PROJECT 2
Performance report with observations and conclusions

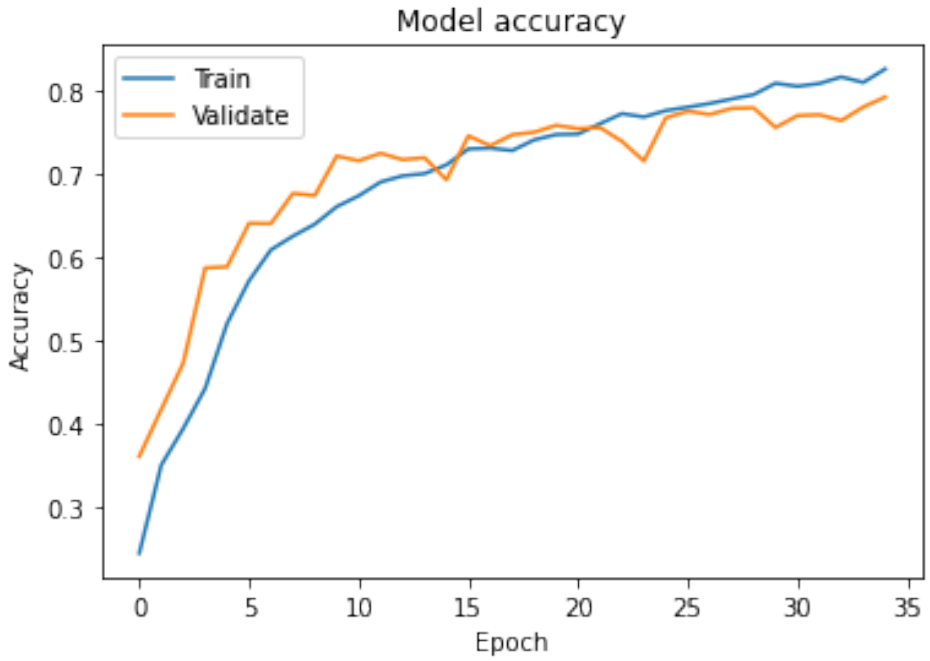
Kanak Tenguria
A02310652

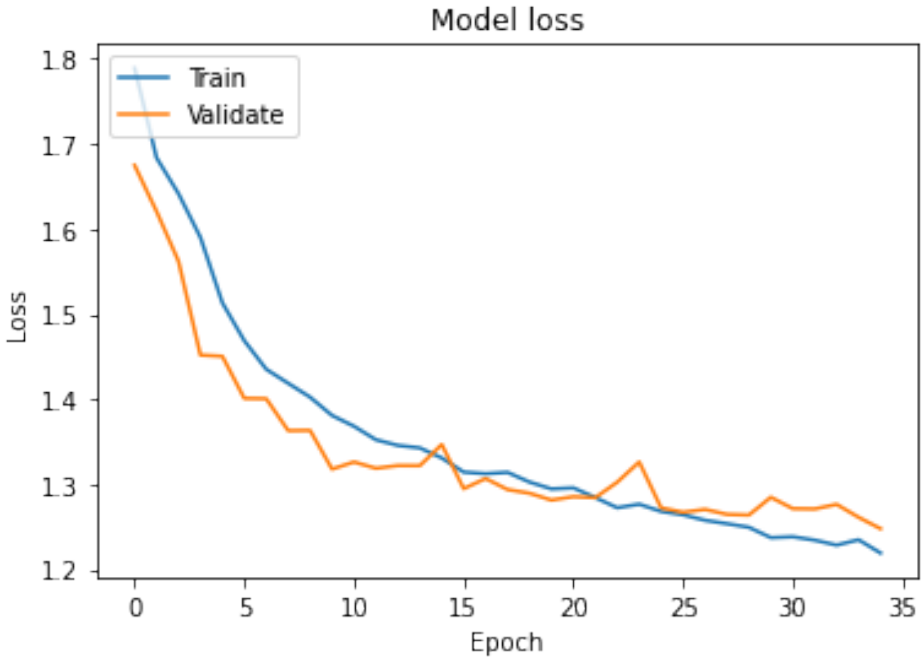
1 Methods and Observations:

To classify the images, I have used CNN, ANN and Random Forest techniques. I have used 11034 images for training, 3000 images for validation and 3000 images for testing. Other than this, I also had 7301 unlabelled images, which were provided by the kaggle for that competition's evaluation. I used random images from this unlabelled data to test how accurate my models are classifying (The determination of accuracy for this data was manual. I saw the images and estimated the performance of the nets). The images were of size 150 x 150 pixels but I reshaped them to 64 x 64 pixels because of computational resource constraints.
For CNN and ANN, the validation loss and validation accuracy along with training loss and training accuracy helped me in early stopping the training and gave me insights about whether my model is overfitting or not. It also helped me determine number of epochs necessary and to tune other hyperparameters.

1.1 CNN Training:

The 2 graphs below show the training accuracy vs validation accuracy and training loss vs validation loss while training.

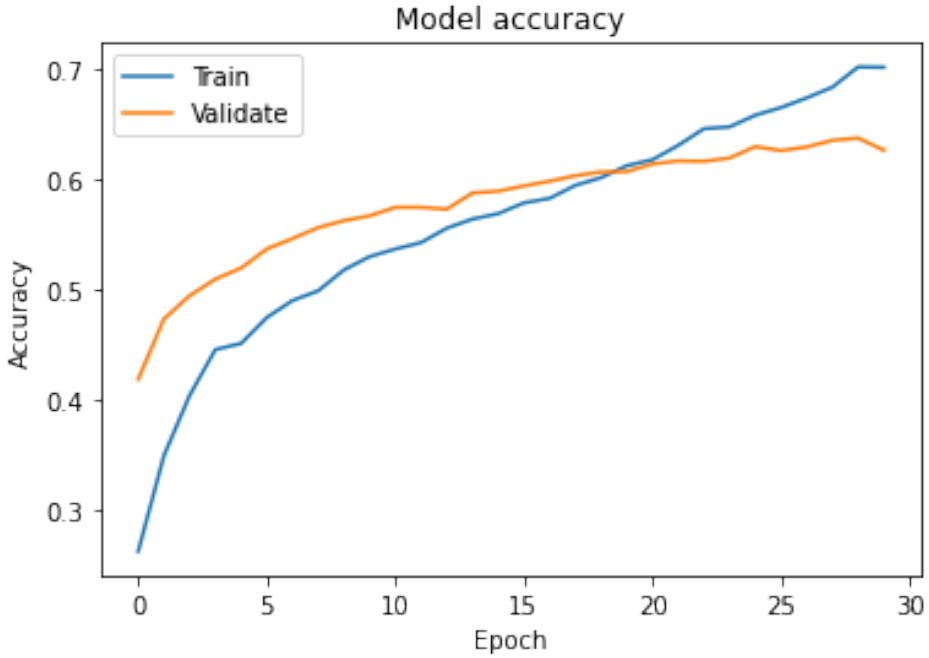


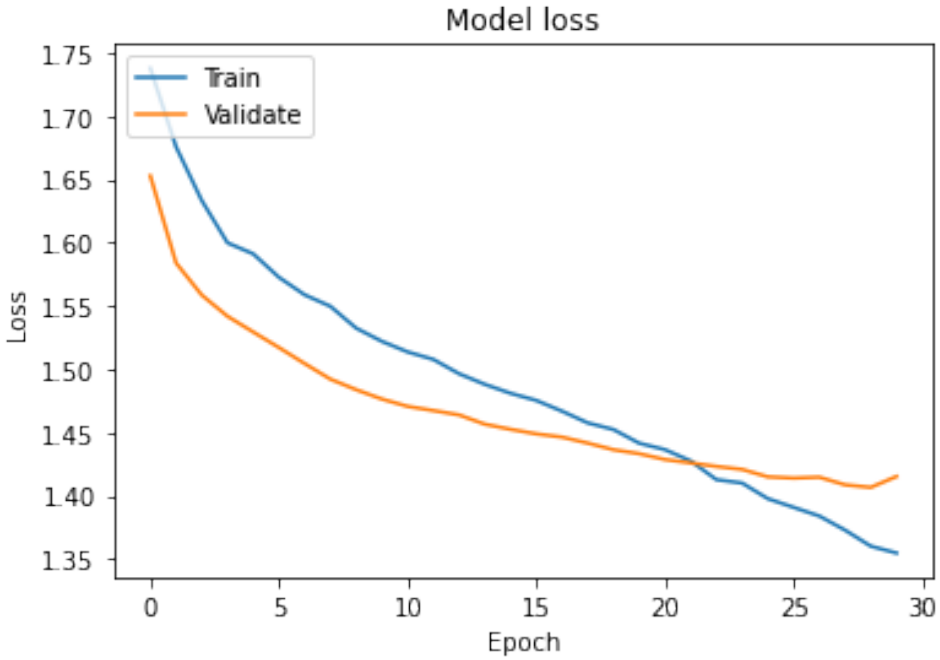


As it is clear from the graph, initially the validation loss and validation accuracy were higher than training loss and training accuracy. I think it is because of the dropout layers. It made the model more robust. As the model started learning more features from the training data, it started recognizing training images better than the validation images. I stopped the training as soon as the validation loss started to fluctuate rather than decrease and training loss kept decreasing to avoid overfitting.

1.2 ANN Training:

The 2 graphs below show the training accuracy vs validation accuracy and training loss vs validation loss while training.





Just like CNN model, ANN model also showed better validation accuracy and loss initially, and as the model started learning features from training data, it started performing poorly on validation data in comparison to training data. Here also, I think this initial behaviour is because of the dropout layers.
For the ANN, I normalized the data to boost up training time.

1.3 Random Forest Training:

I tried 50, 100, 150, 200, 250, 300, 350, 400, 450 and 500 decision trees in the random forest. Best accuracy was achieved using Random Forest with 500 decision trees. Since there was not much improvement, I did not experiment further with more trees.

1.4 Results:

Results obtained using various methods are shown in table below.

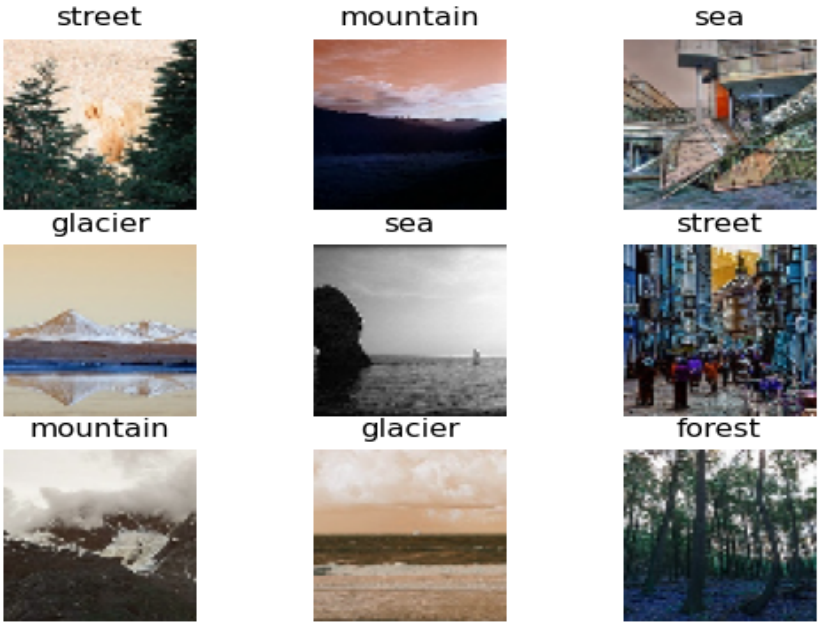
Method	Validation	Testing
CNN	**79.23**	**79.37**
ANN	**62.67**	**61.83**
RF (50)	58.9	57.6
RF (100)	61.93	59.2
RF (150)	61	59.3
RF (200)	61.56	60.1
RF (250)	62.3	60.6
RF (300)	62	60.23
RF (350)	61.8	60.26
RF (400)	61.5	60.8
RF (450)	62.26	60.63
RF (500)	**62.3**	**60.96**

1.4.1 Prediction using CNN:

Below are the labels predicted by CNN for respective images. These images are unlabelled and classifications are verified manually by me. CNN classified most of the images correctly. From the classification report, it is clear that CNN was able to classify forest much better than any other label.

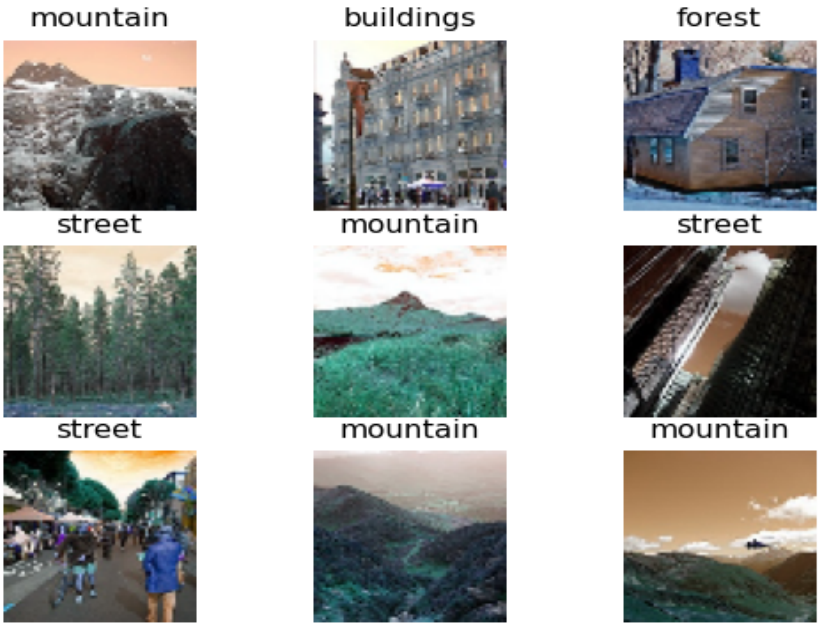
1.4.2 Prediction using ANN:

Below are the labels predicted by ANN for respective images. These images are unlabelled and classifications are verified manually by me. ANN did a poor job at classifying the images. Just like CNN, ANN also classified forest images more precisely and it got confused between glacier and sea.



1.4.3 Prediction using Random Forest:

Below are the labels predicted by Random Forest with 500 trees for respective images. These images are unlabelled and classifications are verified manually by me. Random Forest did not classify many images correctly. From the confusion matrix, it is understood that Random Forest is getting confused between sea and glacier images.



2 Conclusion:

From the results, it is clear that CNN performed better than any other methods. ANN and Random Forest performed almost similarly but ANN showed slightly better performance. Also, CNN and random forest with 500 trees took almost the same amount of time for training. ANN took a lot more time than the other two. Random forest with 500 trees performed similar to ANN but was much faster. From what I noticed, CNN is best for classification of images in my case.
In my understanding, the models would have performed better if I would have used original images of 150 x 150 pixels rather than resized images of size 64 x 64 pixels. It must have lost a lot of features because of resizing. Secondly, I believe that quite a few images in the dataset were too confusing for humans to classify as well, which can be the another reason why the test accuracies were not very high.