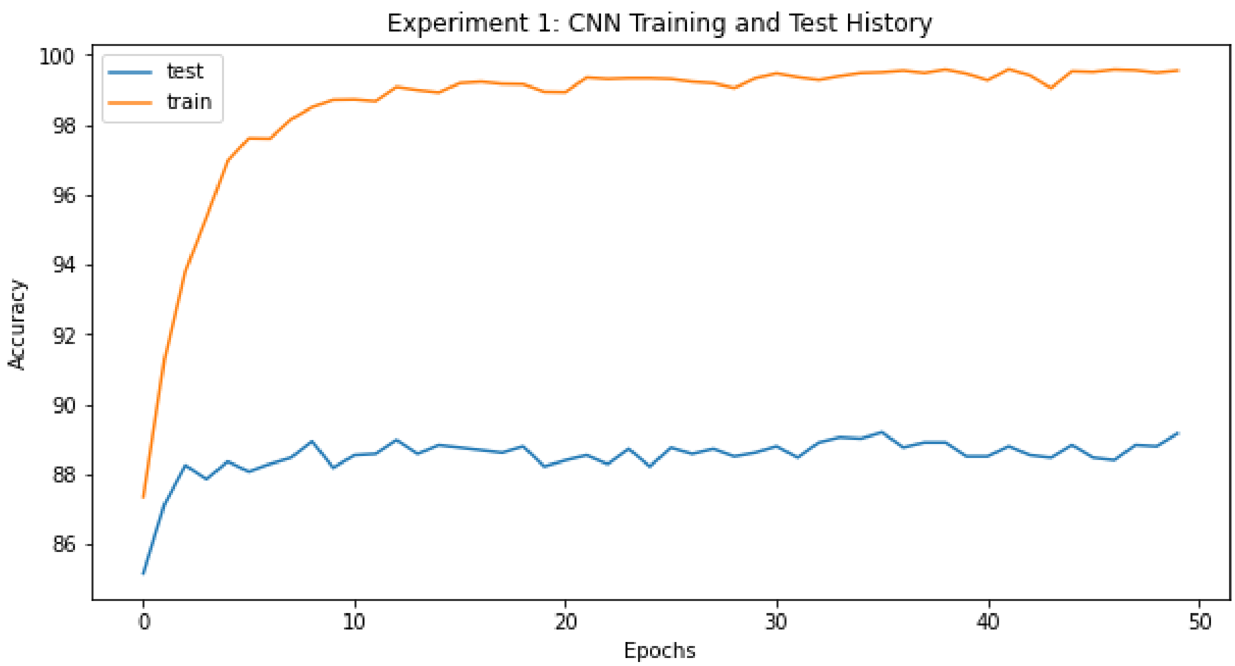
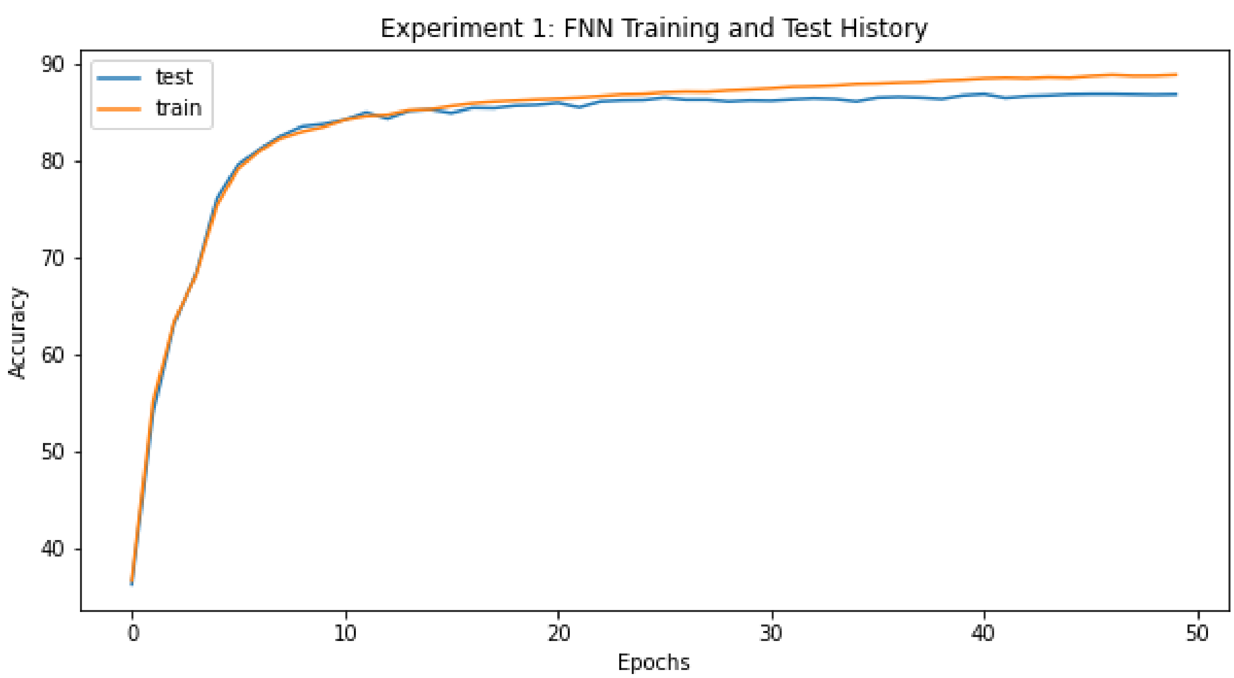
PA3 qa.pdf

Experiment 1: In this experiment, you will compare the performance between CNN and FNN in the image recognition task. Train your CNN and FNN model for a batch size of 32 (default), for 50 epochs (default) and the AdamW optimizer (Link to AdamW) for learning rate of 0.0001. Set the dropout rate=0.0 and weight decay=0.0. Plot and compare the training/testing history of both models. The function name for this experiment will be compare arch().

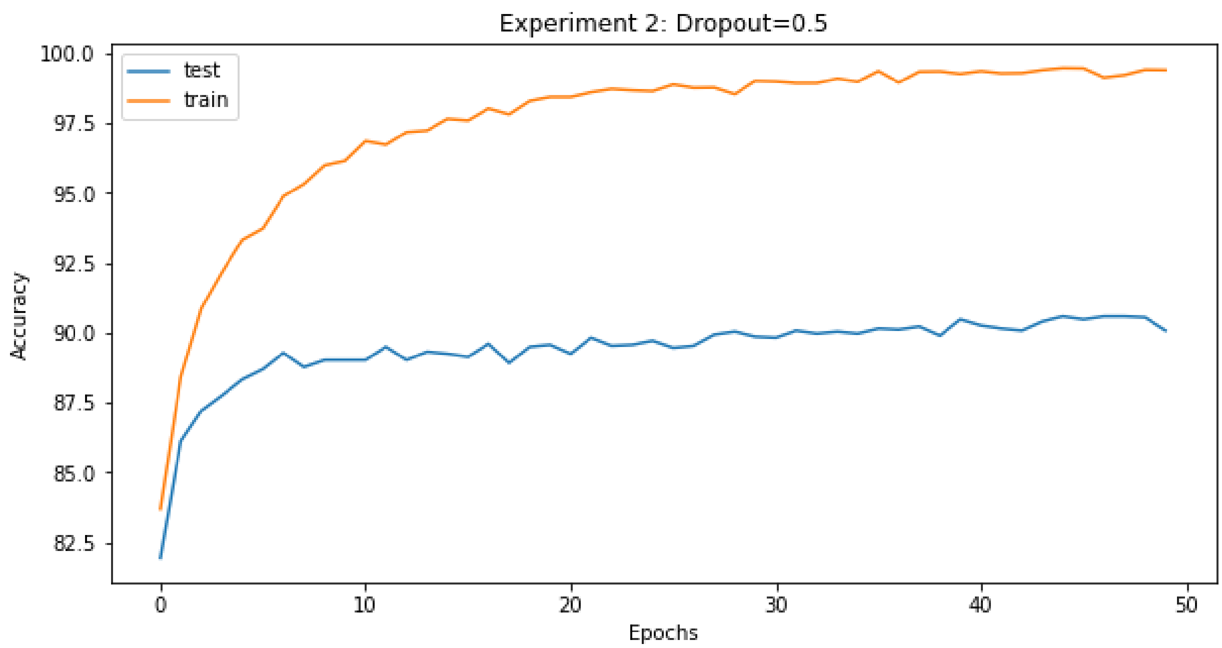
Plots:

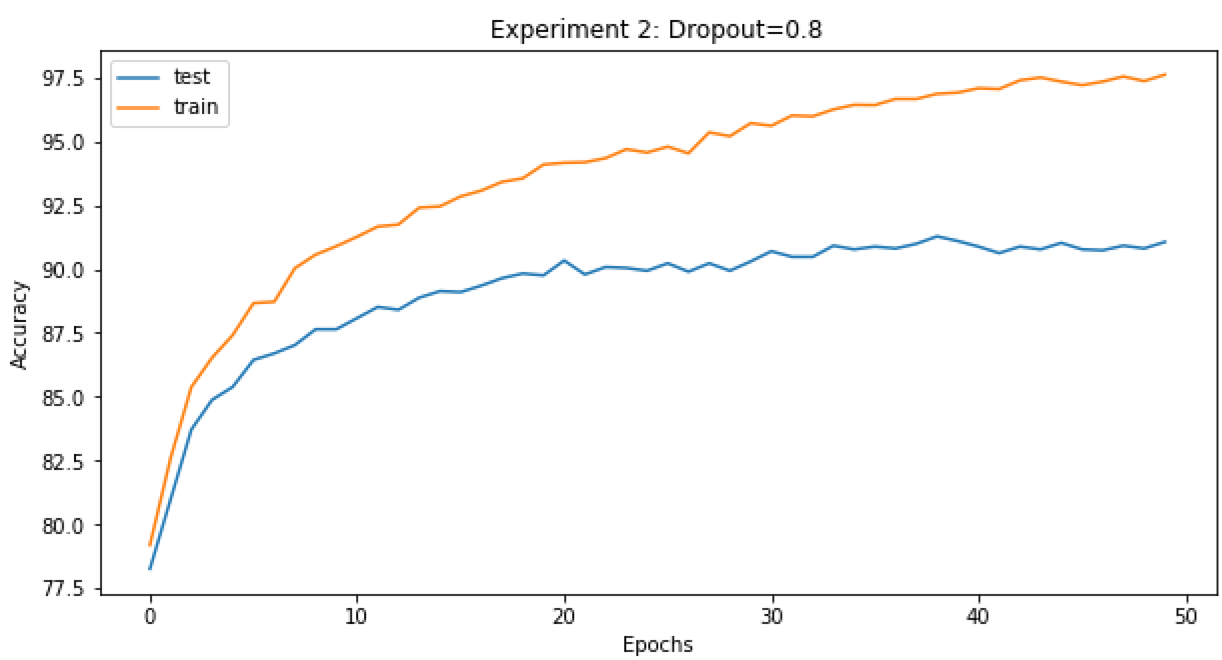


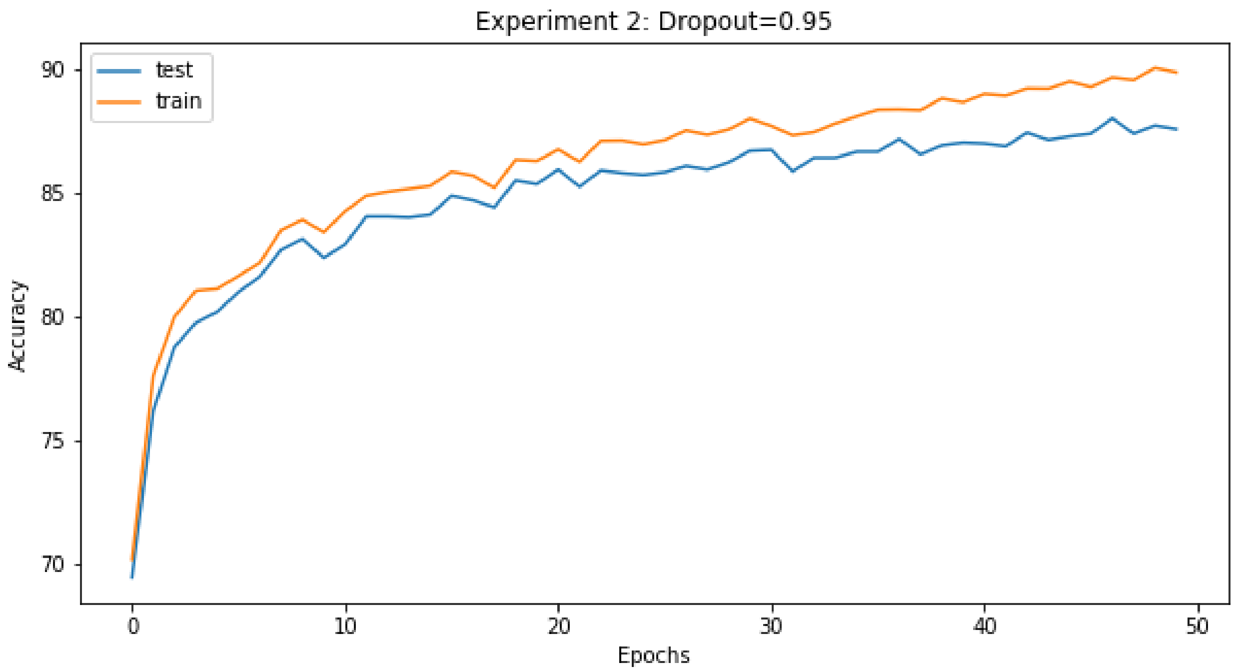


Experiment 2: In this experiment, you will study the effects of dropout rate on your CNN model. Train your CNN model for a batch size of 32 (default), for 50 epochs (default) and the AdamW optimizer (Link to AdamW) for learning rate of 0.0001. Fix your weight decay=0.0 and train 3 different models with dropout rate=0.5,0.8,0.95 respectively. Plot and compare the training/testing history of these three models. Explain the influence of dropout rate to the model's accuracy. The function name for this experiment will be compare dropout().

Plots:





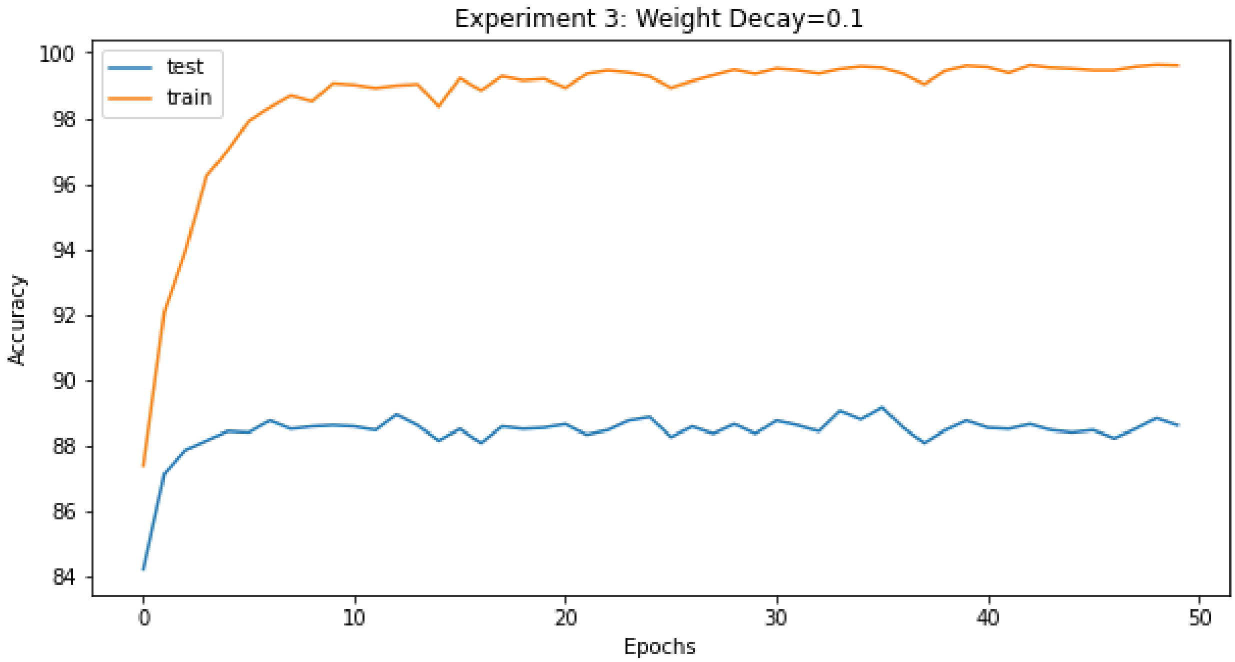


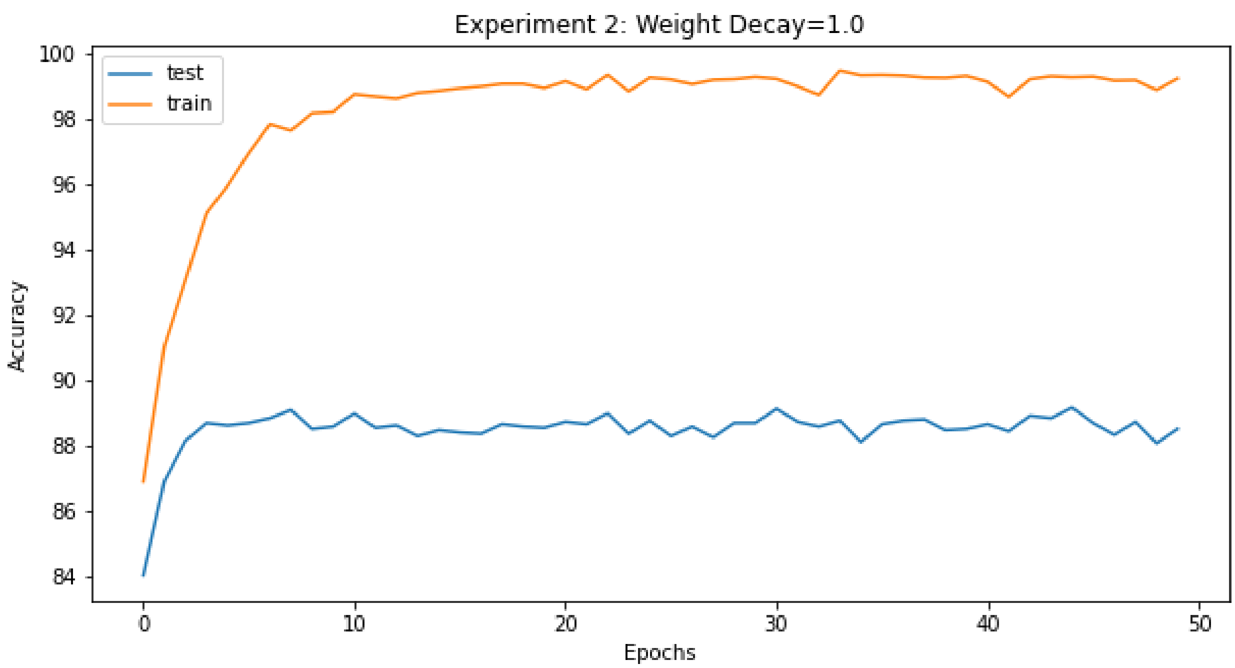
Evaluation:

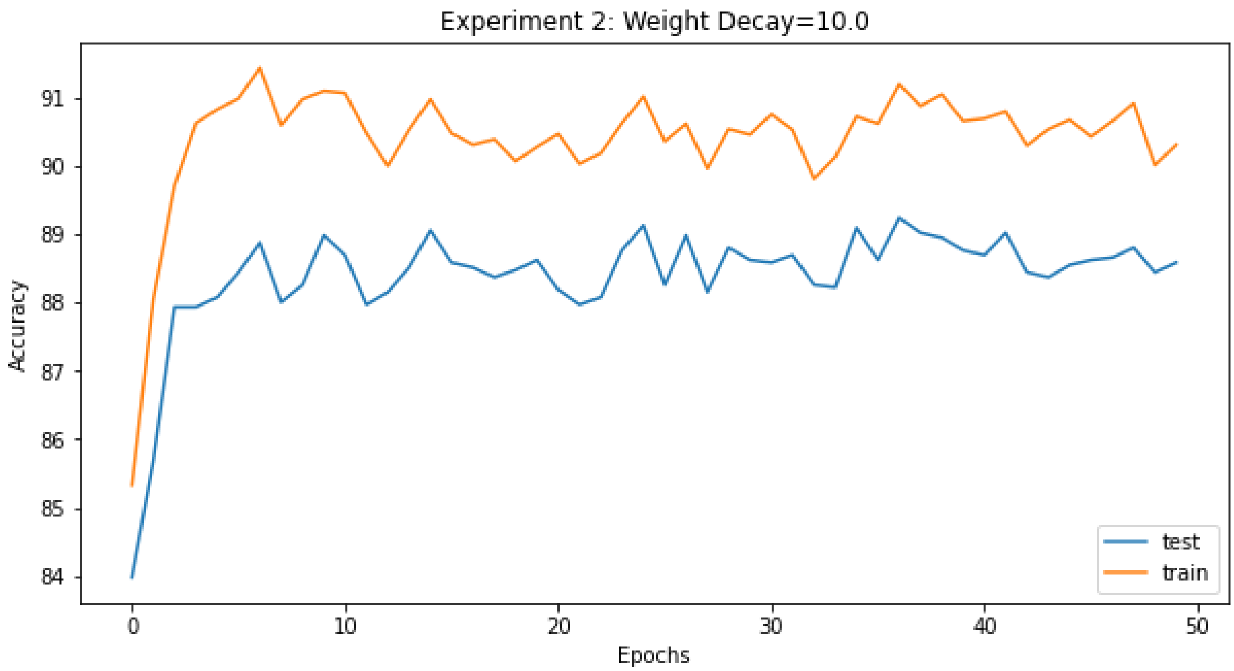
When observing the different plots for 3 different values of dropout, we can tell that the performance of the model when we use dropout of 0.5 and 0.8 are much better than using 0.95. When we used 0.95 the model seemed to be underfitting and this is reflected even on the test data. For values 0.5 and 0.8 the model is performing considerably better. When we compare 0.5 and 0.8, the model seems to be overfitting when dropout value is 0.5 and it performs best at 0.8.

Experiment 3: In this experiment, you will study the effects of weight decay on your CNN model. Train your CNN model for a batch size of 32 (default), for 50 epochs (default) and the AdamW optimizer (Link to AdamW) for learning rate of 0.0001. Fix your drop rate=0.0 and train 3 different models with weight decay=0.1,1.0,10.0 respectively. Plot and compare the training/testing history of these three models. Explain the influence of L2 regularization rate to the model's accuracy. The function name for this experiment will be compare l2().

Plots:


Experiment 3: Weight Decay=0.1
test
train
Accuracy
Epochs
100
98
96
94
92
90
88
86
84
0
10
20
30
40
50


Experiment 2: Weight Decay=1.0
test
train
Accuracy
Epochs
100
98
96
94
92
90
88
86
84
0
10
20
30
40
50



Evaluation:

Weight decay is a regularization technique by adding a small penalty, usually the L2 norm of the weights (the sum of squared parameters, or weights of a model multiplied by some coefficient) is added into the loss function. In all three cases, we see that the model has better performance on the training dataset than the test dataset. When the weight decay is assigned the values of 0.1 and 1.0, the model seems to be overfitted. On increasing the weight decay to 10.0, the overfitting is reduced as expected.