Machine Learning Winter 2024

Programming Assignment #1

Greg Witt

Experiment 1:

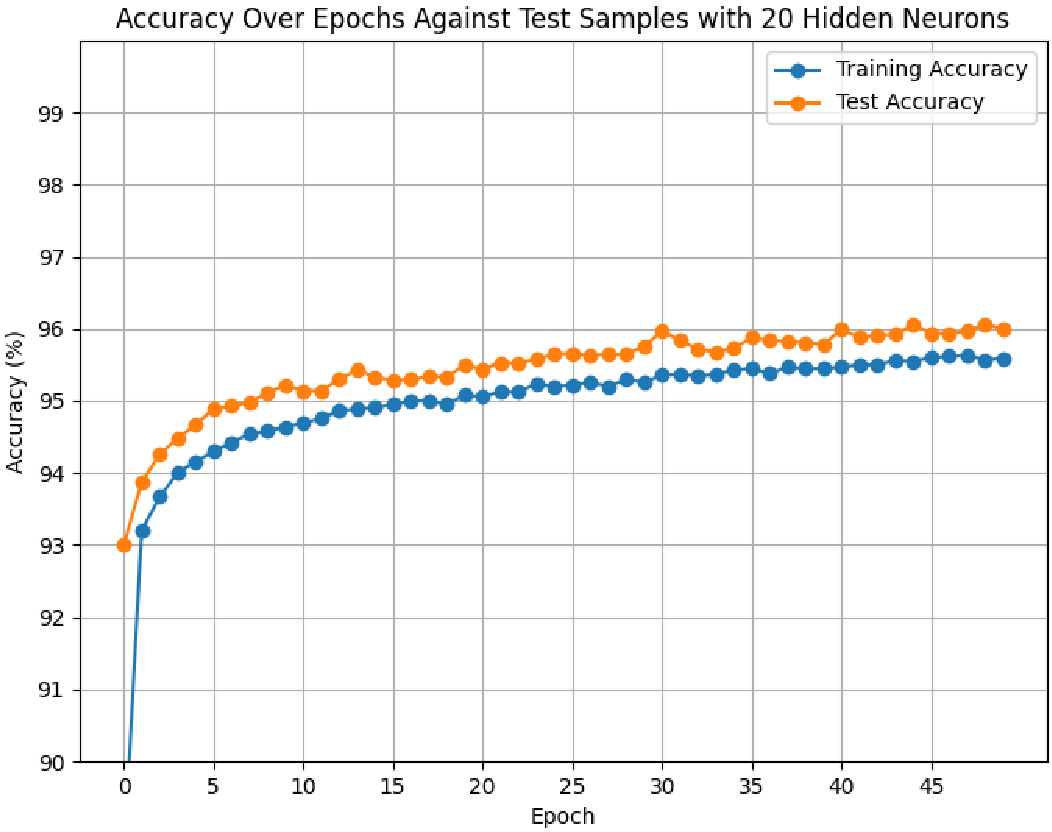
A two layer neural network is constructed with various adjustments made to the hidden layer neurons.

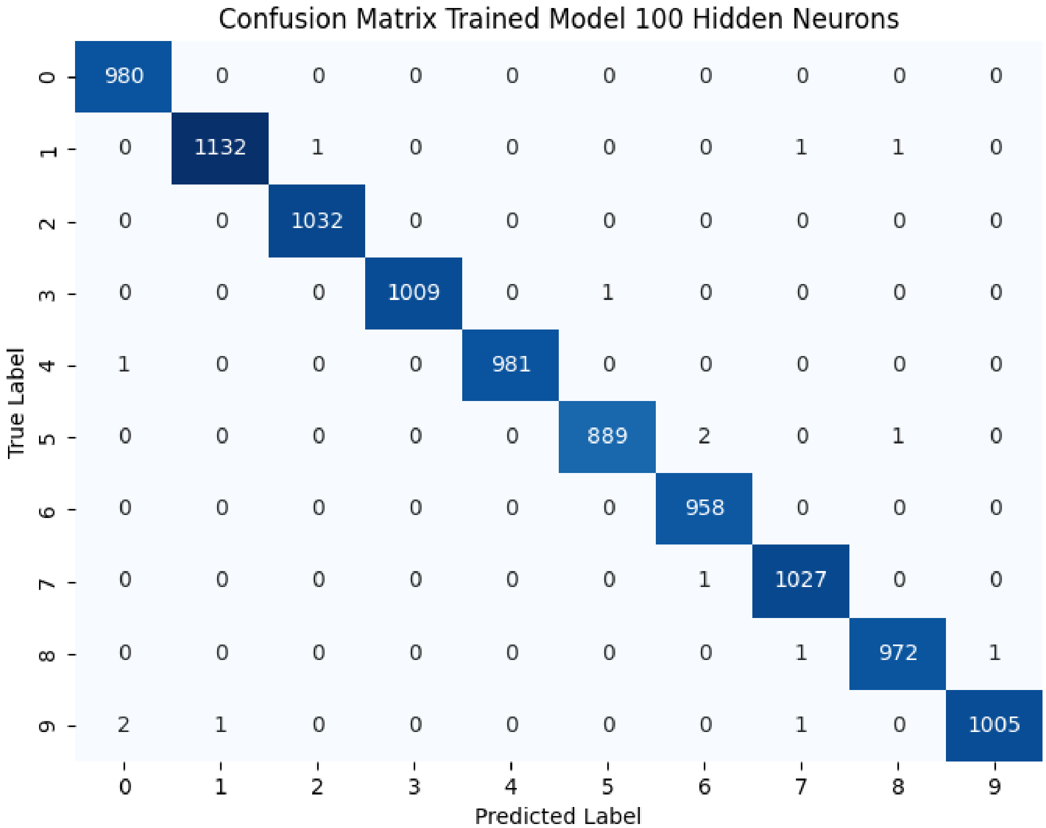
The adjustments were made directly to each model with respect to the number of hidden layers within each model with following numbers 20, 50, 100 hidden units.

The number of hidden neuron units directly affects the resulting accuracy rate and the overall potential accuracy through 50 epochs. From the resulting graphs, it appears 20, 50 and 100 were all impended well enough through backpropagation to achieve accurate results above 90% within the first epoch of against the training and test data sets. The number of neurons directly impacts the rate at which the models can accurately learn to predict the input digits, however this it's intuitive the higher the hidden neurons the more precise the gradient and loss can be reduced. For example, the 100 hidden neuron layer was able to achieve close to 98% accuracy within 10 epochs against both testing and training sets. With 20 hidden neurons, the model was barely able to achieve an accuracy of 96% within 50 epochs. This was not the case however with the 50 hidden neurons model. The 50 hidden neurons model appeared to have learned the training set and converged with the test set almost within 5 epochs. This was the case for the remainder of the experiment. The results from 50 hidden neurons were very close and appear to be overfit against the test and training sets in epochs 15-26 but the presence of new shuffled results from the test set broke this trend. The model continues to appear to re-converge later. The 50 hidden neuron model specifically often confused 4 and 9 and 3s for 7s. These confusion matrix counter were higher in 50 hidden neuron model tests compared to 100 neuron test. This result indicates that having more hidden layers within the model assigns higher feature selection for these numbers. The 100 hidden layer appears to also converge around higher epochs in both test and training samples. However, the neural net with 100 hidden neurons has a much higher accuracy gain over shorter iterations, indicating better performance at the cost of complexity. The change in accuracy is almost identical from training data to test data from epoch 20-50 with this model. It can be noted that the training examples and test examples from 50 hidden layers more closely resemble the performance metric of a single layer perceptron learner from the first experiment in Homework #1. compared to the precision of the 20, and 100 hidden layer model results the 50 hidden model seems to drastically converge and remain slow but constant throughout the entire experiment. The 20 hidden neuron model may have a lower accuracy but appears to be less likely to overfitting compared to the higher numbers of hidden layer experiments.

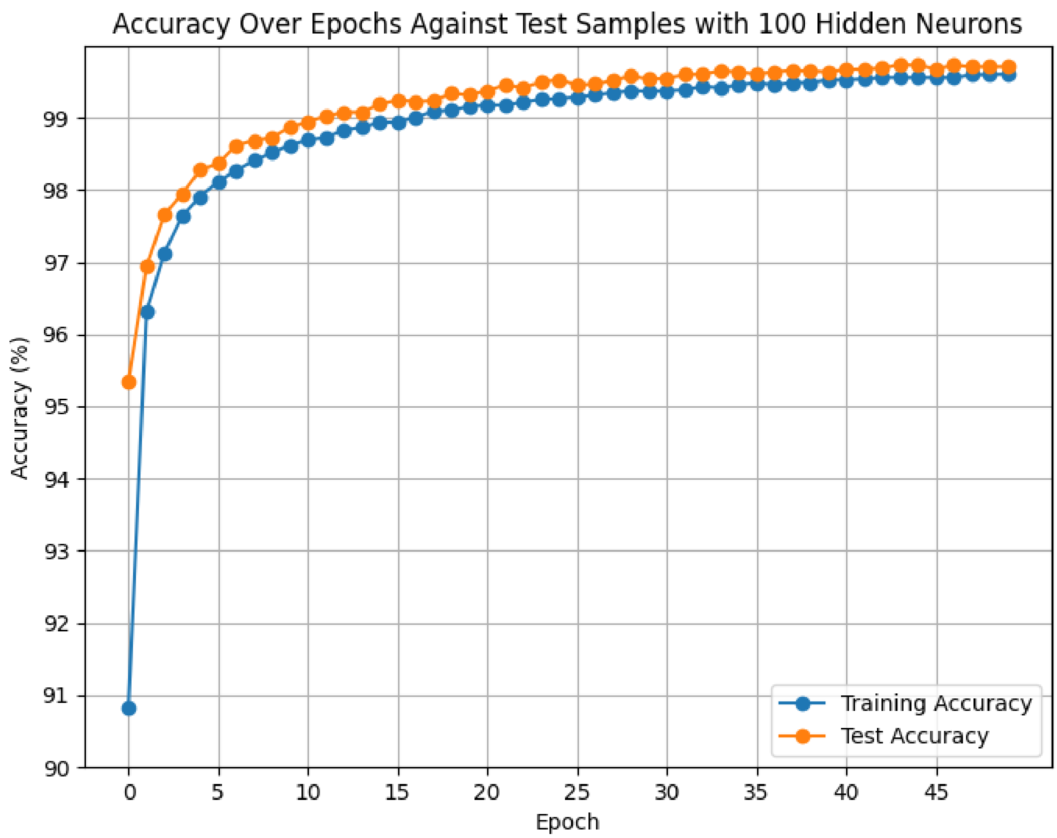
Given all the performances I would choose a model with 20 hidden neurons for this experiment to have performed the best overall. It didn't overfit or converge too quickly until the end of the experiment making it appear that although its accuracy is lower it has more evidence to suggest that its accuracy might improve over more iterations with out overfitting to the training or test data.

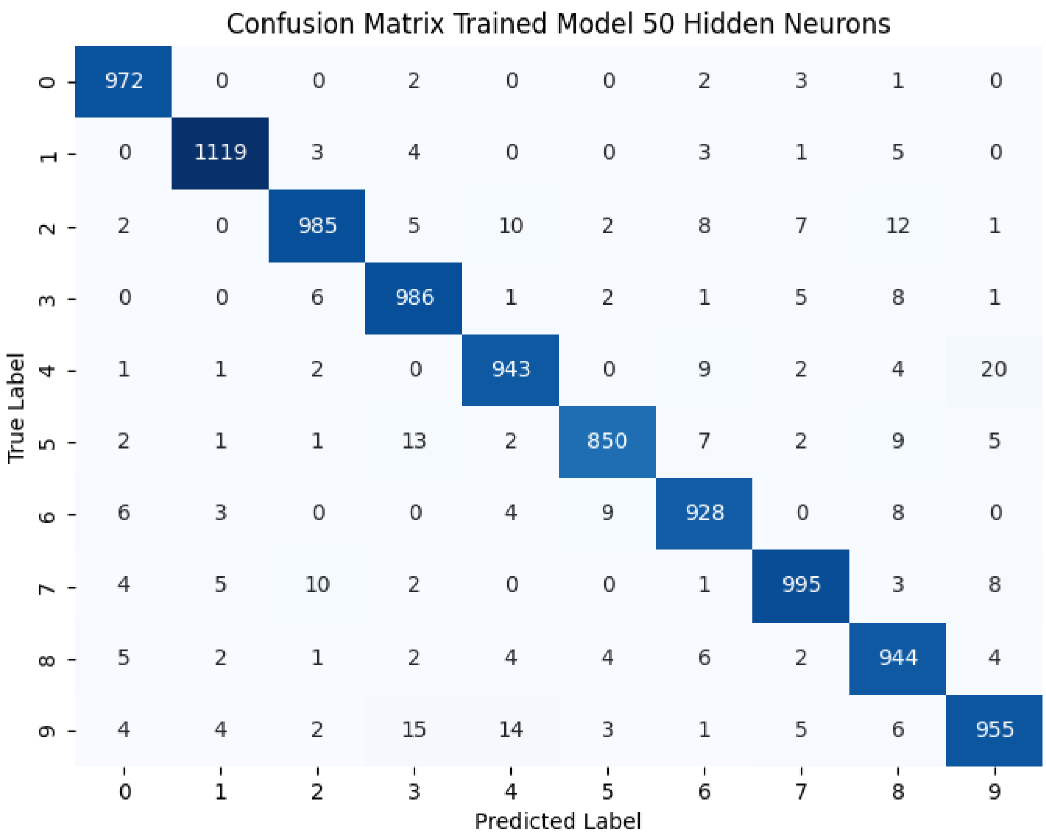
The Results from Experiment 1 are featured below:

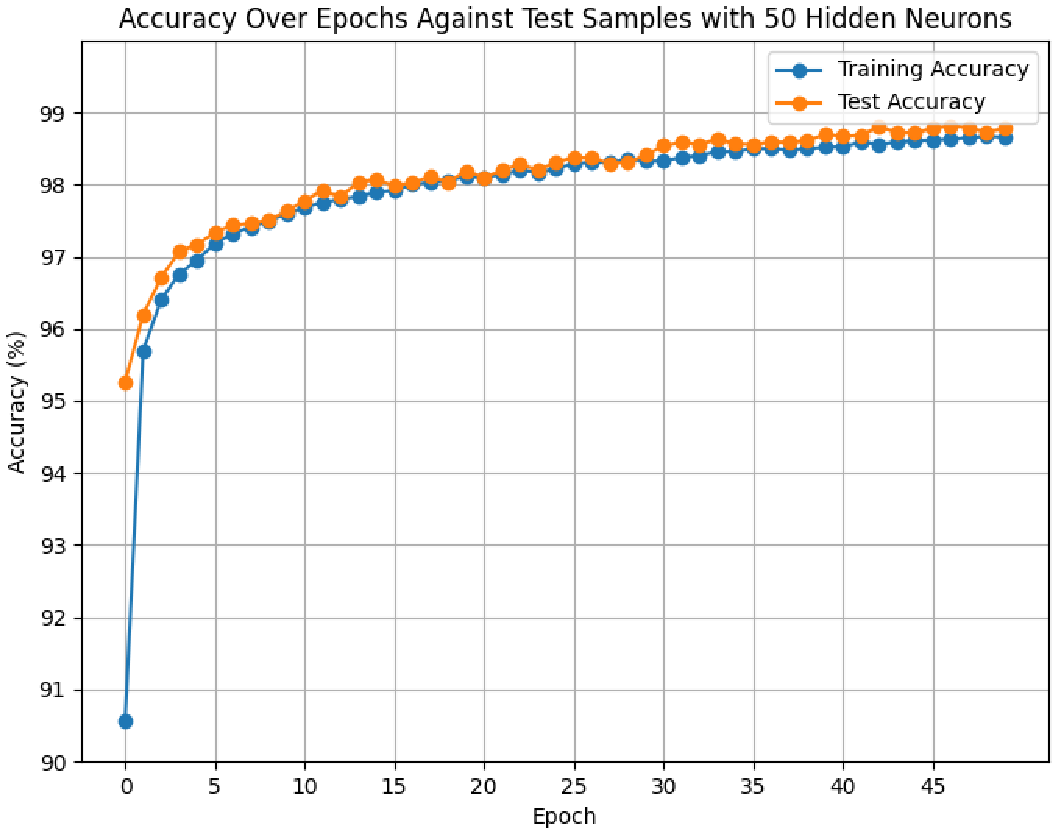

Accuracy Over Epochs Against Test Samples with 20 Hidden Neurons
Training Accuracy
Test Accuracy
Accuracy (%)
99
98
97
96
95
94
93
92
91
90
0
5
10
15
20
25
30
35
40
45
Epoch

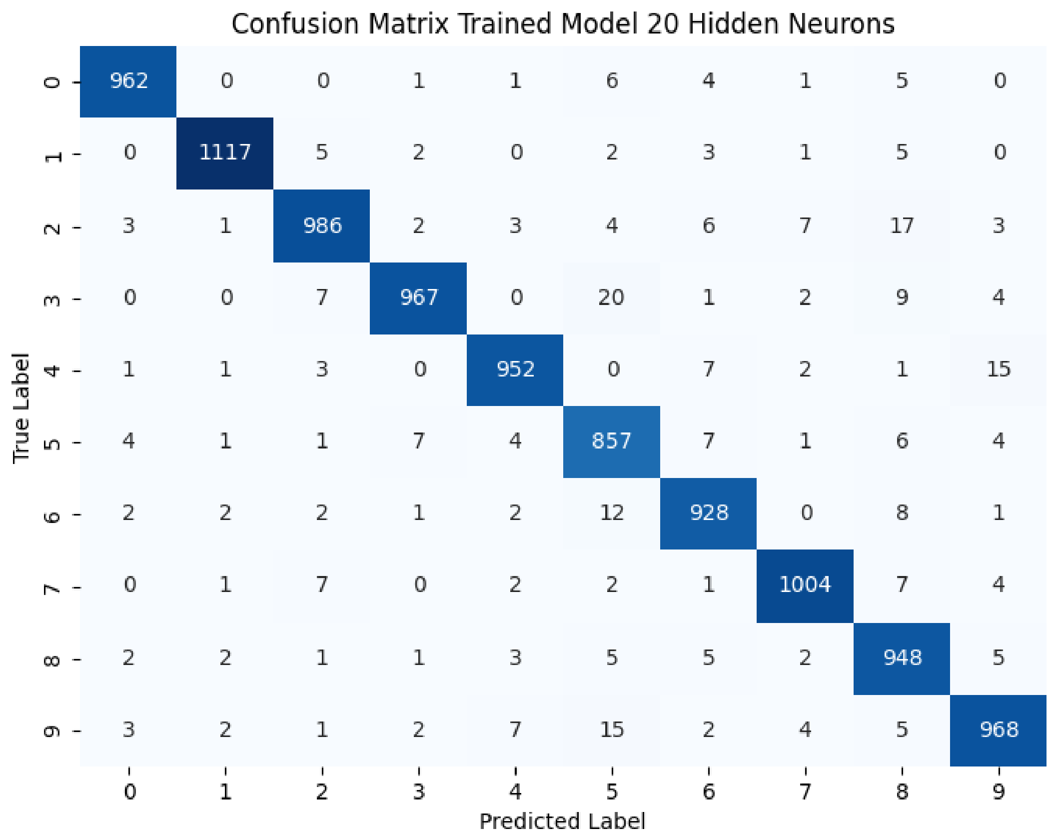


Confusion Matrix Trained Model 100 Hidden Neurons
True Label
Predicted Label
0 1 2 3 4 5 6 7 8 9
980 0 0 0 0 0 0 0 0 0
0 1132 1 0 0 0 0 1 1 0
0 0 1032 0 0 0 0 0 0 0
0 0 0 1009 0 1 0 0 0 0
1 0 0 0 981 0 0 0 0 0
0 0 0 0 0 889 2 0 1 0
0 0 0 0 0 0 958 0 0 0
0 0 0 0 0 0 1 1027 0 0
0 0 0 0 0 0 0 1 972 1
2 1 0 0 0 0 0 1 0 1005


Accuracy Over Epochs Against Test Samples with 100 Hidden Neurons
Accuracy (%)
Epoch
Training Accuracy
Test Accuracy
90
91
92
93
94
95
96
97
98
99
0
5
10
15
20
25
30
35
40
45

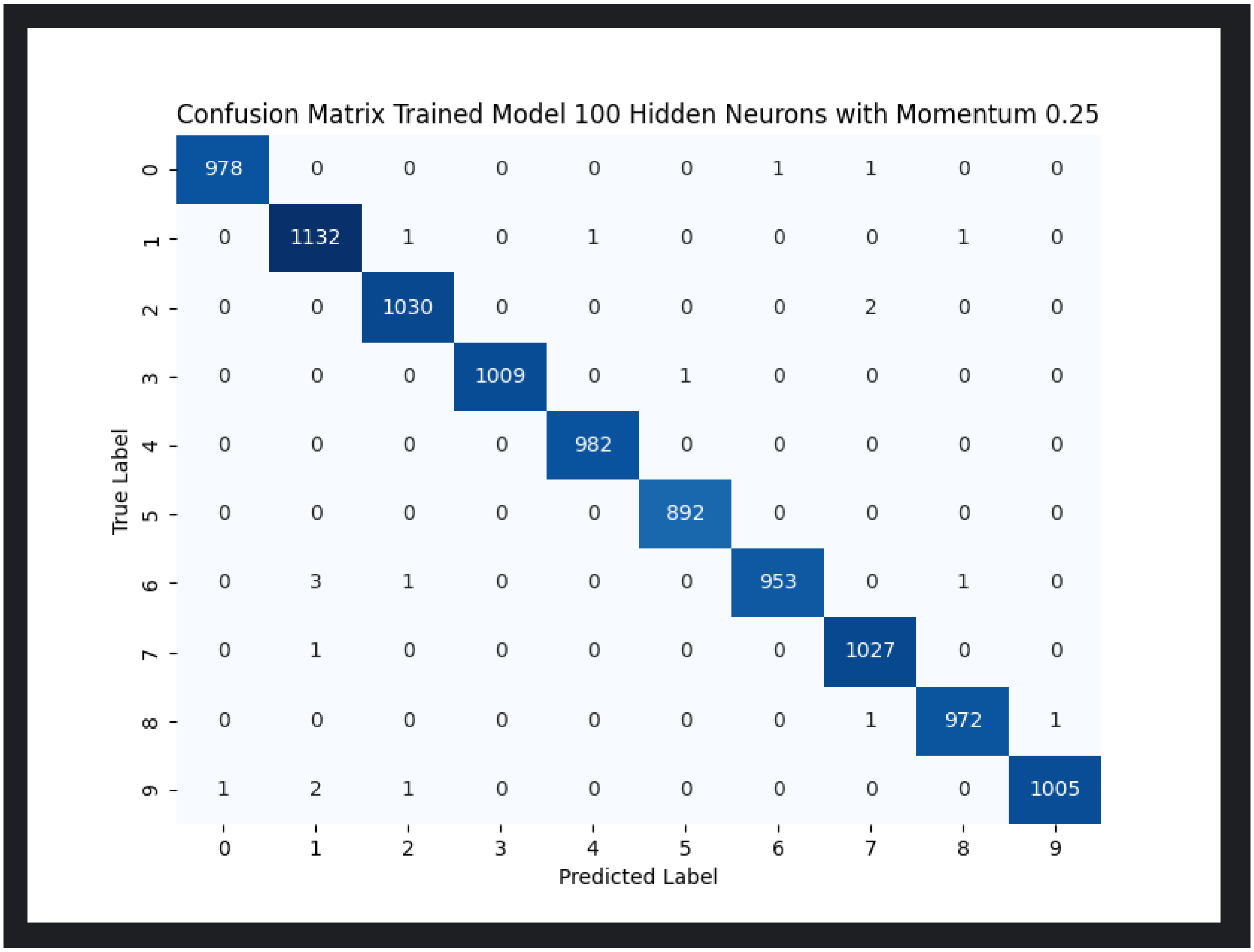

Confusion Matrix Trained Model 50 Hidden Neurons
True Label
Predicted Label
0 1 2 3 4 5 6 7 8 9
0 972 0 0 2 0 0 2 3 1 0
1 0 1119 3 4 0 0 3 1 5 0
2 2 0 985 5 10 2 8 7 12 1
3 0 0 6 986 1 2 1 5 8 1
4 1 1 2 0 943 0 9 2 4 20
5 2 1 1 13 2 850 7 2 9 5
6 6 3 0 0 4 9 928 0 8 0
7 4 5 10 2 0 0 1 995 3 8
8 5 2 1 2 4 4 6 2 944 4
9 4 4 2 15 14 3 1 5 6 955

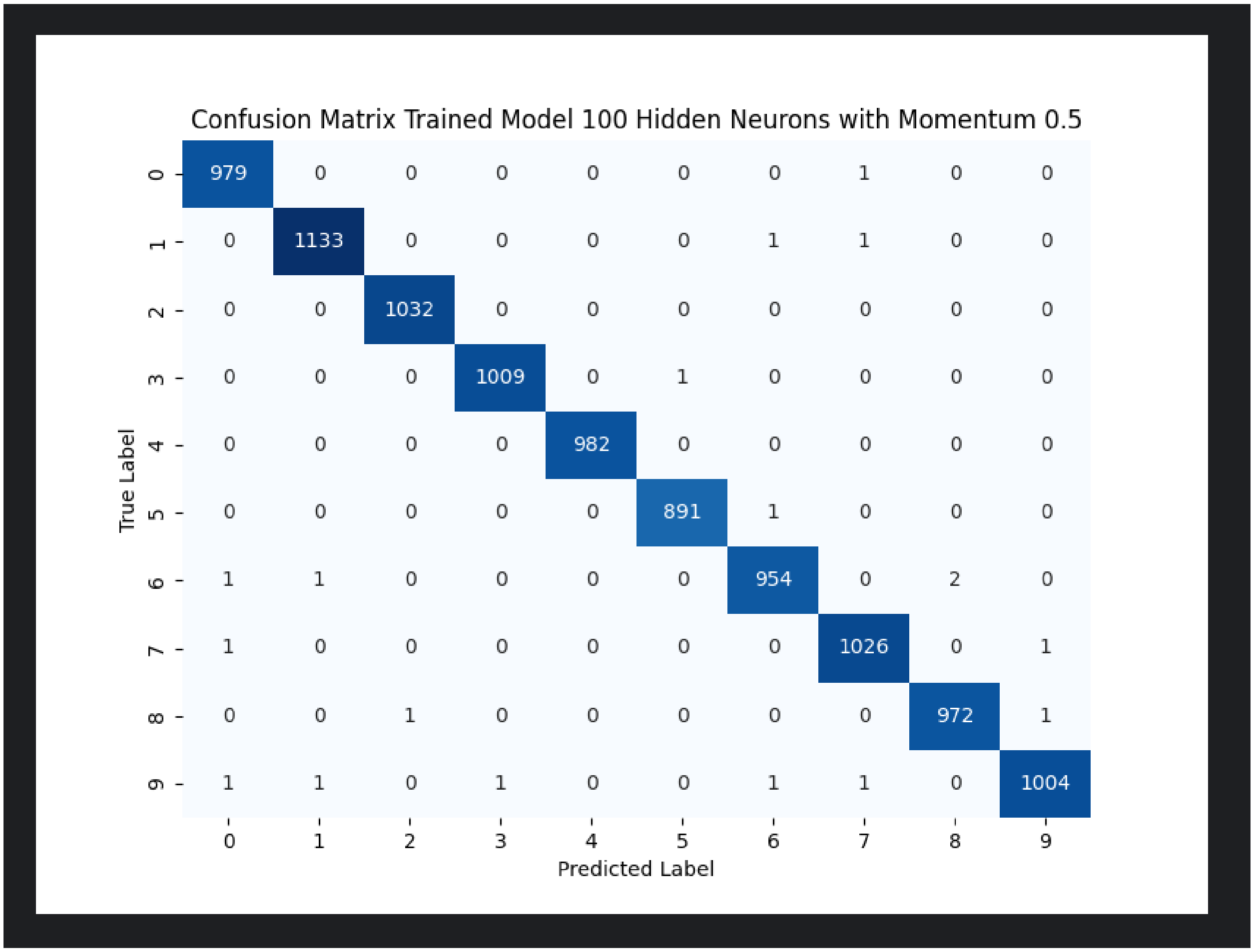

Accuracy Over Epochs Against Test Samples with 50 Hidden Neurons
Training Accuracy
Test Accuracy
Accuracy (%)
Epoch
90
91
92
93
94
95
96
97
98
99
0
5
10
15
20
25
30
35
40
45

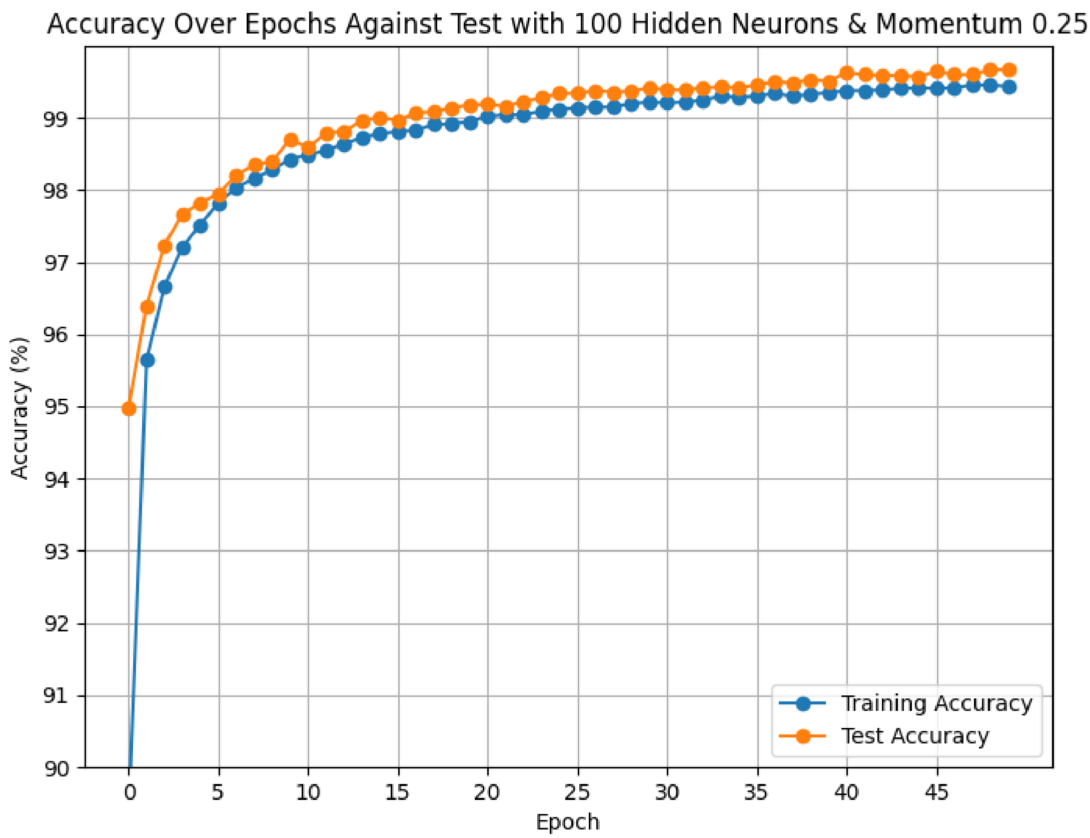


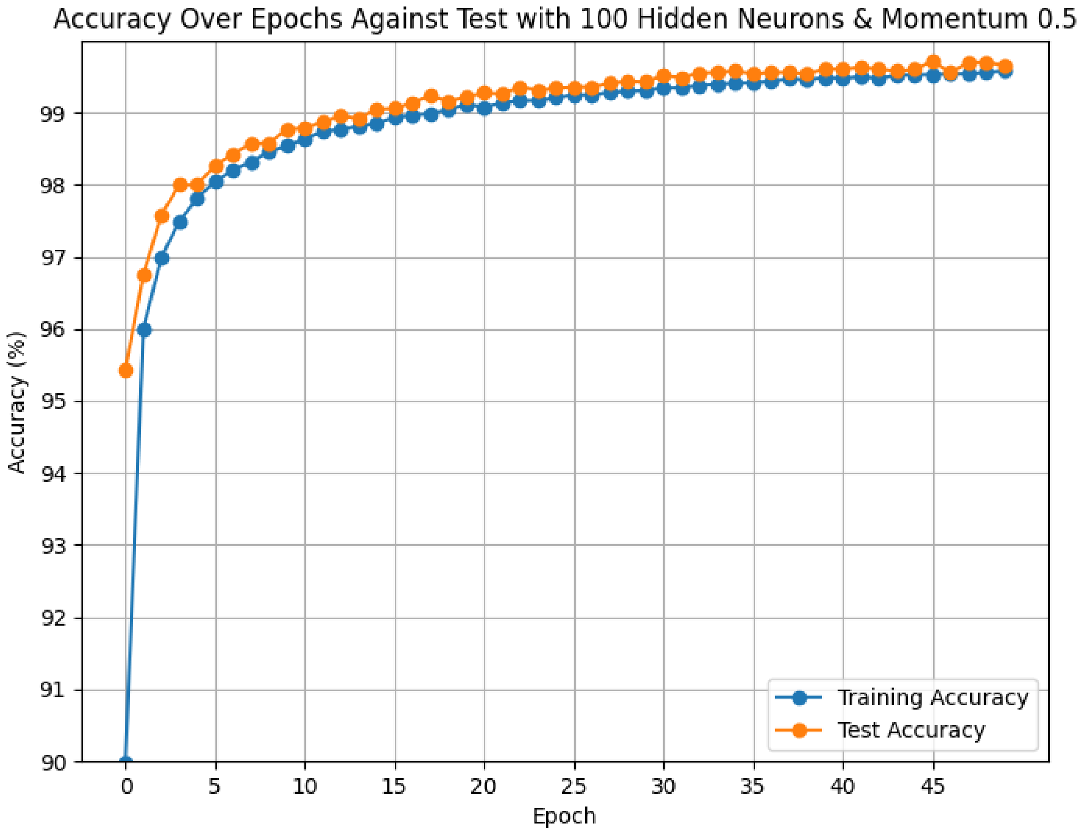
Experiment 2:

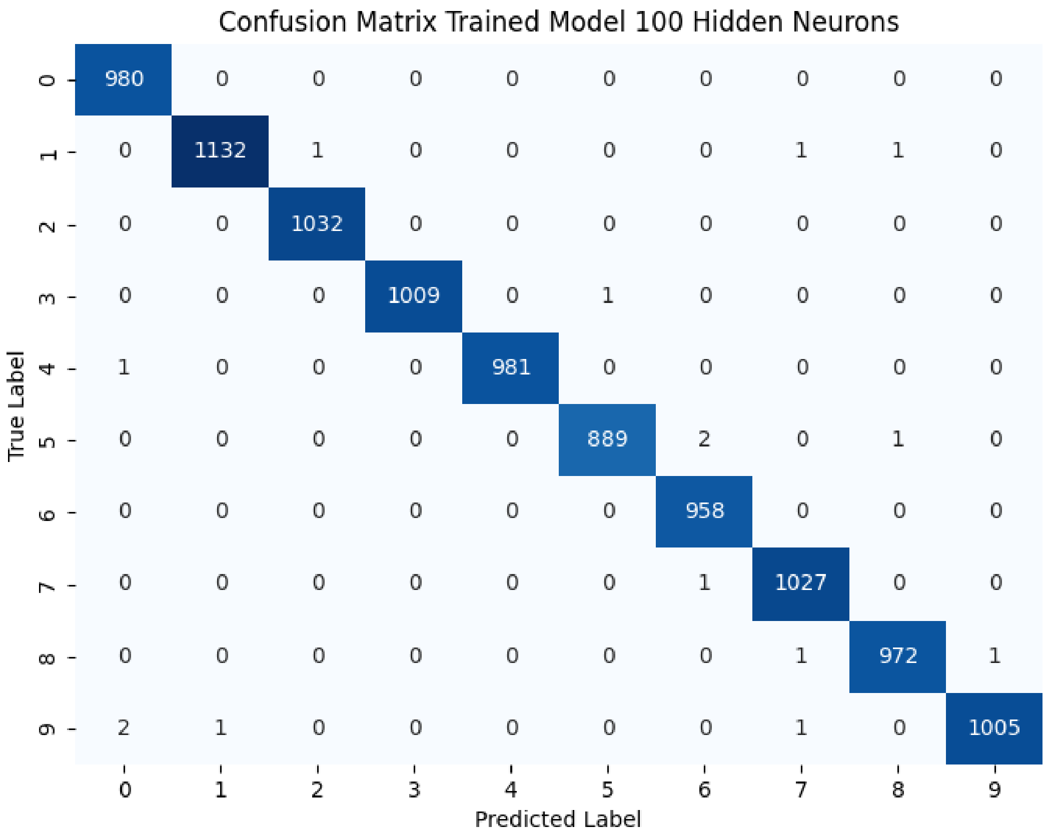
When momentum was adjusted and compared to the control result of 100 hidden neurons and a momentum set to .09 the model had a base line performance above 90% within the first 15 epochs. After modifying the training data percentages being fed to the model for training, as expected the momentum term kept the convergence less shaky and maintained a constant climb converging with the test results and maintaining an accuracy prediction of above 99% after epoch 15. The behavior was slightly different in the experiment with a momentum rate of 0.25 the 0^{th} epoch had a performance well below 90% then accelerated to 95.8% suggesting the momentum however small pushed the performance out of a possible local min by increasing accuracy by over 5% in the preceding epoch. The performance there after was slightly worse. It took 20 epochs to reach 99% and for the remainder of the training and testing the results didn't match the performance of the first control experiment with 0.9 for a momentum value. The resulting 0.25 model's peak performance recorded in epoch 50 was roughly 99.3%-99.4% however the small momentum appeared to help prevent possible over fitting. There were less epochs where the performance on the training matched the performance on the testing. Indicating the model closely bordered the same results but did not converge and remain the same percentage. Compared to the 0.5 momentum model, which within 10 epochs was matching the exact accuracy of the training and test data. The only results with a more separate margin in accuracy compared to the test accuracy was the 0.25 rate of momentum model. This rate appears to not converge as quickly but also doesn't fully exceed or match the test results in later epochs of 30-50 and remains distant in the training data in epochs 20 and below. Which was different for the .50 momentum model and the 0.9 control momentum model which performed better overall. With respect to the proximity to the testing performance accuracies, the 0.9 and 0.5 model's momentums might have fewer oscillations, but they might be performing too well and would suffer if given a new unseen set of data to test with.

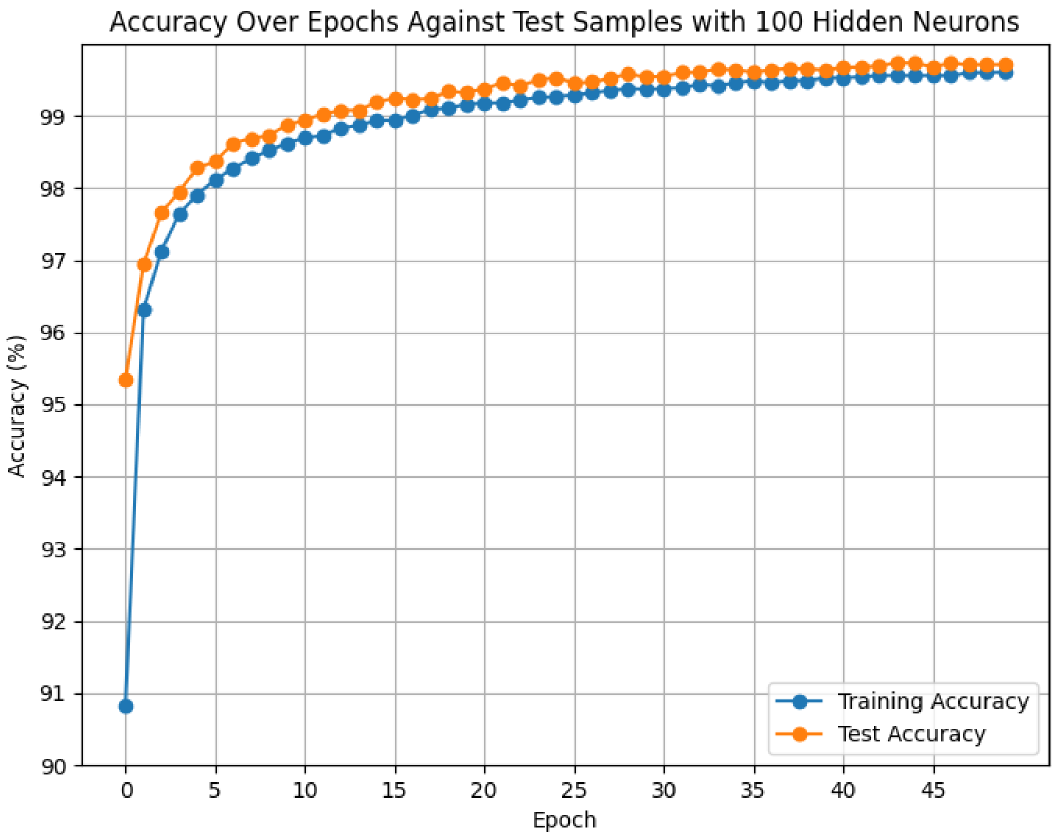

Confusion Matrix Trained Model 100 Hidden Neurons with Momentum 0.25
True Label
Predicted Label


Confusion Matrix Trained Model 100 Hidden Neurons with Momentum 0.5
True Label
Predicted Label
0 1 2 3 4 5 6 7 8 9
979 0 0 0 0 0 0 1 0 0
0 1133 0 0 0 0 1 1 0 0
0 0 1032 0 0 0 0 0 0 0
0 0 0 1009 0 1 0 0 0 0
0 0 0 0 982 0 0 0 0 0
0 0 0 0 0 891 1 0 0 0
1 1 0 0 0 0 954 0 2 0
1 0 0 0 0 0 0 1026 0 1
0 0 1 0 0 0 0 0 972 1
1 1 0 1 0 0 1 1 0 1004


Accuracy Over Epochs Against Test with 100 Hidden Neurons & Momentum 0.25
Accuracy (%)
Epoch
Training Accuracy
Test Accuracy


Accuracy Over Epochs Against Test with 100 Hidden Neurons & Momentum 0.5
Accuracy (%)
Epoch
Training Accuracy
Test Accuracy
90
91
92
93
94
95
96
97
98
99
0
5
10
15
20
25
30
35
40
45

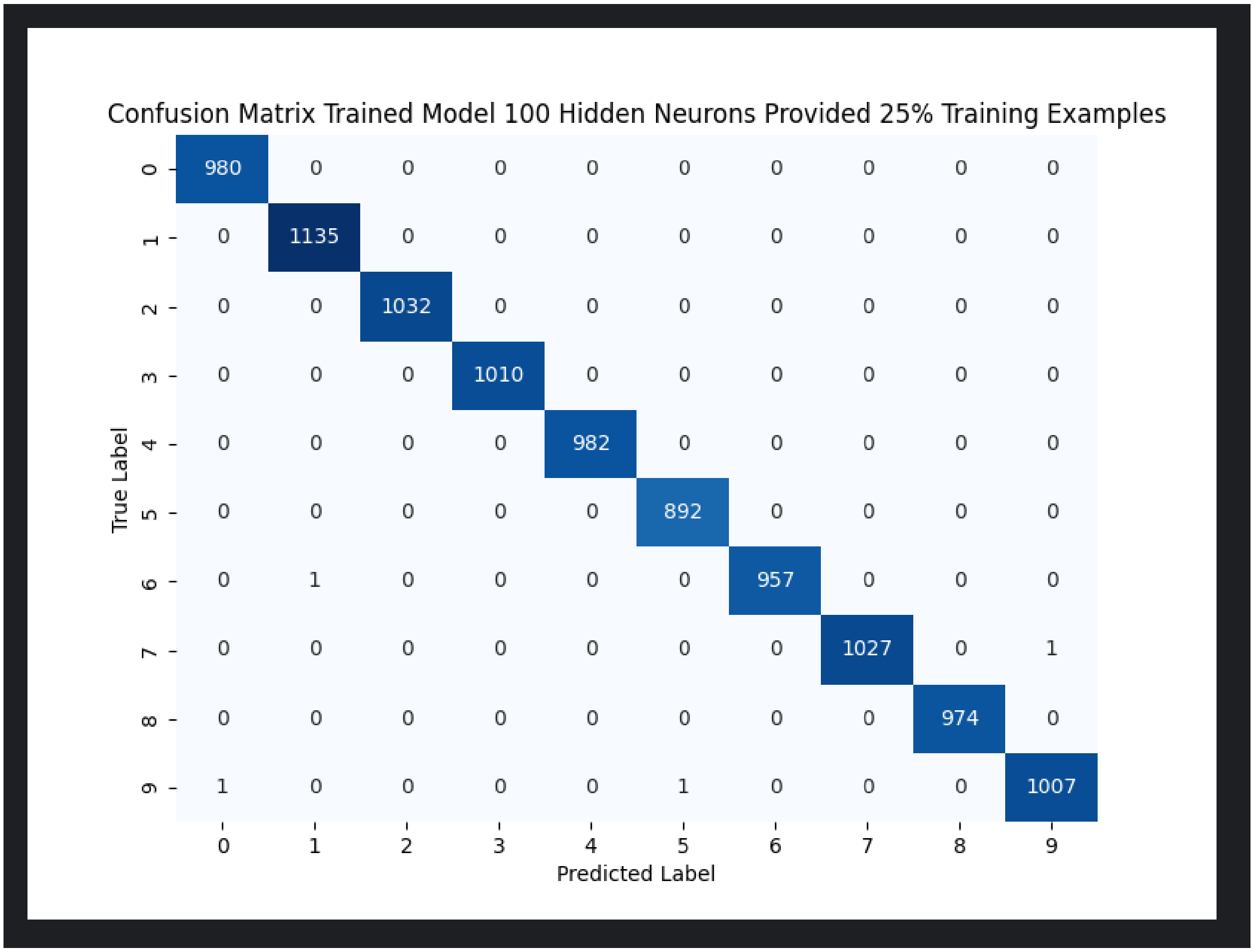

Confusion Matrix Trained Model 100 Hidden Neurons
True Label
Predicted Label
0 1 2 3 4 5 6 7 8 9
980 0 0 0 0 0 0 0 0 0
0 1132 1 0 0 0 0 1 1 0
0 0 1032 0 0 0 0 0 0 0
0 0 0 1009 0 1 0 0 0 0
1 0 0 0 981 0 0 0 0 0
0 0 0 0 0 889 2 0 1 0
0 0 0 0 0 0 958 0 0 0
0 0 0 0 0 0 1 1027 0 0
0 0 0 0 0 0 0 1 972 1
2 1 0 0 0 0 0 1 0 1005

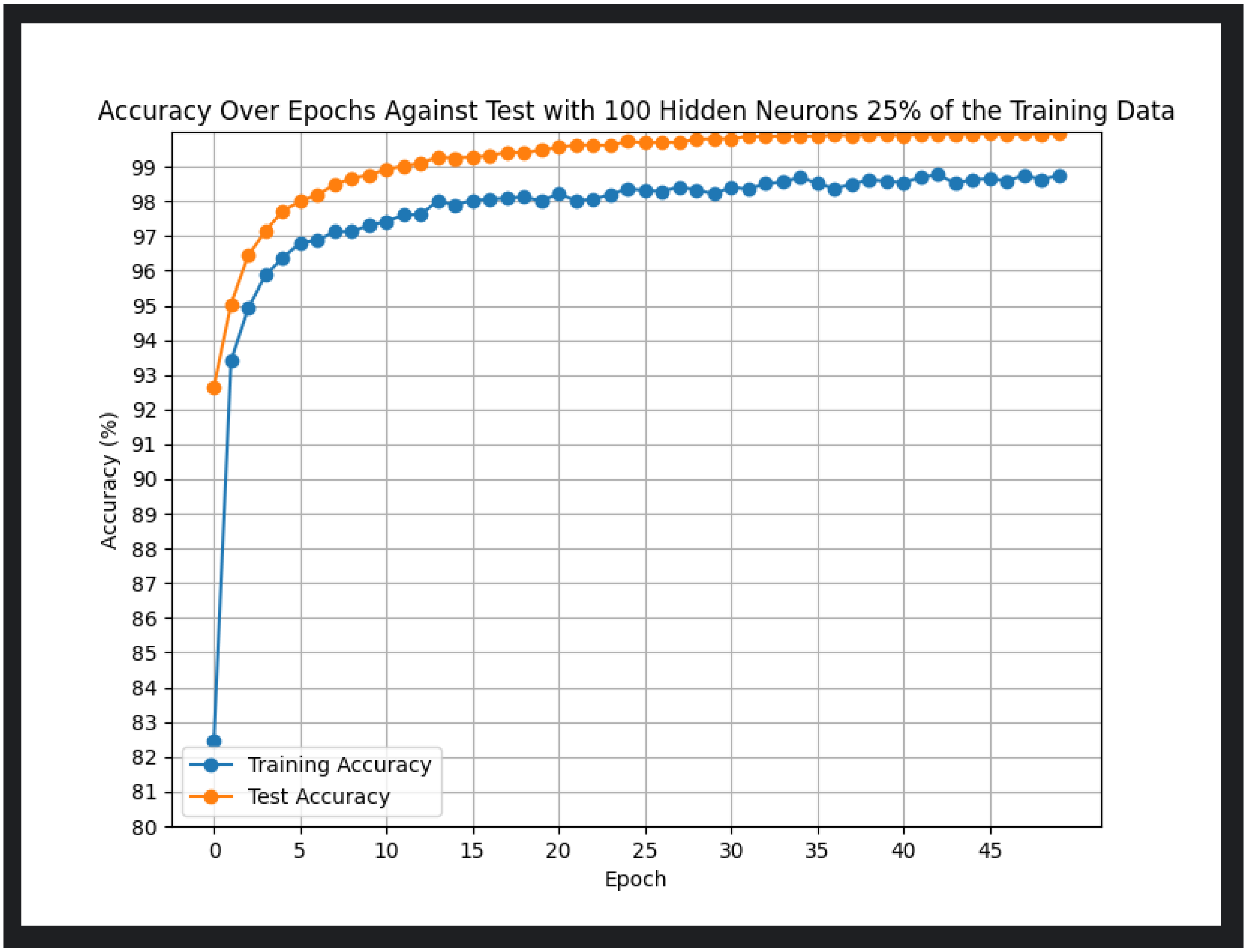


Experiment 3:

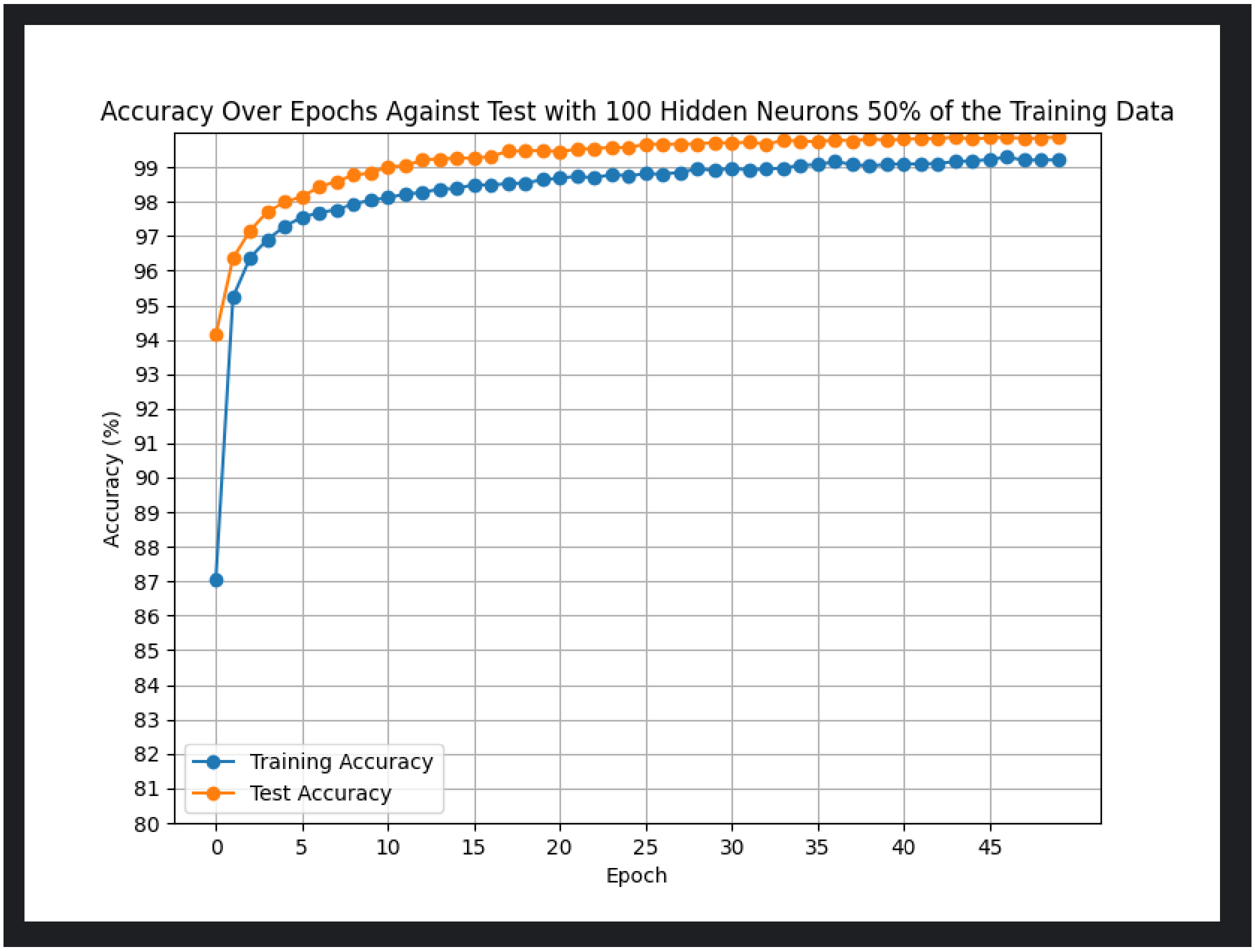
Varying the number of training data resulted in much higher accuracy in the testing set amongst the 50% and 25% trained models. Overall, the performance was impressive in both trained models. Which were able to achieve 99% accuracy against the test set within 10 epochs. The resulting accuracies on the training data were significantly lower never fully reaching 99% in the 25% trained model, and with the 50% model it just barely surpasses 99% within the 50 epochs. The number epochs needed to converge with the 25% seems to be higher. There were still slight oscillations in the training results after epoch 25 but the same cannot be said for the testing accuracy of the same model. Within 25 epochs the 25% model appears to remain at the asymptotic limit below 100. If the test data had new numbers added for this experiment it could be probable the model would maintain this result, with little oscillations. The model with 50% of the training set has a higher rate of learning accuracy for both the training and test set and converges quickly around 15 constantly growing and climbing. But this experiment still shows no overfitting for the test data in either model as their results are drastically separated. This approach is intuitively optimal as the main concept to help to prevent overfitting is to prevent patterns in the data from being learned. With limited training data being expressed to the model it will have more variety to train against in the test set. This approach has increased the accuracy against unseen test data.


Confusion Matrix Trained Model 100 Hidden Neurons Provided 50% Training Examples
True Label
Predicted Label
0 1 2 3 4 5 6 7 8 9
979 0 0 0 0 0 0 1 0 0
0 1133 1 0 0 0 0 0 1 0
0 1 1030 0 0 0 0 0 1 0
0 0 0 1010 0 0 0 0 0 0
0 0 0 0 981 0 1 0 0 0
0 0 0 0 0 892 0 0 0 0
0 1 0 0 0 0 957 0 0 0
0 0 0 0 0 0 0 1028 0 0
0 0 0 0 0 0 0 0 974 0
0 0 0 0 0 0 0 0 0 1009


Confusion Matrix Trained Model 100 Hidden Neurons Provided 25% Training Examples
True Label
Predicted Label
0 1 2 3 4 5 6 7 8 9
980 0 0 0 0 0 0 0 0 0
0 1135 0 0 0 0 0 0 0 0
0 0 1032 0 0 0 0 0 0 0
0 0 0 1010 0 0 0 0 0 0
0 0 0 0 982 0 0 0 0 0
0 0 0 0 0 892 0 0 0 0
0 1 0 0 0 0 957 0 0 0
0 0 0 0 0 0 0 1027 0 1
0 0 0 0 0 0 0 0 974 0
1 0 0 0 0 1 0 0 0 1007



Accuracy Over Epochs Against Test with 100 Hidden Neurons 25% of the Training Data
Accuracy (%)
Epoch
Training Accuracy
Test Accuracy
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
0
5
10
15
20
25
30
35
40
45


Accuracy Over Epochs Against Test with 100 Hidden Neurons 50% of the Training Data
Accuracy (%)
Epoch
Training Accuracy
Test Accuracy
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
0
5
10
15
20
25
30
35
40
45