**Accuracy results:**

Step 1: 97.83 %

Step 2: 99.01 %

Step 3: 99.27 %

Step 4: 99.15 %

Step 5: <training did not complete in time>

Step 6: 99.5 %

Step 7: 99.48

**Discussion:**

As can be seen from the results, accuracy increases with the addition of convolutional layers in step 2 and 3, as training the weights in these layers enable the extraction of more features. The replacement of Sigmoid with ReLU activation in step 4 should slightly increase accuracy. This is probably because of the non-saturation of its gradient, which accelerates the convergence of gradient descent compared to the sigmoid function. The sparsity effects of ReLU is also another advantage it has over sigmoid activation. Accuracy improves in step 5 because of augmenting the training images, so the convolutional kernels extract more information from the images. Adding another fully connected layer in step 6 increases the number of trainable parameters and therefore enables more features to be learned. This is also the case when the size of the fully connected layer is increased in step 7. The use of Dropout in step 7 acts as a regularizer by reducing overfitting of data. Hence, accuracy improves in steps 6 and 7.

Due to fluctuations when training these NNs, there is some variability in the expected results, especially due to the fine margins of improvement. However, a general trend of improving accuracy can be seen from steps 1 through 7 because of the reasons outlined above.