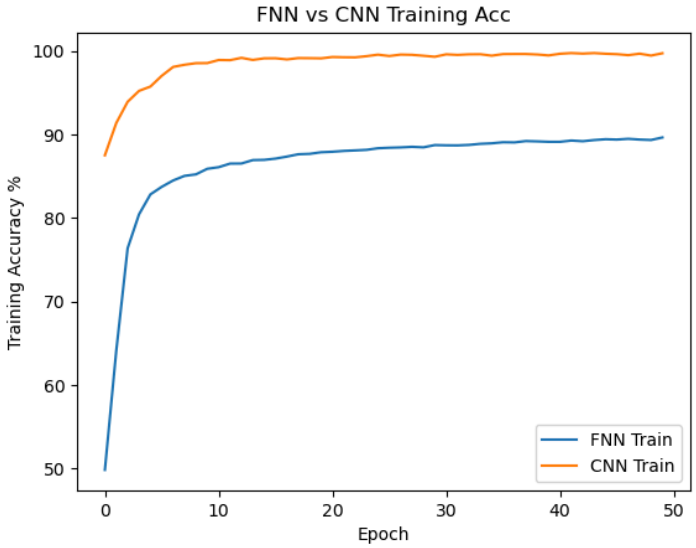
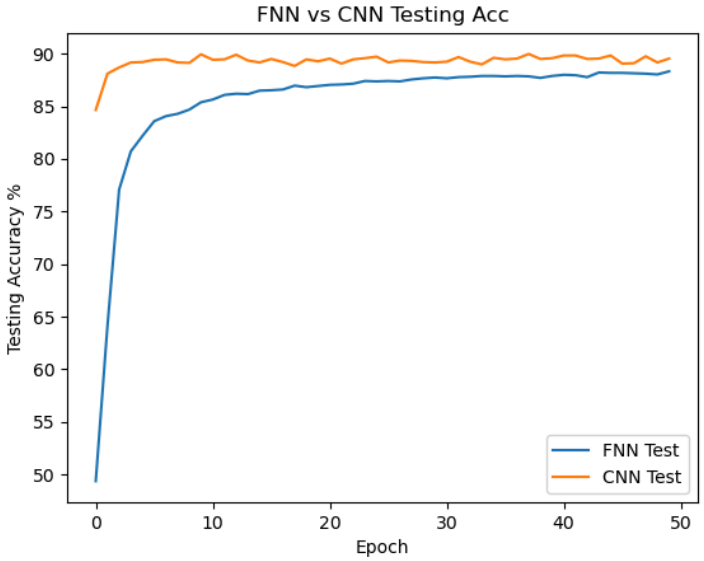
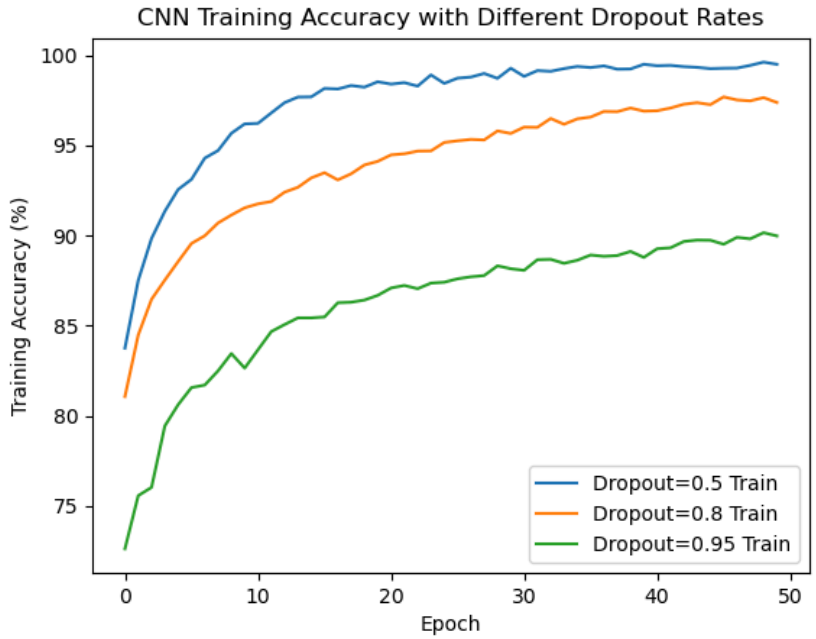
Experiment 1:

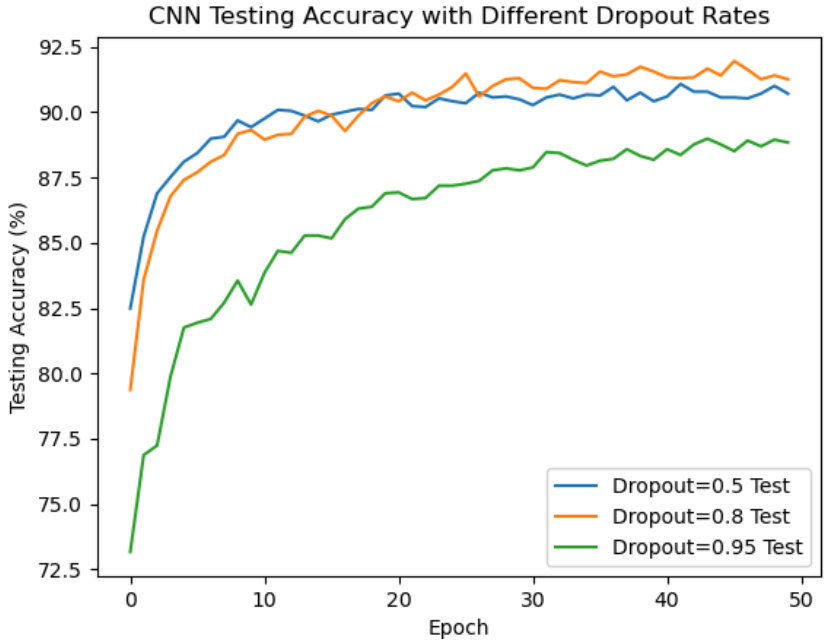




Overall, the CNN model performs better at the task of classifying images of letters. This is likely due to how CNN is designed with kernels and pooling, which obtains important features of an image better than that of a FNN.

Experiment 2:

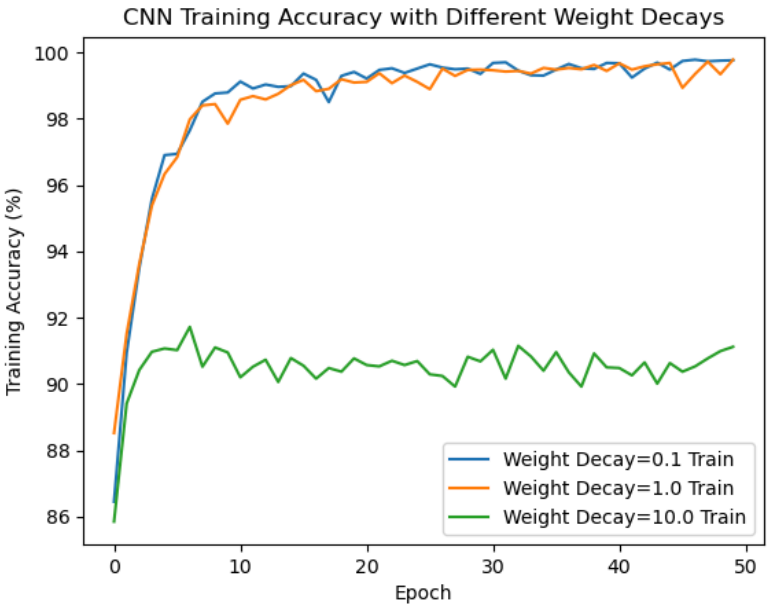


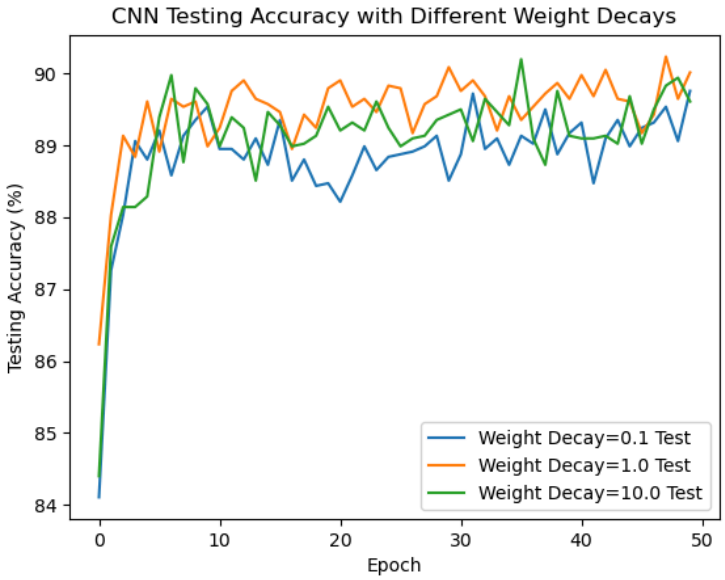


Influence of dropout rate:

The lowest dropout model achieves the highest training accuracy. The training accuracy decreases drops as the dropout rate increases. This is because a high dropout rate deactivates more neurons and forces the model to memorize less.
The middle dropout model achieves the highest testing accuracy, suggesting that a moderate amount of dropout can help the model generalize better.

Experiment 3:





Influence of Weight Decay:

Too small of a weight decay causes a bit of overfitting (training accuracy high whereas testing accuracy is relatively low). Too high of a weight decay can stop the model from learning well (training accuracy is stuck at around 89-91%). As a result, a moderate weight decay performs the best overall.