Melanoma Case Study

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.

Goal:

Reduce a lot of manual effort needed in diagnosis

Model:

Convolutional Neural Network Case Study

Data Understanding

Data Description:

Images – 2357

Data Set Contains images of the mentioned diseases:

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

Model Creation Consideration:

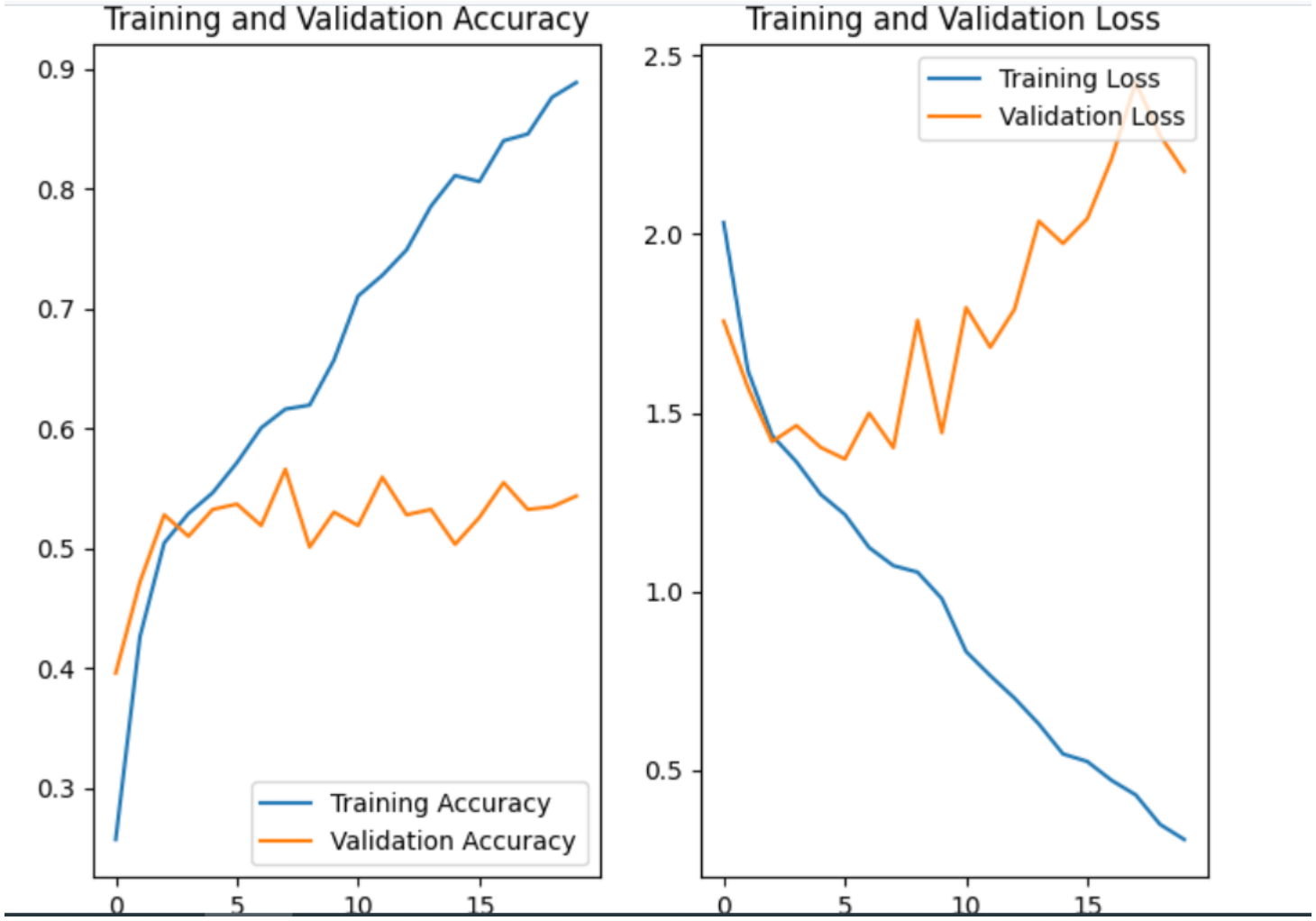
- Data Reading/Data Understanding
- Dataset creation
- Dataset visualization
- Model Building & training
- Data augmentation
- Model Building & training
- Class distribution
- Handling class imbalances
- Model Building & training
- Inferences

1st Model

- There is huge variance between the Training Accuracy & Validation Accuracy.

Training Accuracy is at approx. 90% and validation accuracy at approx. 55% is indicating that there is an overfitting issue.

Applying Data Augmentation Technique to resolve the overfitting issue.

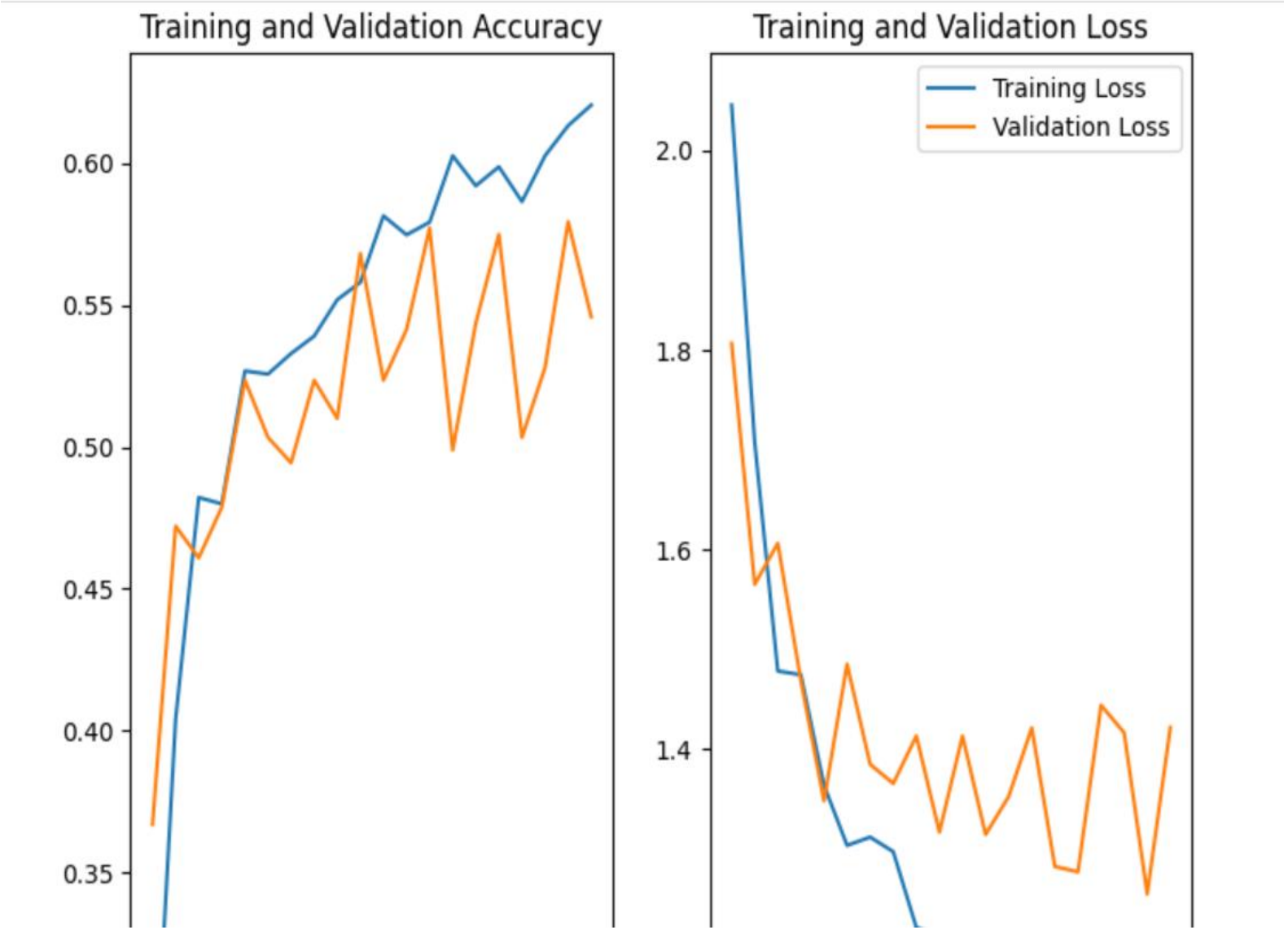


2nd Model

- The variance in Training Accuracy & Validation Accuracy has substantially reduced from Training Accuracy is at approx. 90% and validation accuracy at approx. 55% without data augmentation.

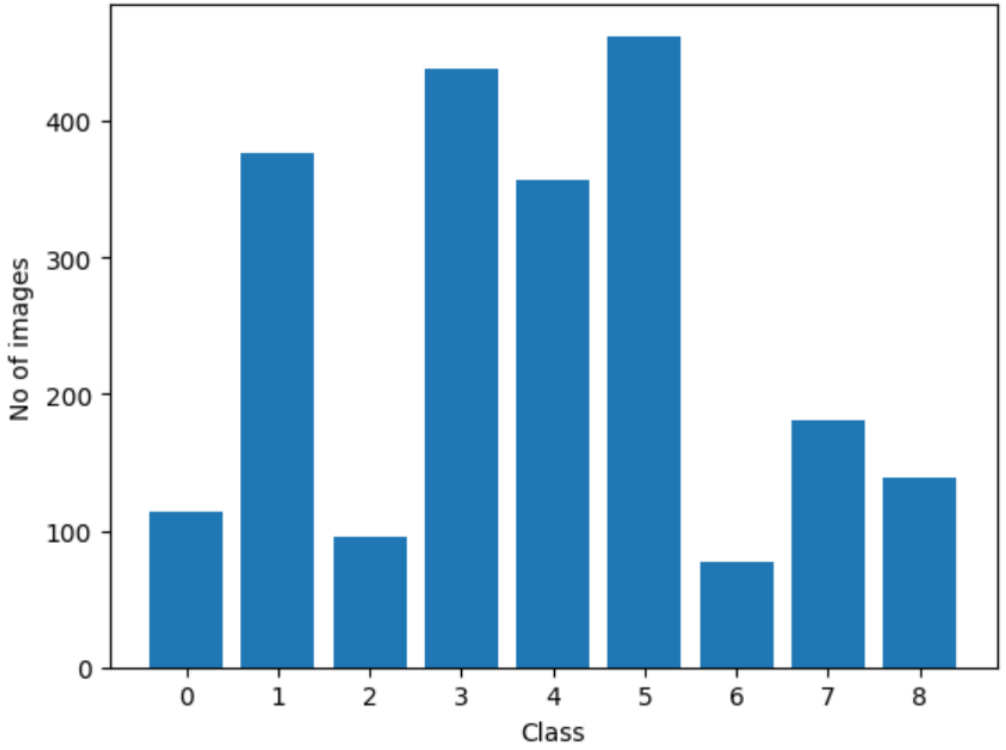
The current training accuracy approx 60% and Test Accuracy greater than 50%.

Looking at enhancing the accuracy of the model using Class Distribution and reducing the imbalances



Class Imbalance

```
Class=0, n=114 (5.092%)
Class=1, n=376 (16.793%)
Class=2, n=95 (4.243%)
Class=3, n=438 (19.562%)
Class=4, n=357 (15.945%)
Class=5, n=462 (20.634%)
Class=6, n=77 (3.439%)
Class=7, n=181 (8.084%)
Class=8, n=139 (6.208%)
```



3rd Model

- The model accuracy for Training is around 60% and validation accuracy is around greater than 55%.

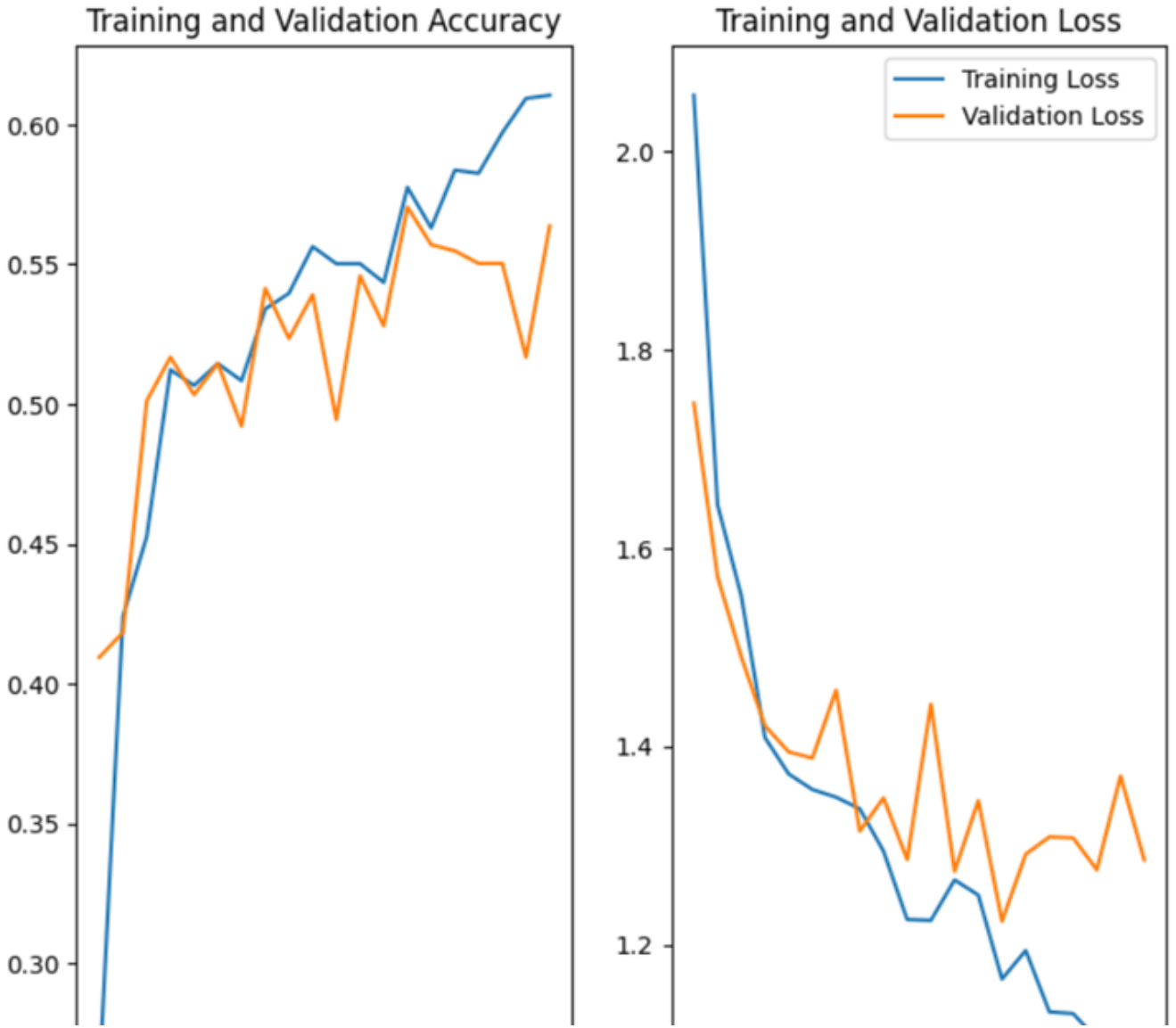
There variance has been reduced from the initial model of Training Accuracy at approx. 90% and validation accuracy at approx. 55% without data augmentation.

After using data augmentation, it was around training accuracy approx. 60% and Test Accuracy around 50%.

After reducing the class imbalances and data augmentation the training accuracy is greater than 60% and Test Accuracy above 56%, this has helped in better accuracy and resolving overfitting.

With the limited total population sample of 2239 and training sample size of 1792, the results generated have proved that is a good model for prediction.

Multiple code runs had been done



Thank You