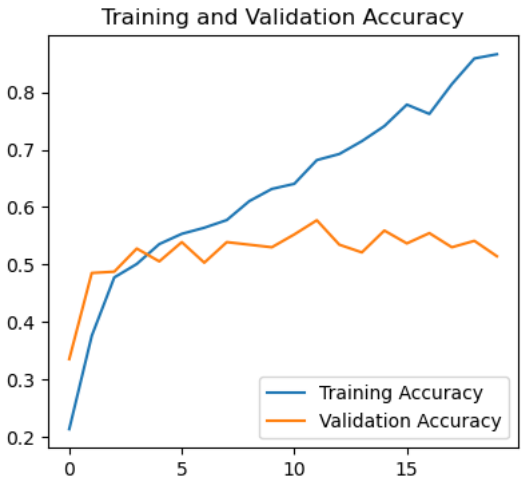
After Data Reading and understanding, this is how the sample images from each class looked like:

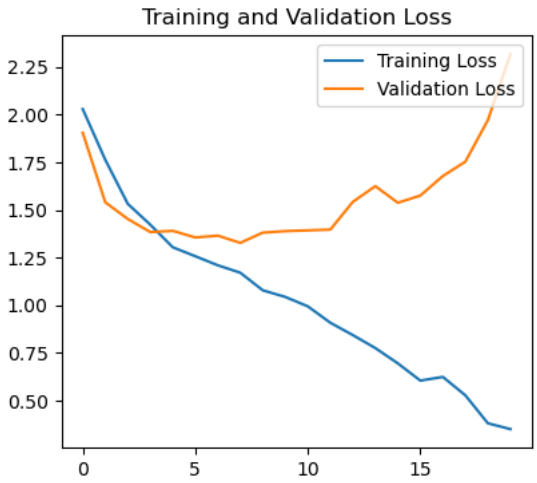


As can be seen from the dataset most of the images have their key information in only a small region around the infection.

Model Building:

Model 1: Comprised of 4 Convolutional and Maxpooling layers followed by flatten layer and a Dense layer of 512 neurons. However the results showed a lot overfitting as well as low accuracy.

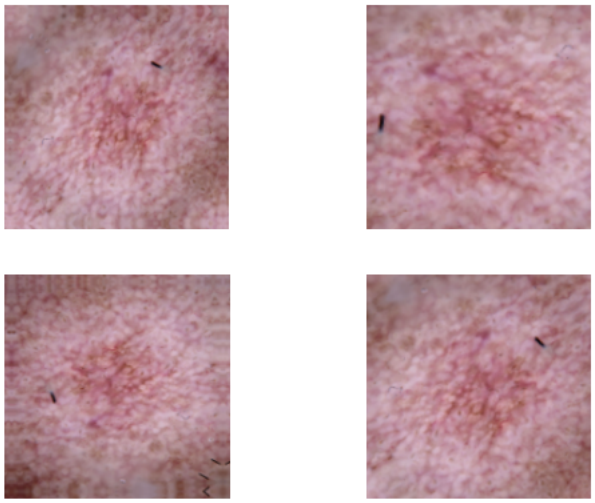




Model 2:

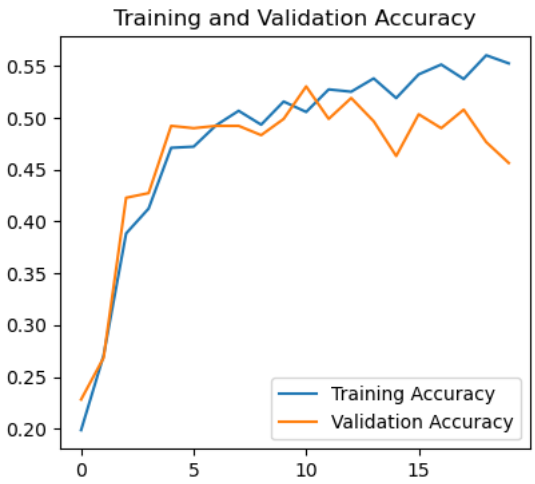
Decided to go with data augmentation and dropout layer to check for improvement.

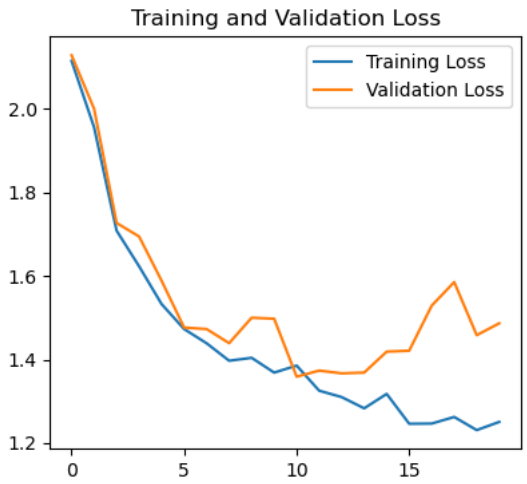
Data augmentation helps to create many rotated ,flipped ,cropped or zoomed images from each of the sample of images as can be seen below:



This can help in providing more data as well more learning.

The results of the model helped to improve the overfitting to some extent but still accuracy was not upto the mark.

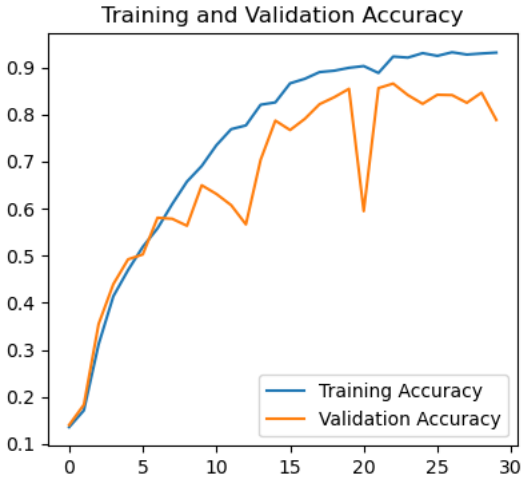


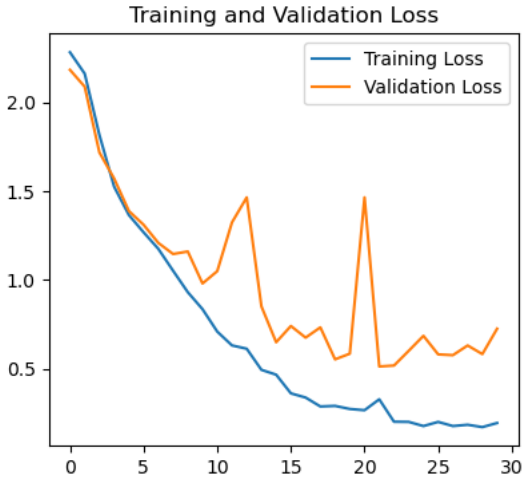


Model 3:

In order to further improve the model, the Augmentor library from python was used to handle with class imbalance as few classes like **melanoma, pigmented benign keratosis, basal cell carcinoma** dominated the data. With the help of Augmentor 500 images were added per class to deal with imbalance.

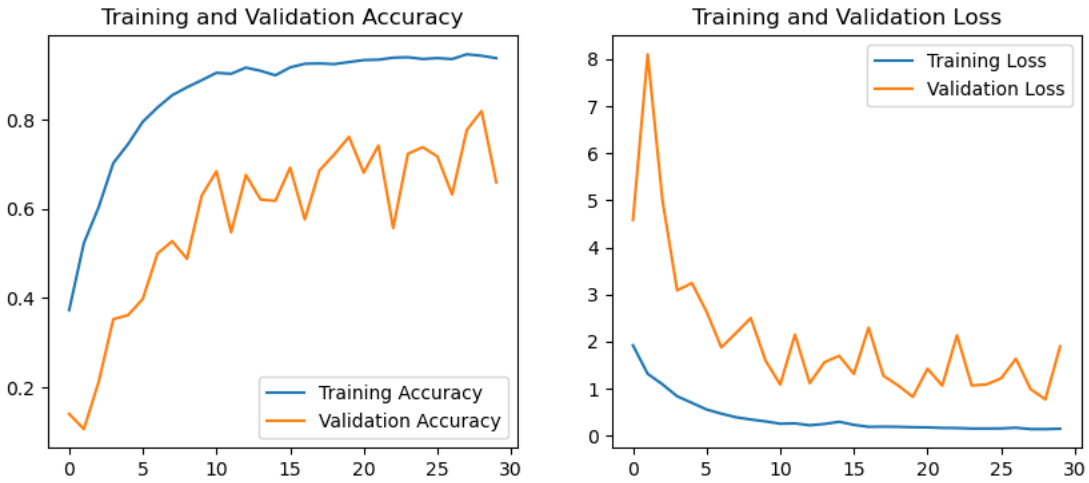
As can be seen from below result the class imbalance has improved overfitting and also accuracy increased greatly.





Model 4:

Tried with Batch Normalization to see for any improvements in performance.



As can be seen from above graph that the model becomes quite unstable and accuracy drops with batch normalization as compared to without normalization.

Conclusion:

So as compared to all the models, Model 3 turns out to be the best, which contains the imbalance handling and dropouts.