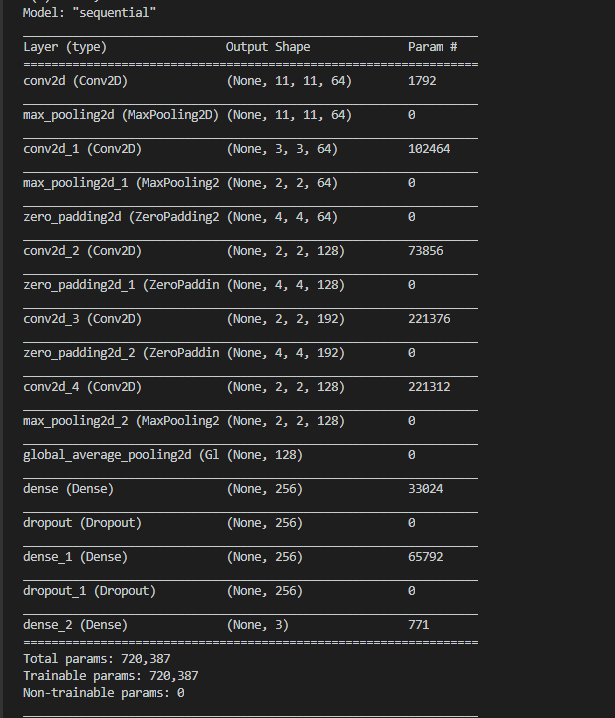
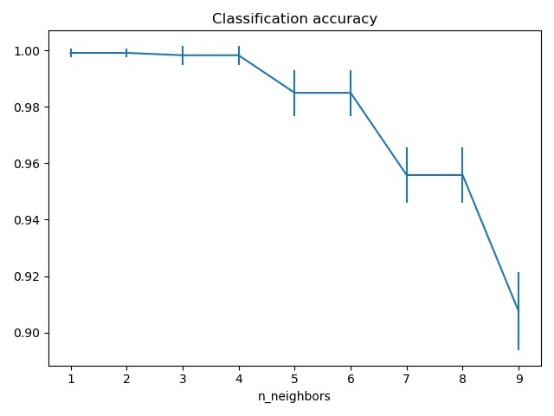
Assignment 3

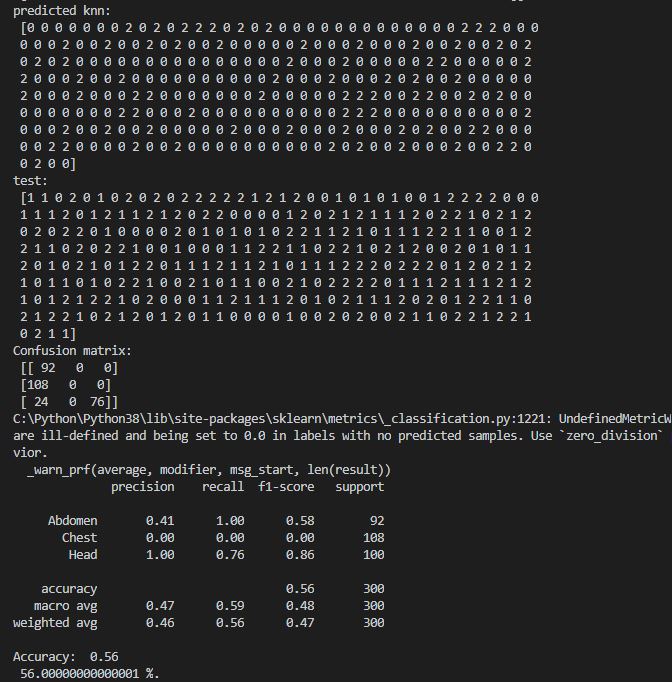
1. Pre-processing step includes resize the image into 32x32 and label each image. I define a *imgLabel()* function it returns images as array and its labels. Before fitting the labels into CNN, they are encoded into one-hot-vector.
2. CNN model has 5 convolutional layers and 3 dense layers



1. Total 1500 images are spilt by 80:20 rate, 1200 for training and 300 for testing. Only use training data for training. When training the CNN cross validation methods is applied, using *“StratifiedKFold”*. *K = 5*. The numbers of *epochs = 5*.
2. Features are extracted from the first dense layer of CNN.
3. Training KNN classifier and find the optimal value of K, from 1 to 9 when *K = 1* the performance is optimal. 
4. For Random Forest, two hyperparemeters are *n\_ estimators* (number of trees) and *max\_features* (the maximum number of features to consider while looking for split. When *n\_ estimators=20, max\_features=’sqrt’-sqrt(number of features).* It will have the best performance.

Performance of KNN

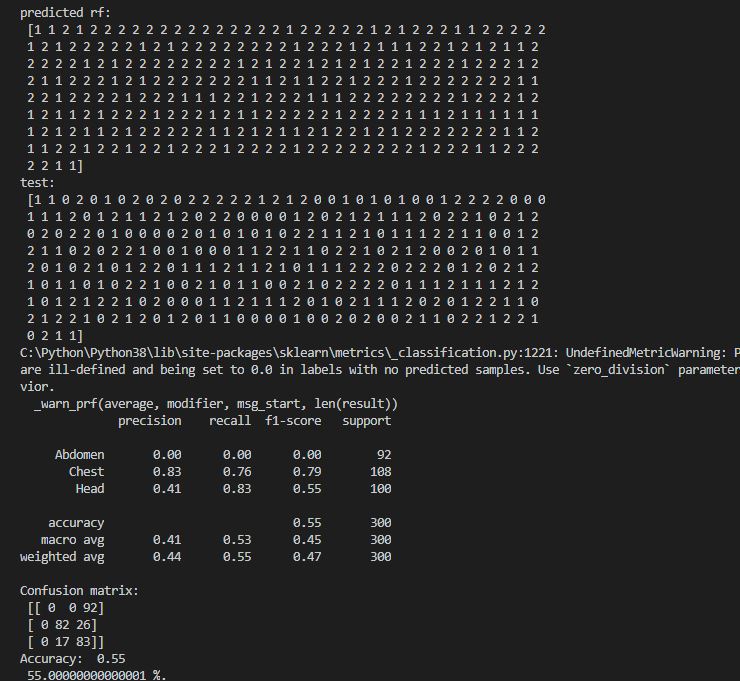
Load the saved KNN model get the accuracy of 56%:



The first class are all classified correctly, the second class are fully misclassified, the third class has mixed first class and its own class.

(without using pickled model, using the trained and saved model in the same session the accuracy is almost 100%.)

Performance of RF

Same situation as KNN the performance of RF is 55%, but without using the pickled model, the performance is 100%.

Conclusion

After training and testing in split dataset, there is no overfitting issue or underfitting, the performance is valid.

The only problem here is that loading from pkl file cause the lower performance and it is because the order of features from the pickled model is differ from the test model, so when load the pickle file to test, we can see from the confusion matrix that one of the class are fully misclassified. That could be explained as the order of features are altered and it cannot simple classify correctly, Pickle file is not reliable.