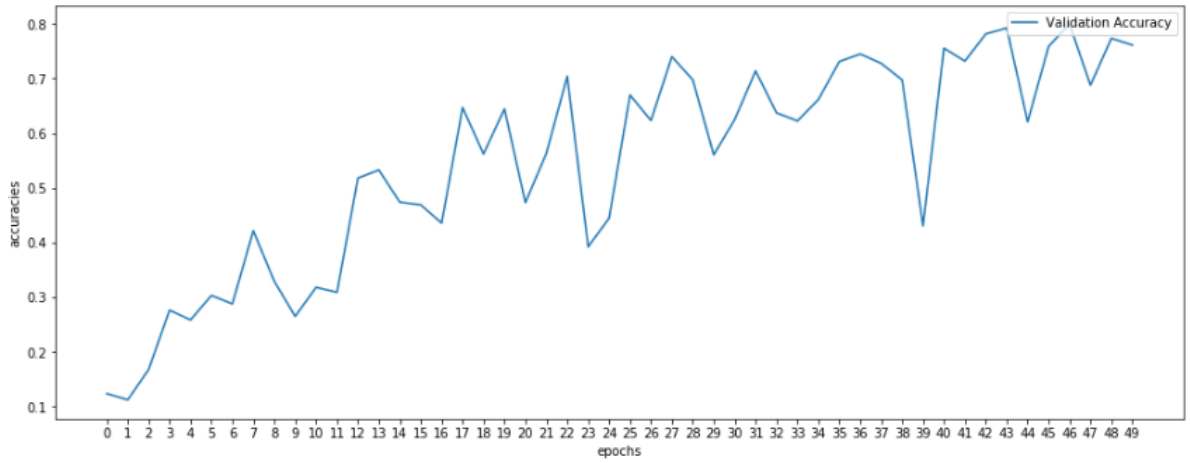
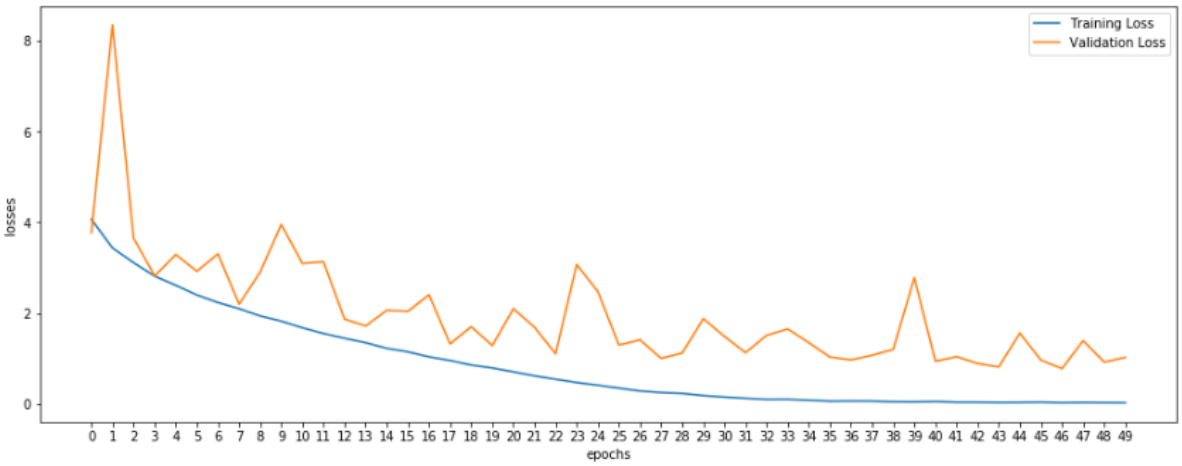
Mode A: Without loading pretrained weights, train all layers

Validation Accuracy



Training and Validation Loss



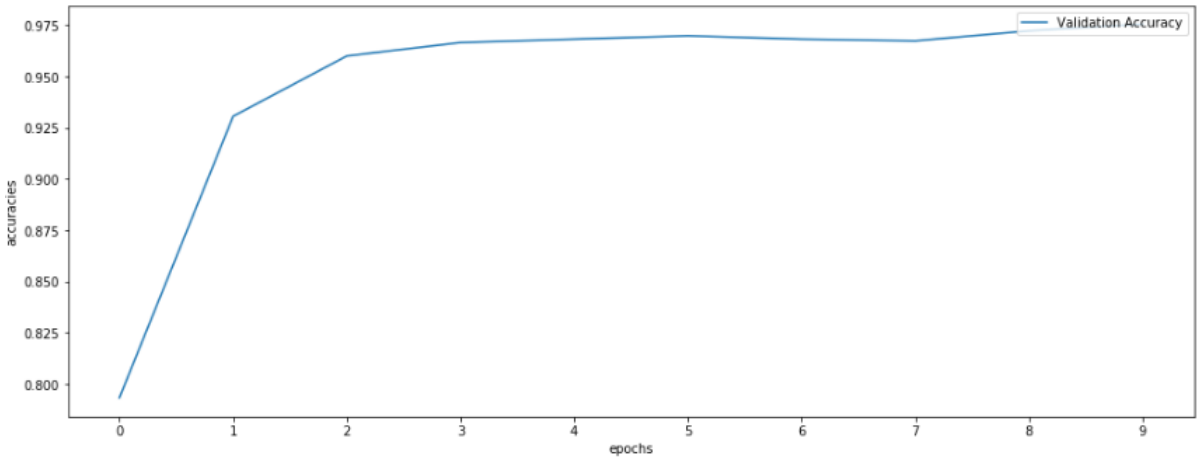
Results:

After training for 50 epochs, we got the best model parameters from Epoch 46, with the highest validation accuracy of 79.98%.

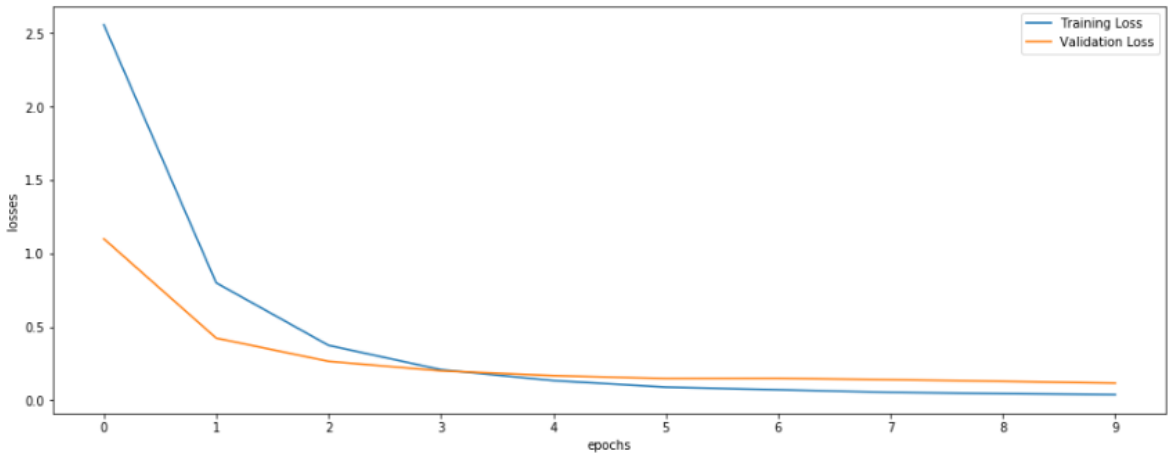
Testing these model parameters against the test set, we got an accuracy of 70.58%

Mode B: Load pretrained weights, train all layers

Validation Accuracy



Training and Validation Loss



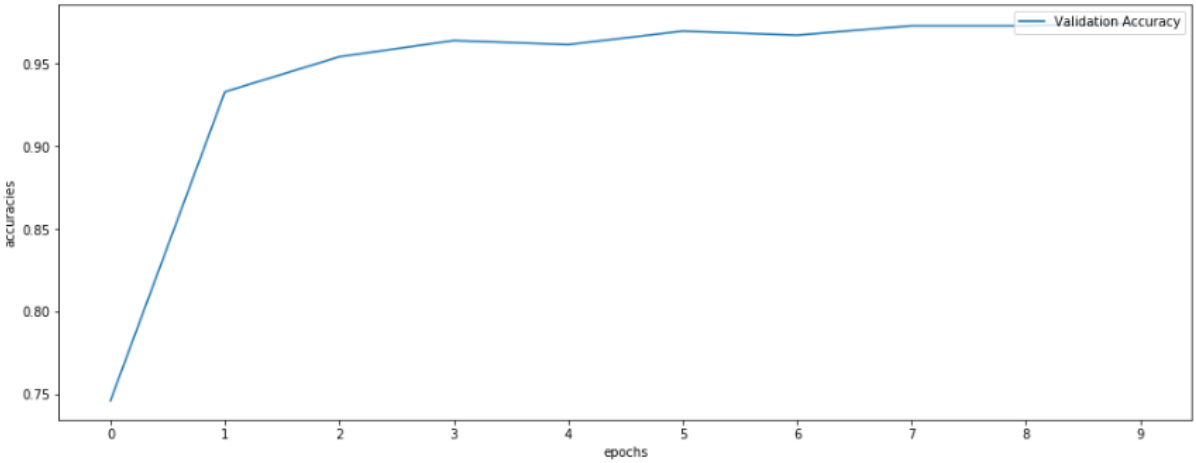
Results:

After training for 10 epochs, we got the best model parameters from Epoch 9, with the highest validation accuracy of 97.55%.

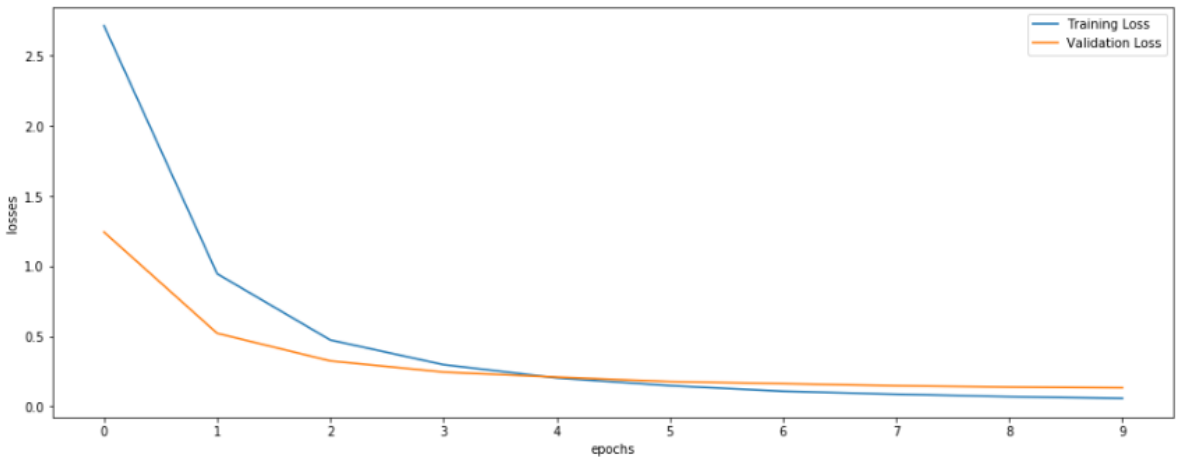
Testing these model parameters against the test set, we got an accuracy of 96.95%

Mode C: Load pretrained weights, train only last 2 trainable layers ('layer4' and 'fc' layers)

Validation Accuracy



Training and Validation Loss



Results:

After training for 10 epochs, we got the best model parameters from Epoch 9, with the highest validation accuracy of 97.47%.

Testing these model parameters against the test set, we got an accuracy of 96.24%

Observation of the differences between validation and test accuracies:

Despite training the models in their respective distinct modes, A, B and C, one observation made was that throughout those trainings, the test accuracy tended to be closest to the model that had weights giving rise to the highest validation accuracy over all the epochs.

Another observation made is that the test accuracy was usually lower than the validation accuracy. This could be due to the model having generalized well to the validation set used. Hence when using the test set which assumedly has hidden data more generalized to real world circumstances, it would make sense for the test set accuracy to be lower since it is less generalized compared to the validation set.