



CSE-472

*(Machine Learning
Sessional)*

Report on Assignment-3

Submitted by:

Tanjeem Azwad Zaman

Roll: 1805006

Dept. of CSE, BUET

Date of Submission:

03.01.2024

1. Models (training and validation)

1.1 Model 1 (Best):

1.1.1 Architecture

```
MyModel = FNNmodel()
L1 = DenseLayer(784, 1024)
L1.init_weights_he()
MyModel.addLayer(L1)
# CHANGE
L2 = ReLUActivationLayer()
MyModel.addLayer(L2)

L2_5 = DropoutLayer(0.3)
MyModel.addLayer(L2_5)

L3 = DenseLayer(1024, 26)
L3.init_weights_he()

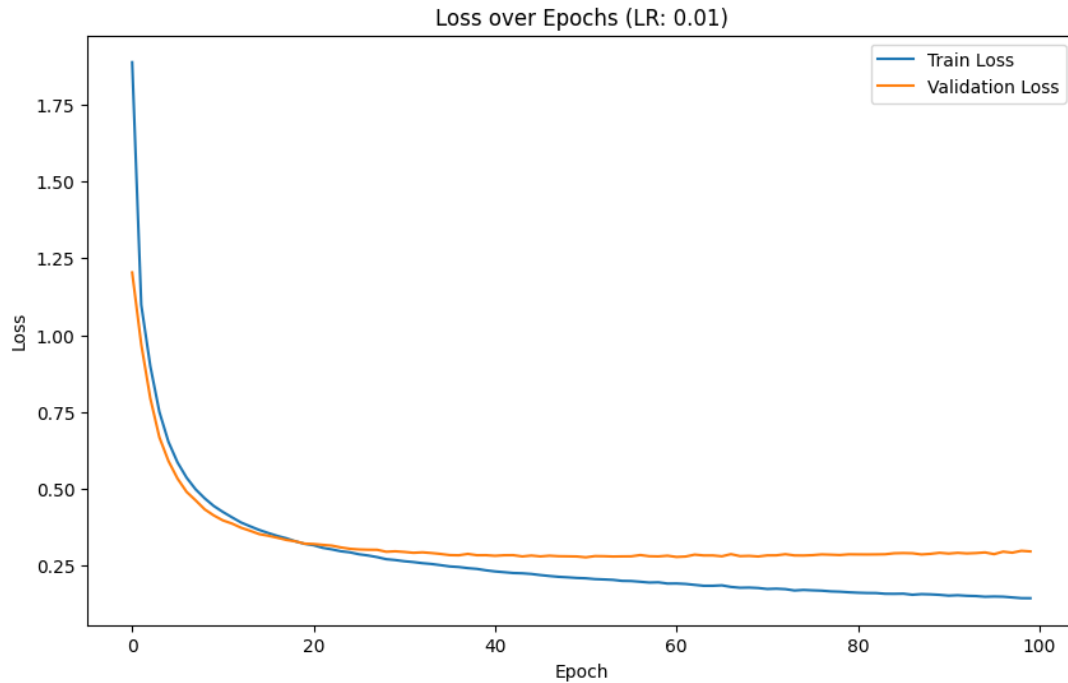
MyModel.addLayer(L3)
MyModel.addFinalLayer(SoftmaxCrossEntropyLayer())
```

1.1.2 Best Metrics:

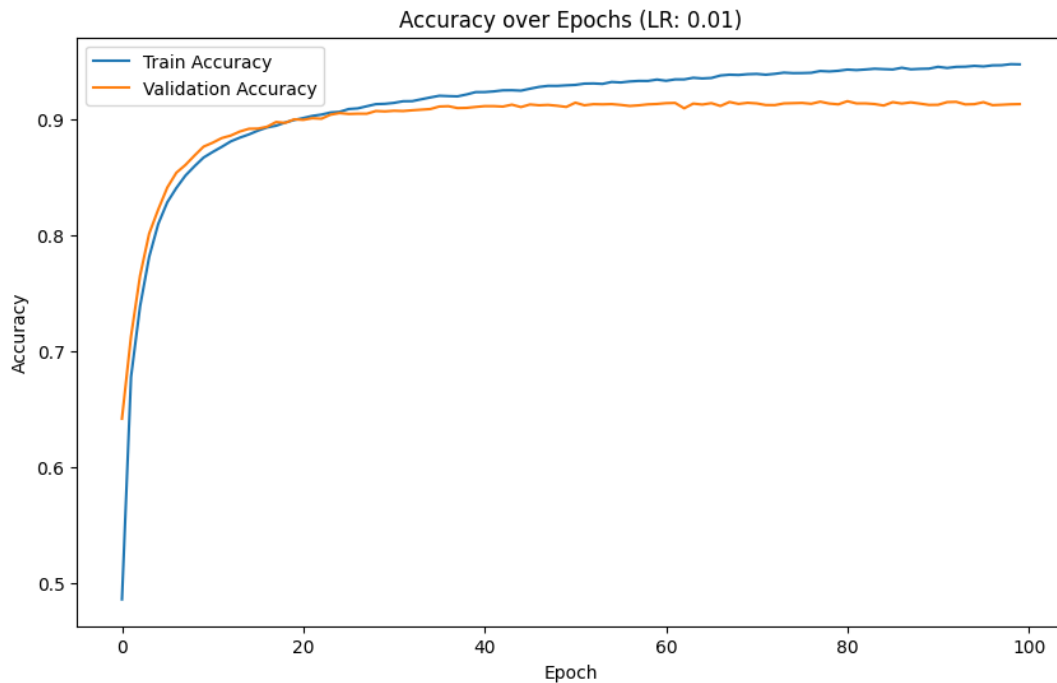
learning rate	loss		accuracy		macro_f1	
	training	validation	training	validation	training	validation
0.01	0.162	0.287	0.943	0.916	0.929	0.916
0.005	0.183	0.278	0.939	0.913	0.936	0.914
0.001	0.370	0.367	0.888	0.888	0.914	0.888
0.0005	0.520	0.505	0.847	0.850	0.853	0.849
5e-5	1.454	1.434	0.584	0.590	0.594	0.584

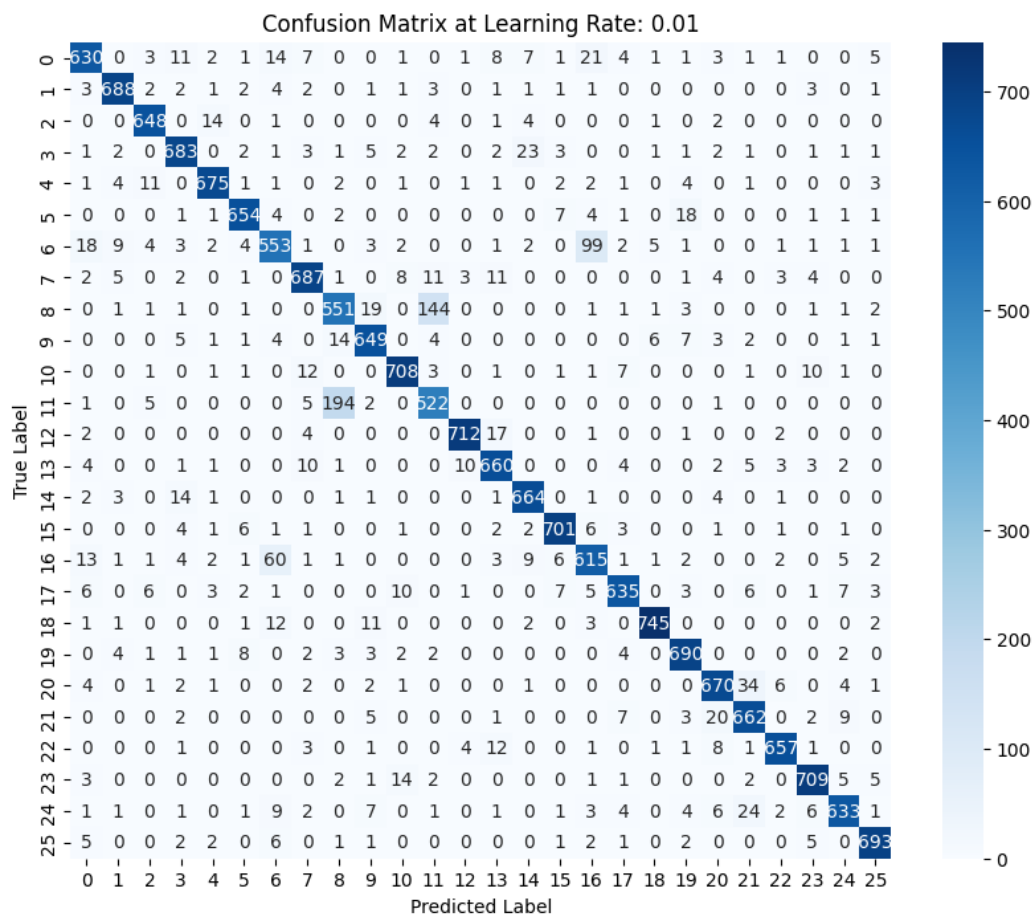
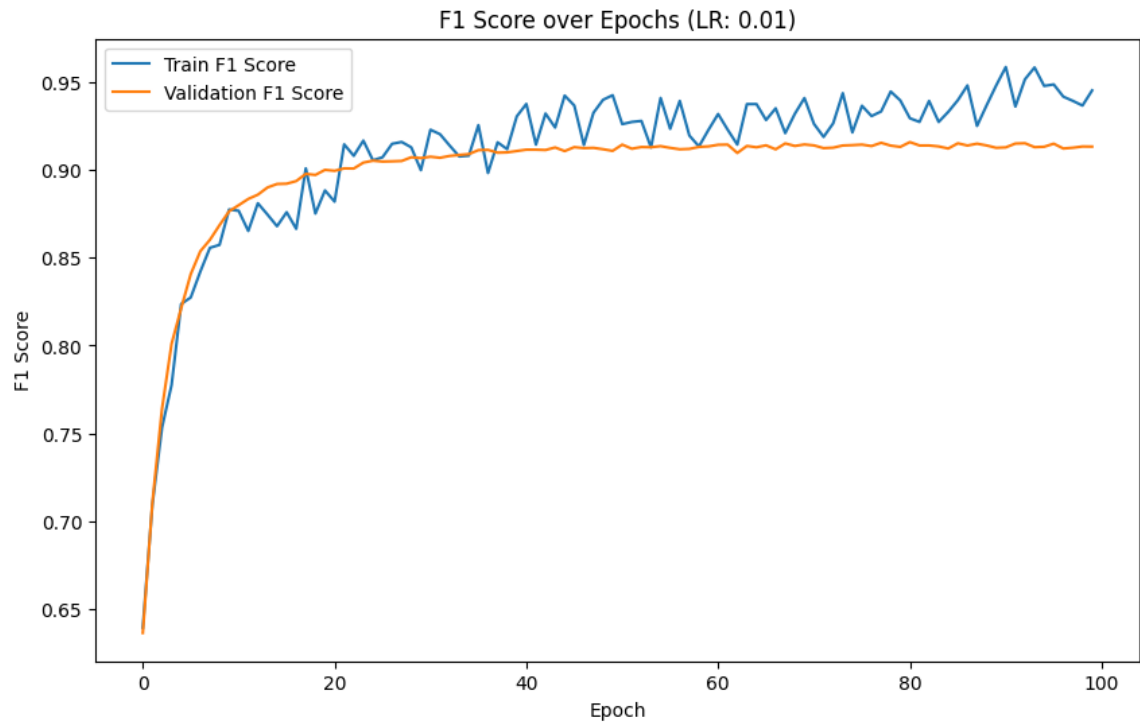
1.1.3 Graphs of Metrics:

1.1.3.1 LR = 0.01 (Best)

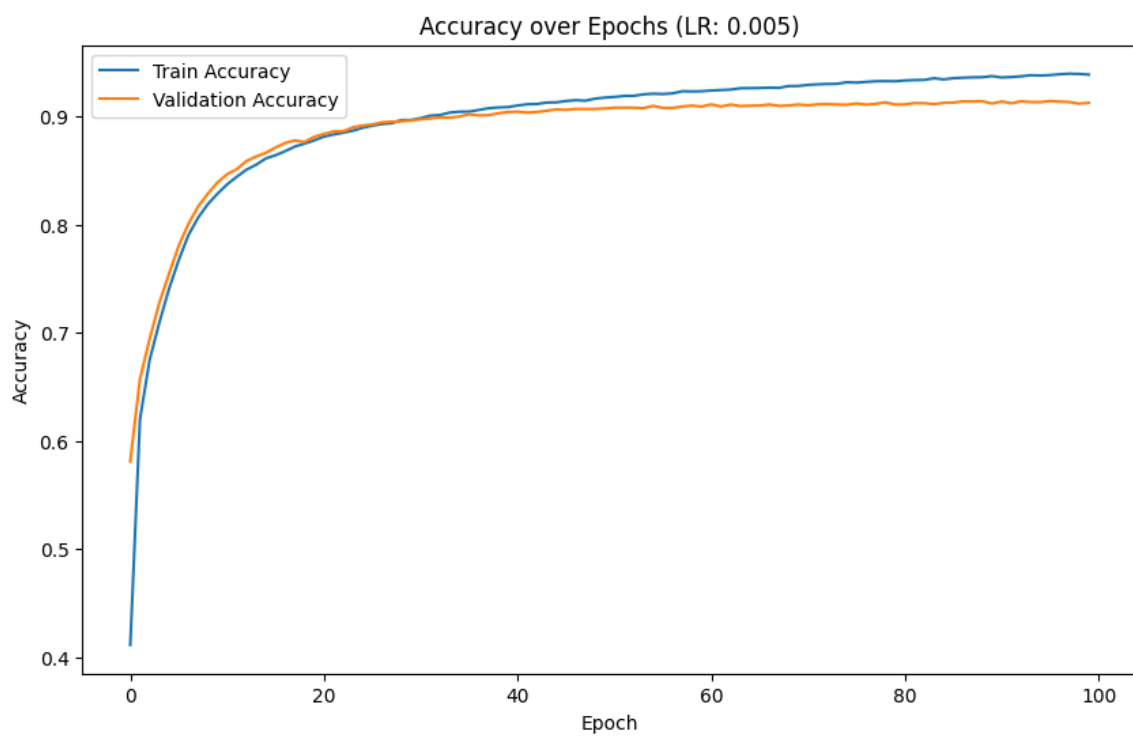
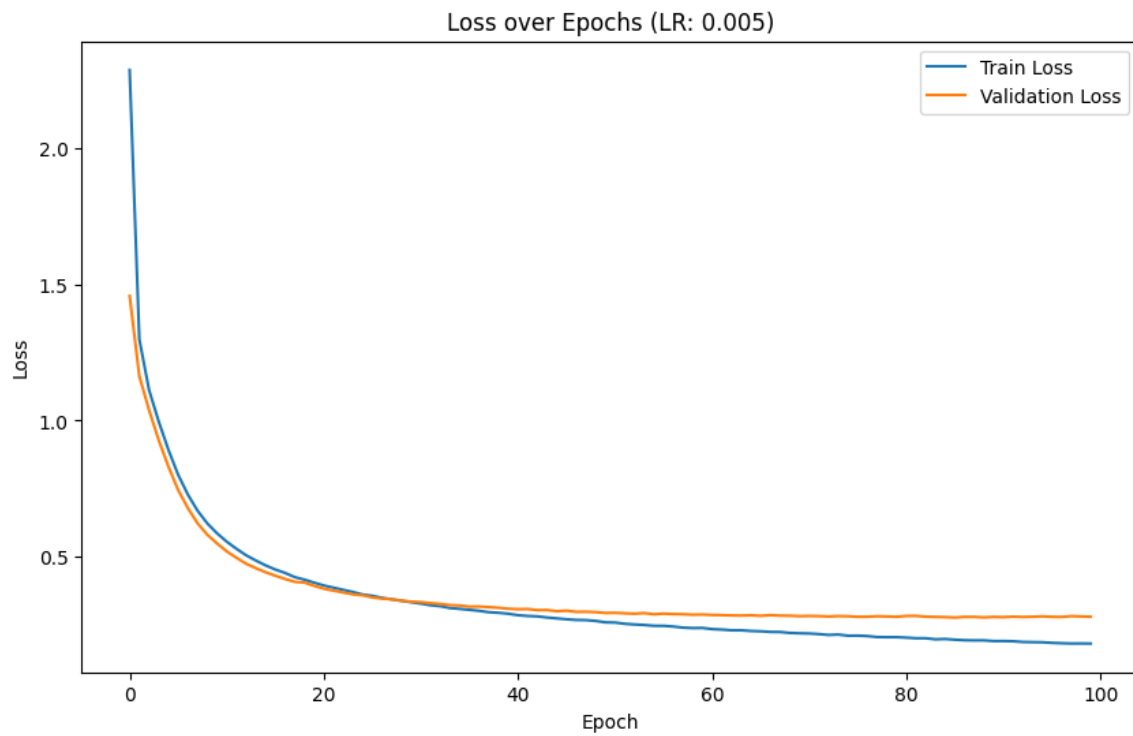


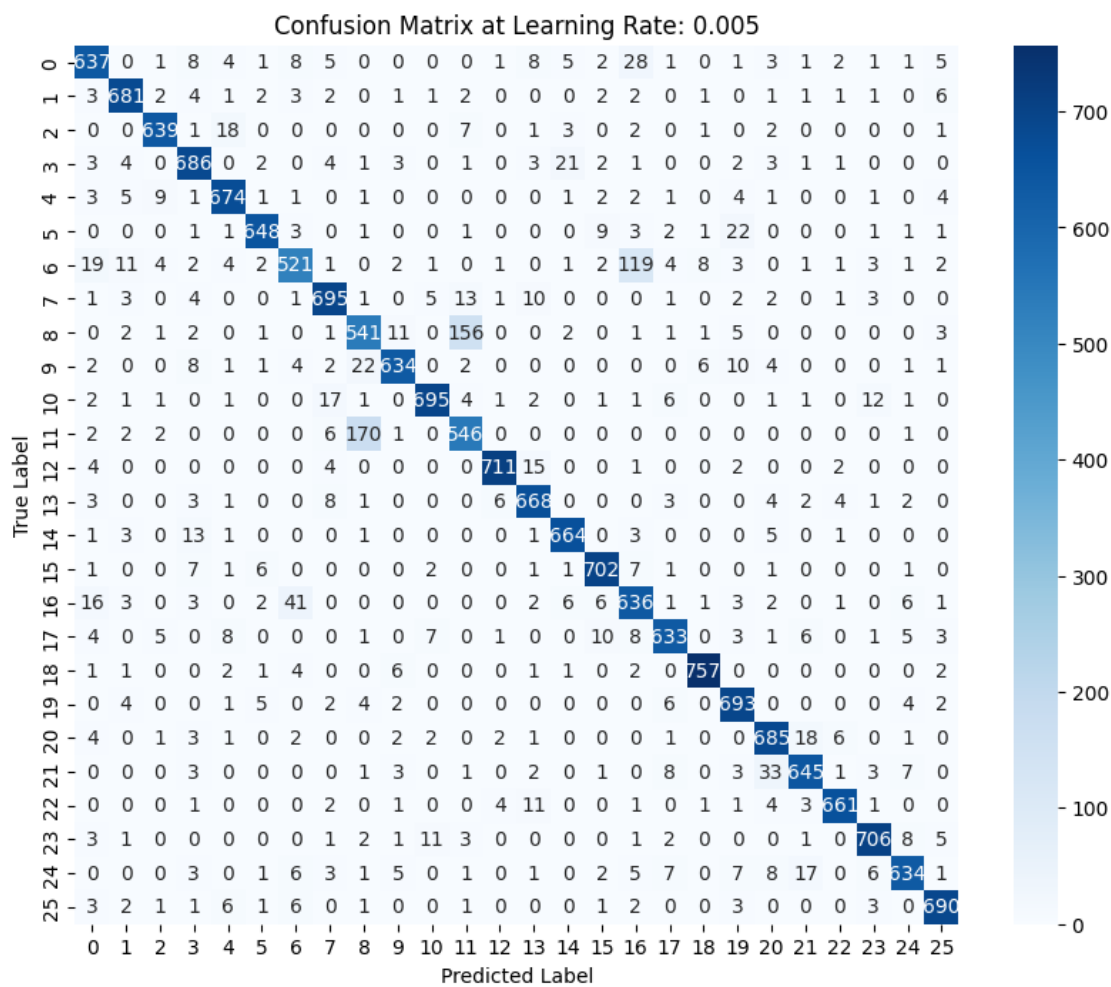
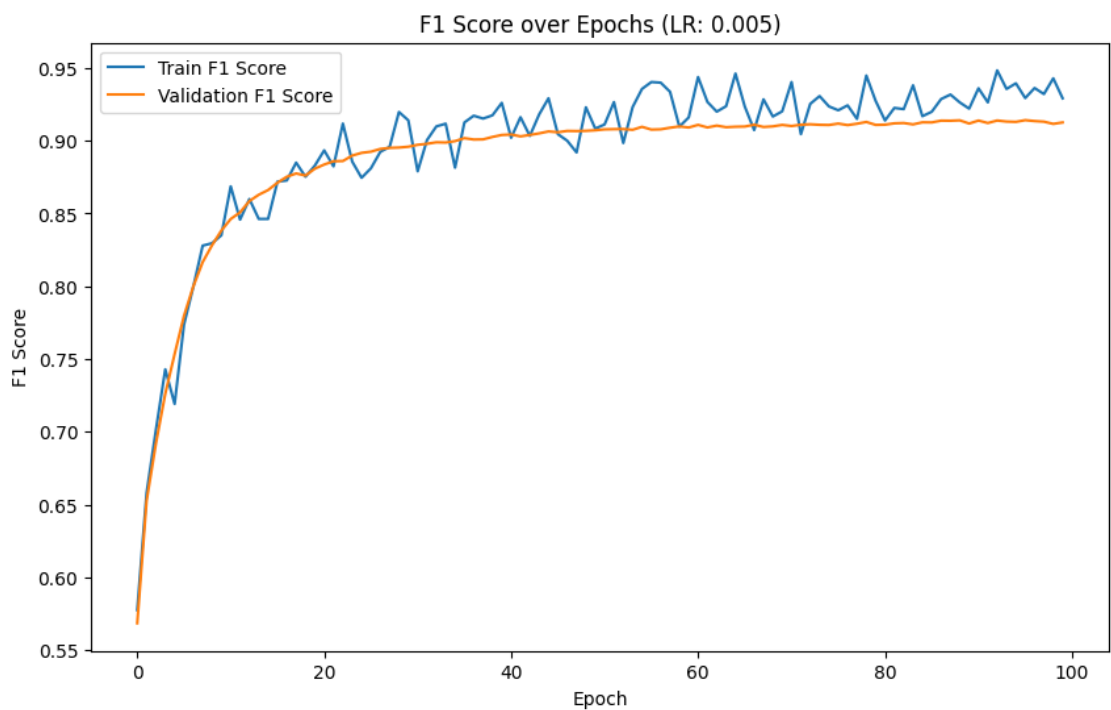
1.3.2 Accuracy



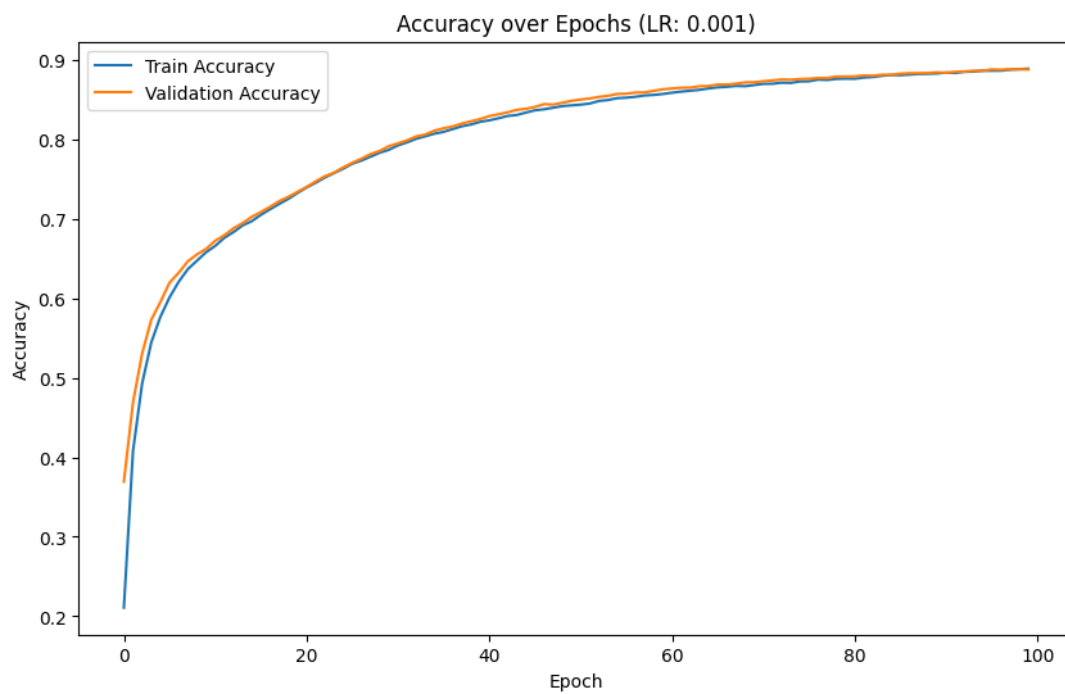
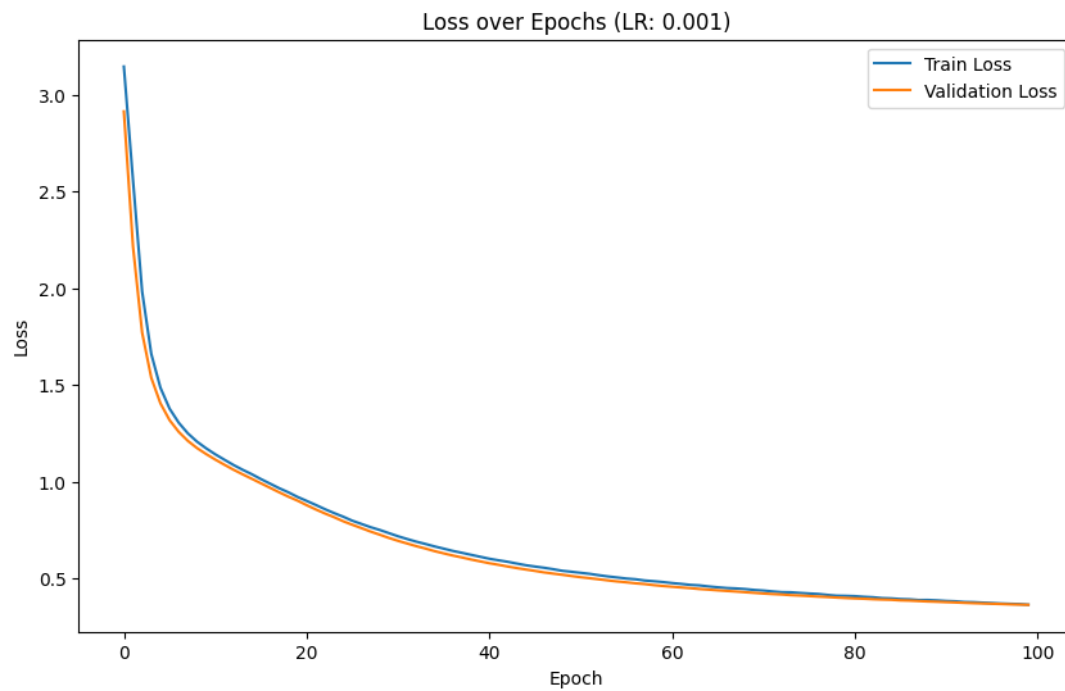


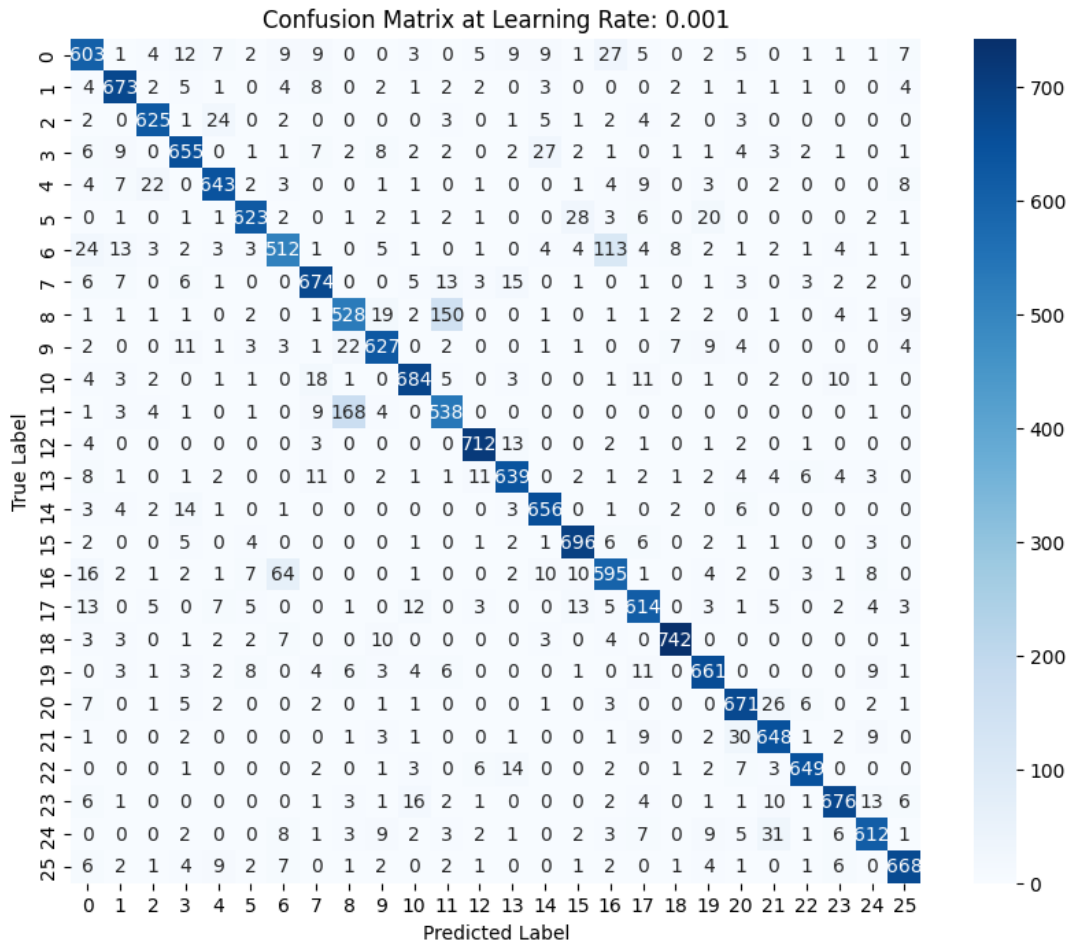
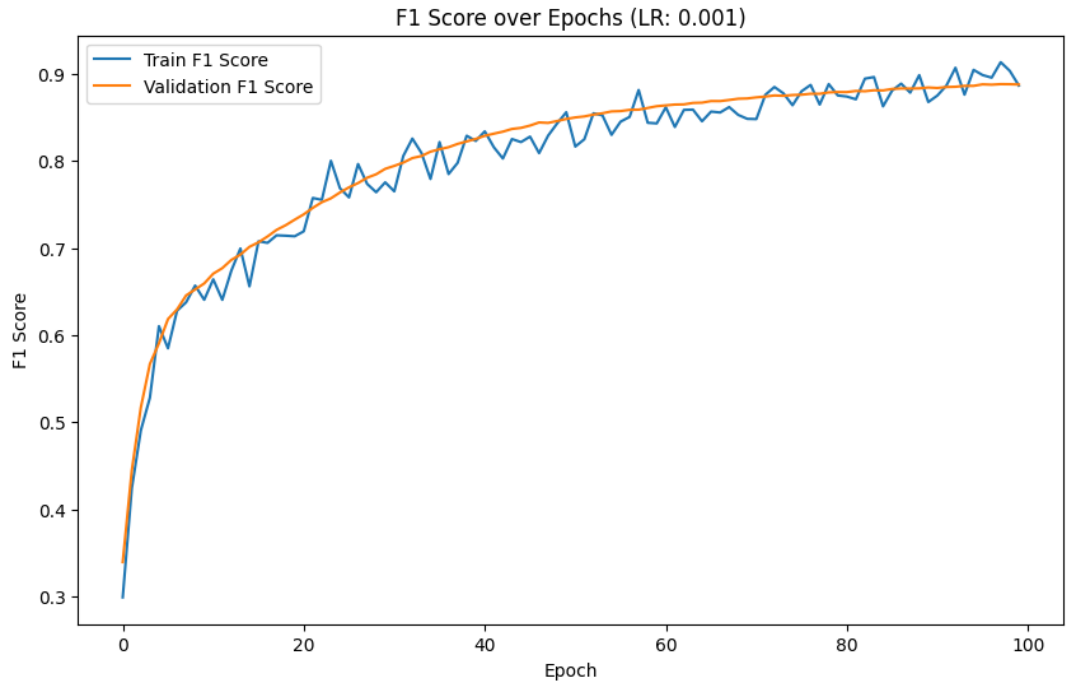
1.1.3.2 LR = 0.005



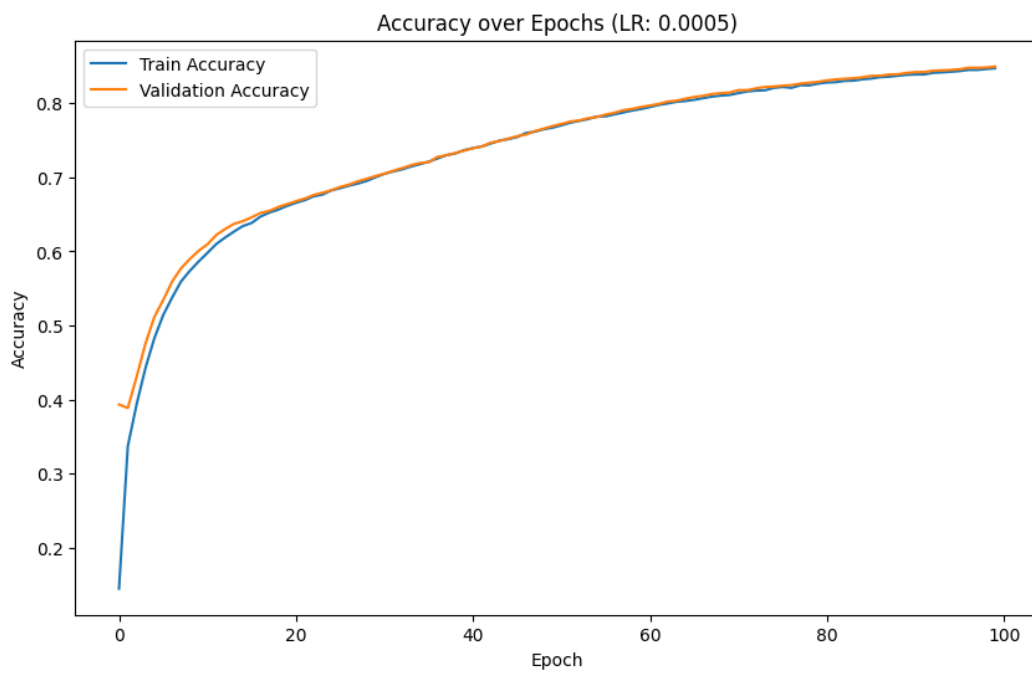
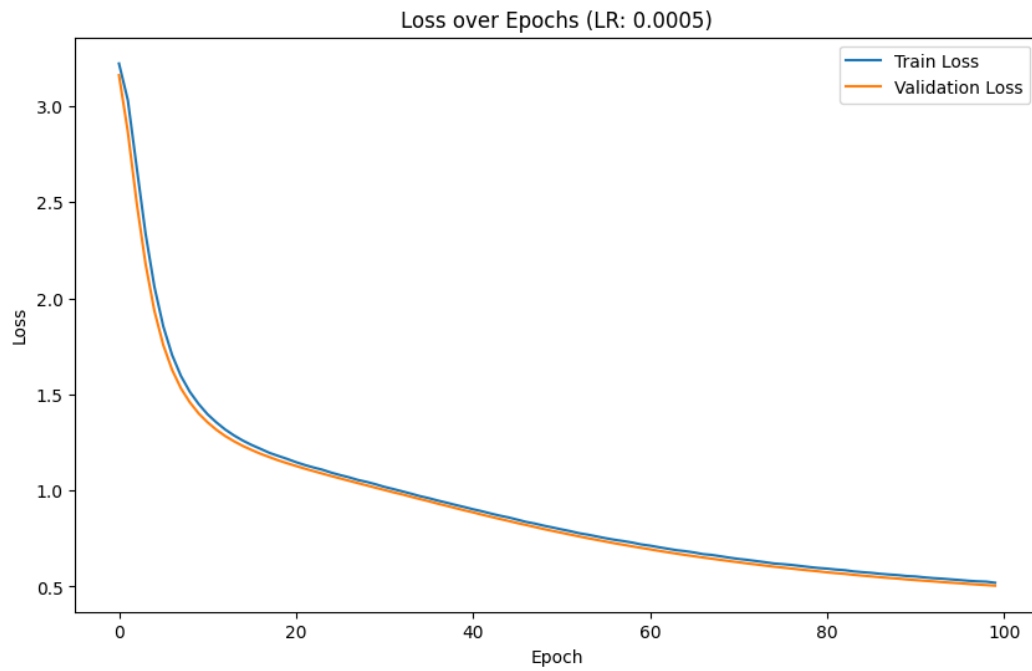


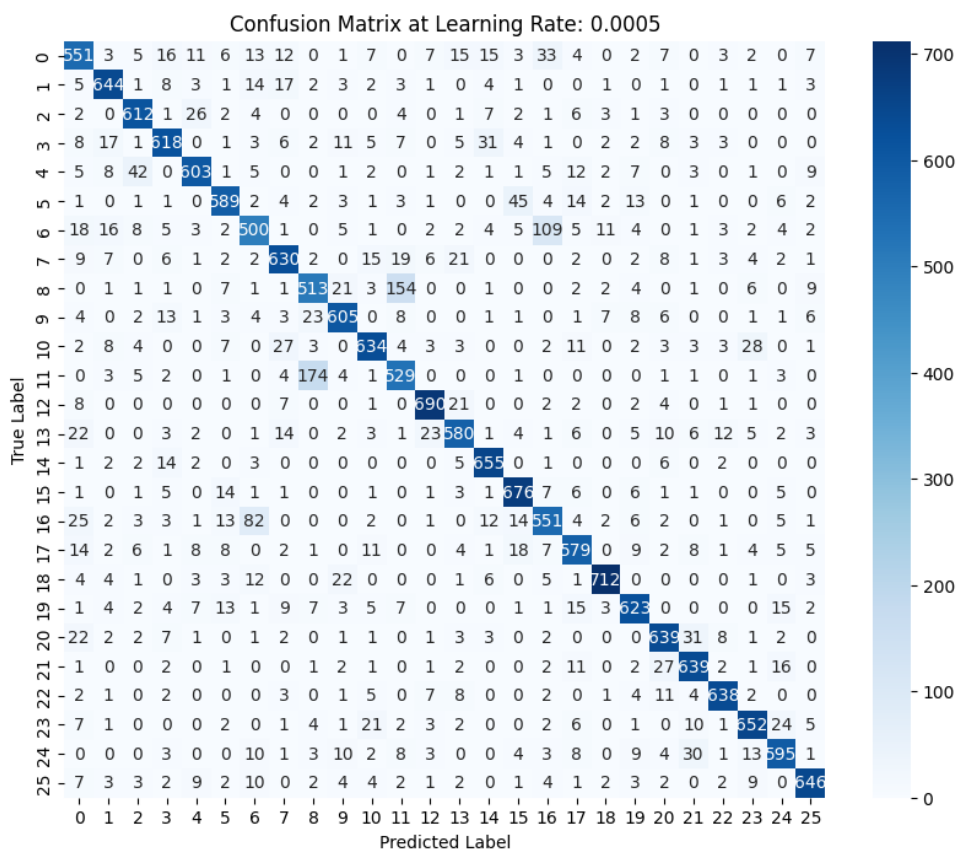
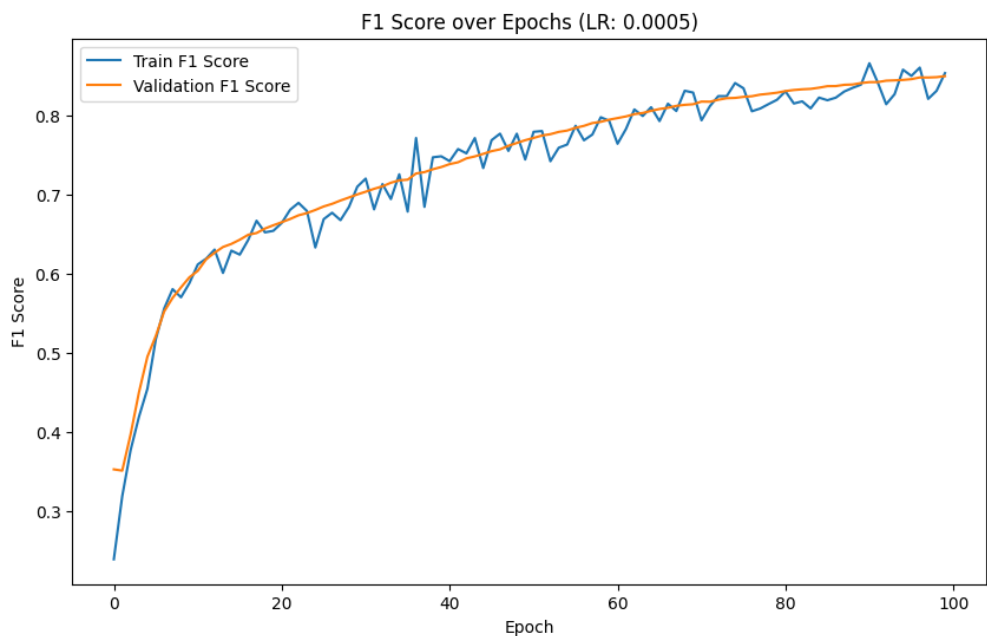
1.1.3.3 LR = 0.001



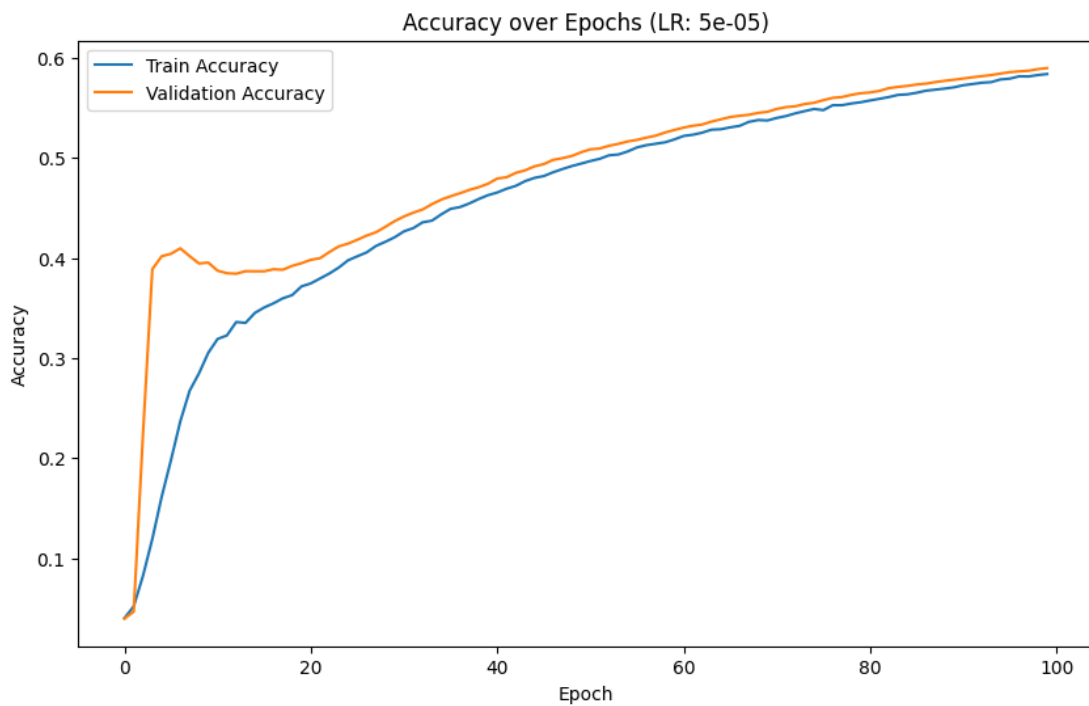
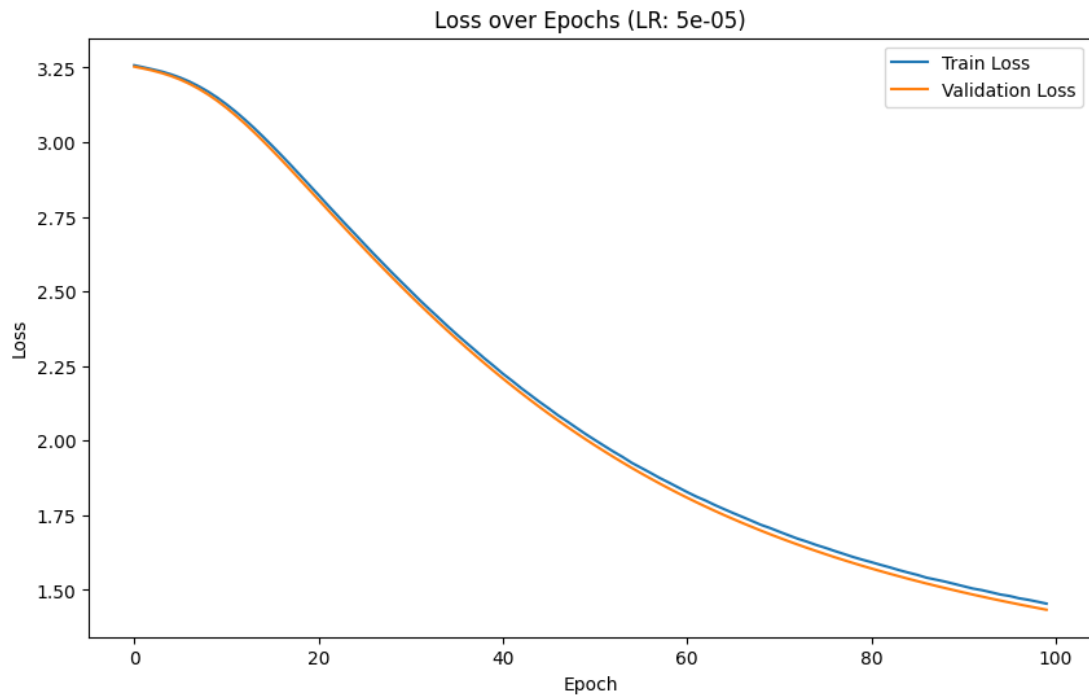


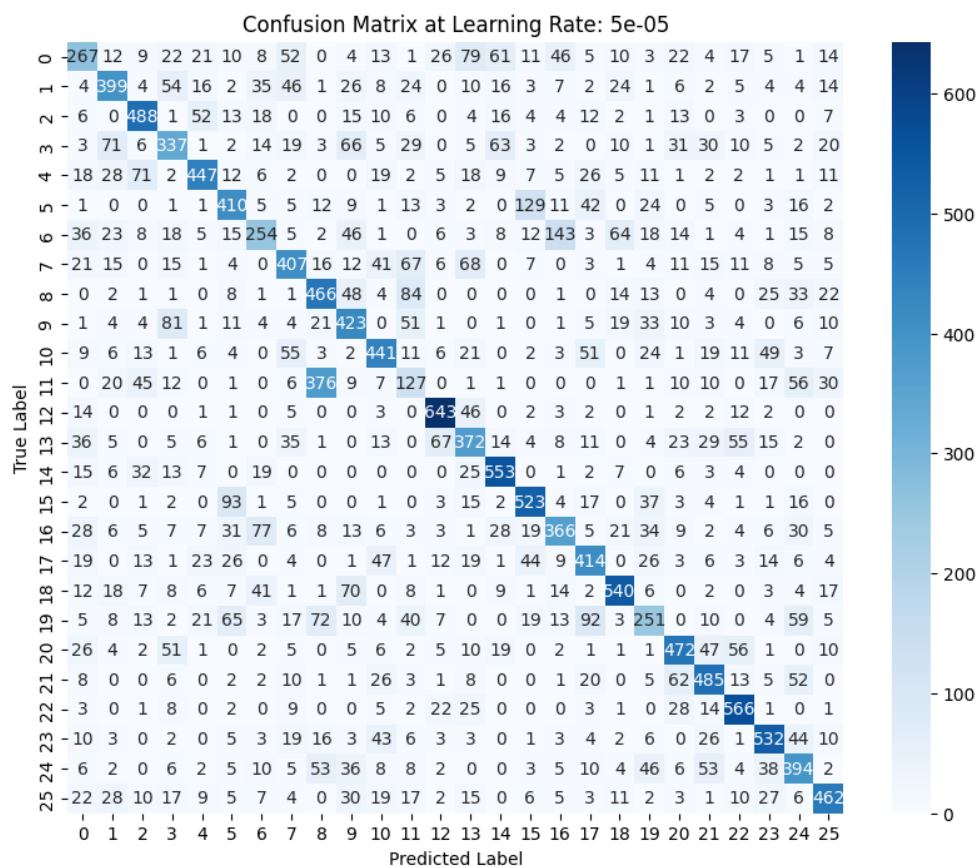
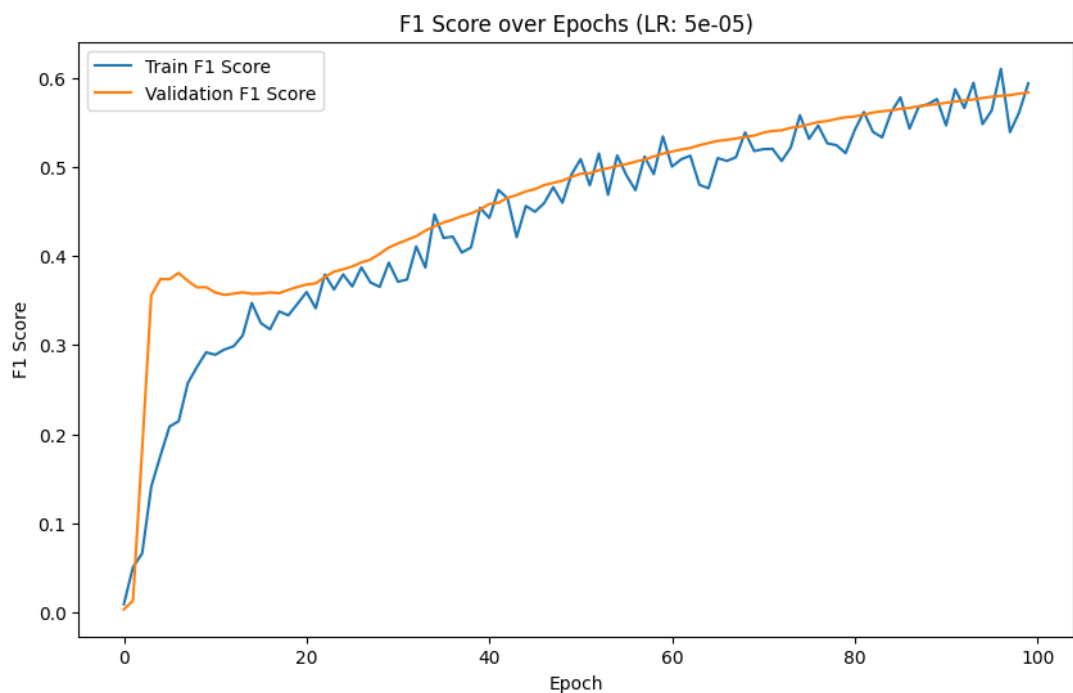
1.1.3.4 LR = 0.0005





1.1.3.5 LR = 5e-5





1.2. Model 2:

1.2.1 Architecture

```
MyModel = FNNmodel()
L1 = DenseLayer(784, 254)
L1.init_weights_he()
MyModel.addLayer(L1)
# CHANGE
L2 = ReLUActivationLayer()
MyModel.addLayer(L2)

L2_5 = DropoutLayer(0.3)
MyModel.addLayer(L2_5)

L3 = DenseLayer(254, 26)
L3.init_weights_he()

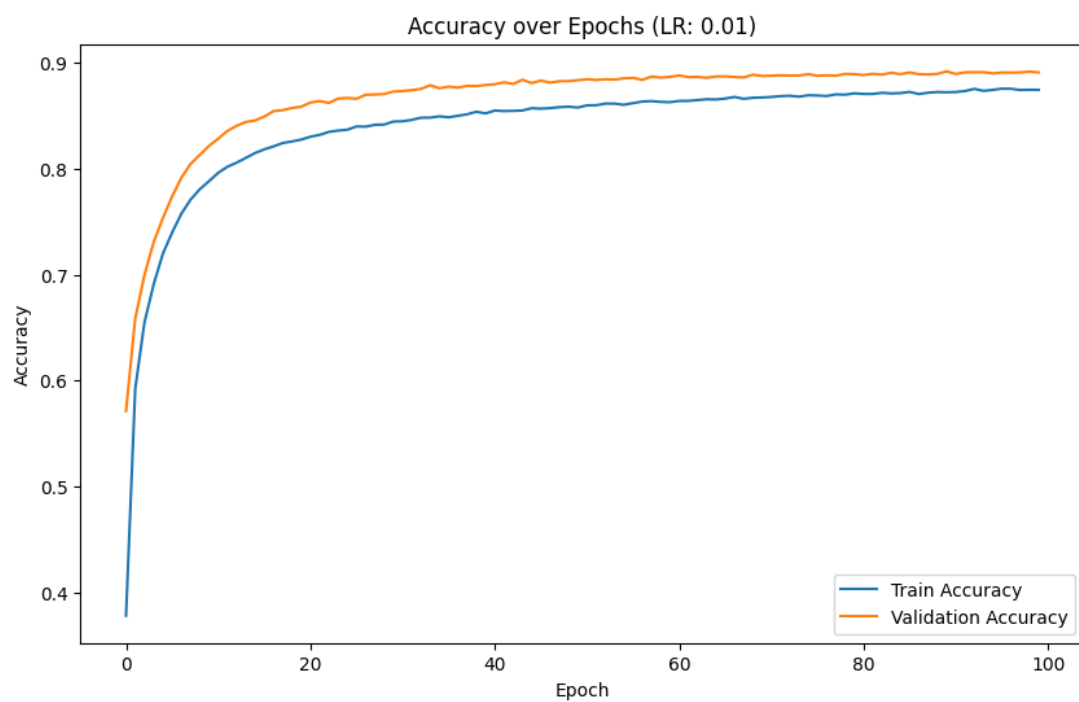
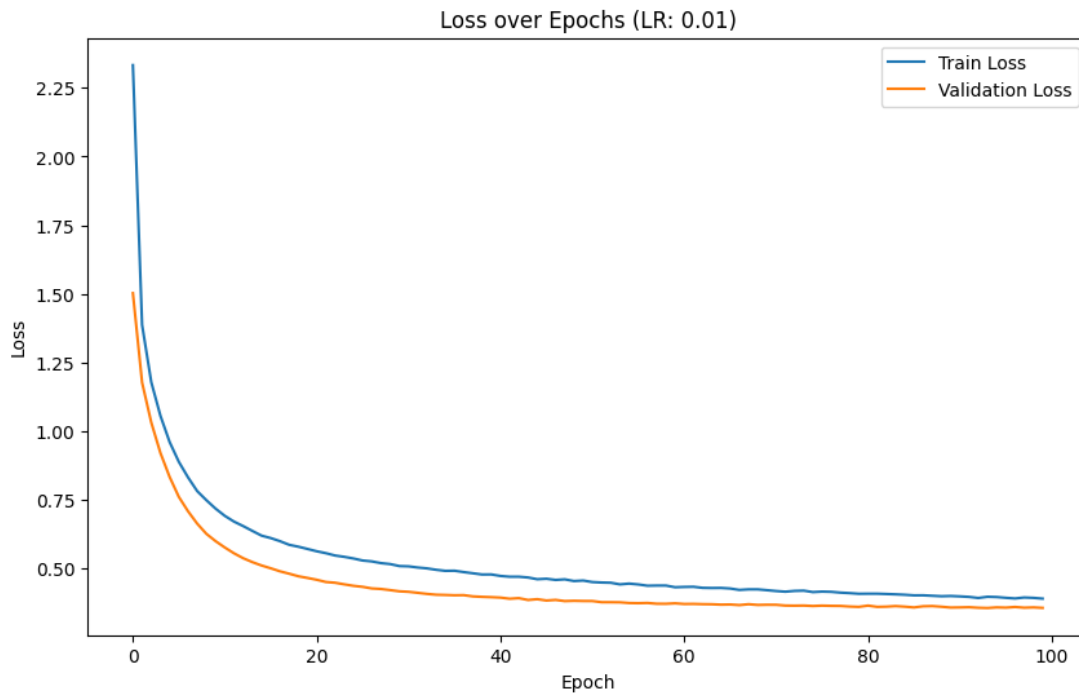
MyModel.addLayer(L3)
MyModel.addFinalLayer(SoftmaxCrossEntropyLayer())
```

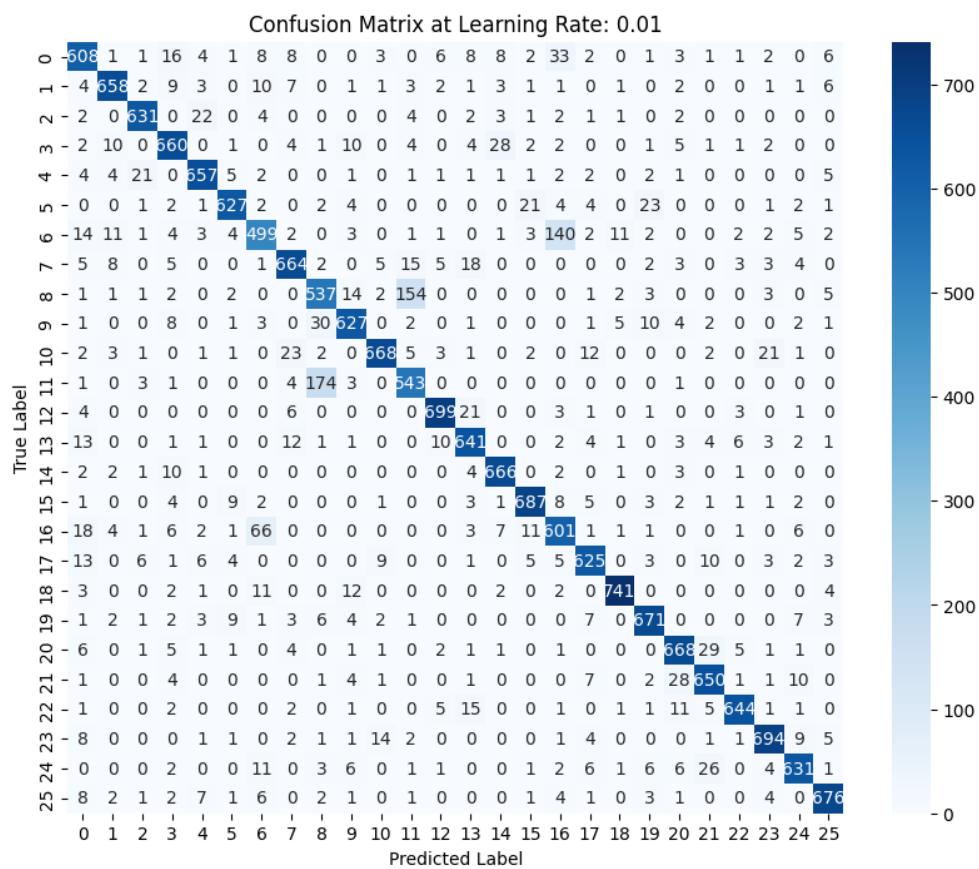
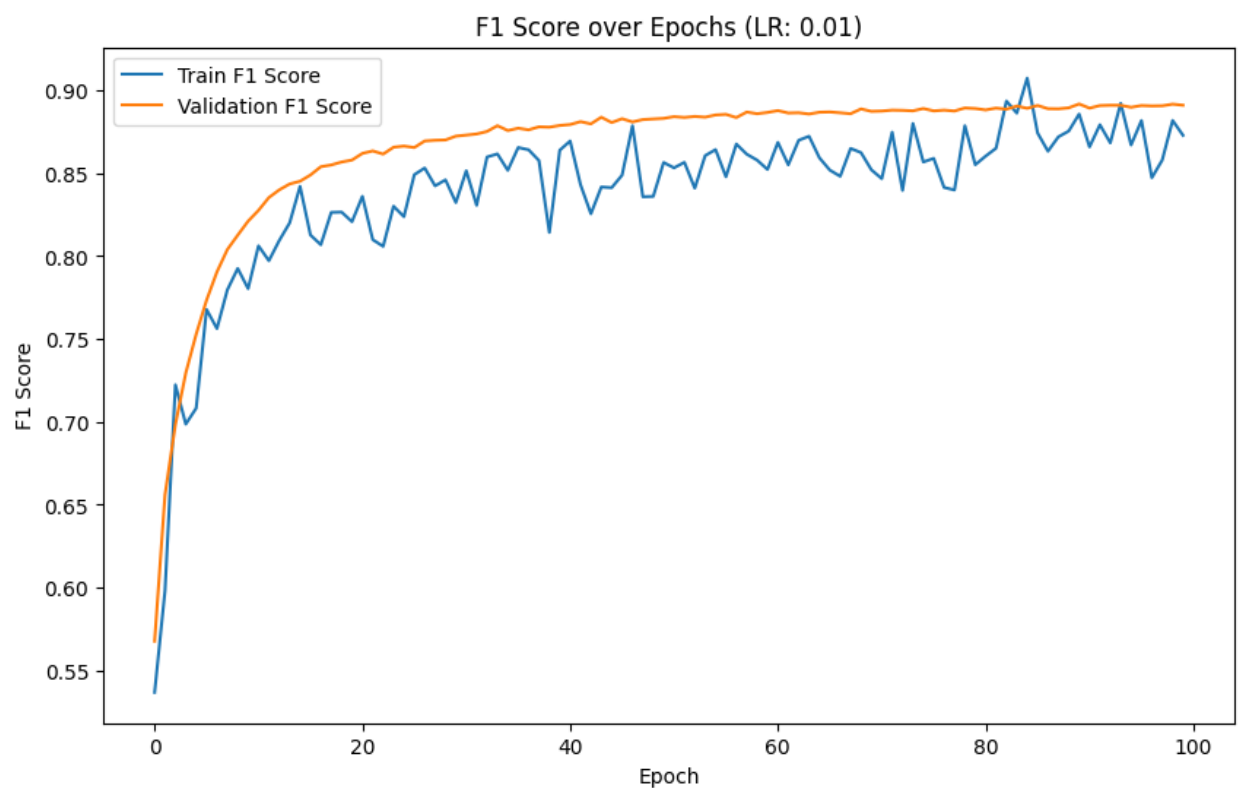
1.2.2 Best Metrics:

learning rate	loss		accuracy		macro_f1	
	training	validation	training	validation	training	validation
0.01	0.390	0.356	0.874	0.891	0.874	0.891
0.005	0.420	0.363	0.868	0.887	0.886	0.887
0.001	0.653	0.557	0.807	0.834	0.797	0.834
0.0005	0.827	0.746	0.758	0.780	0.758	0.779

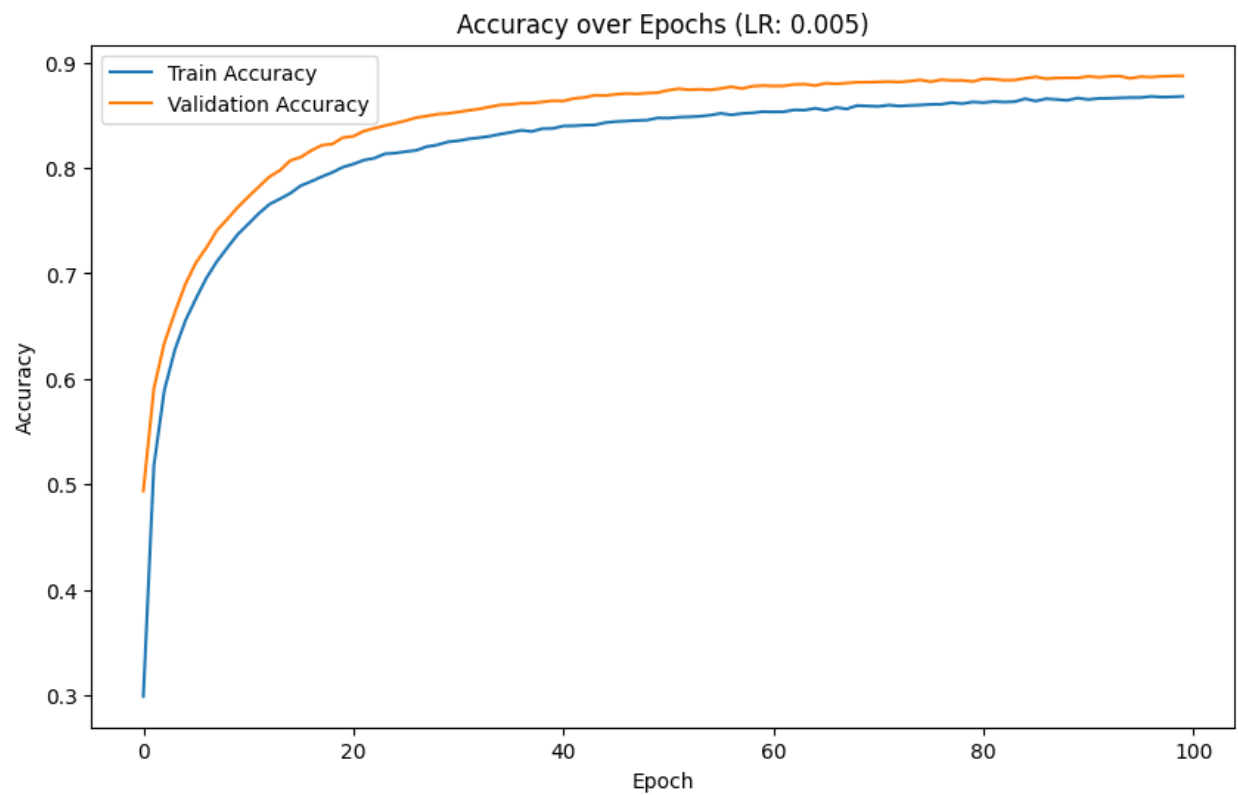
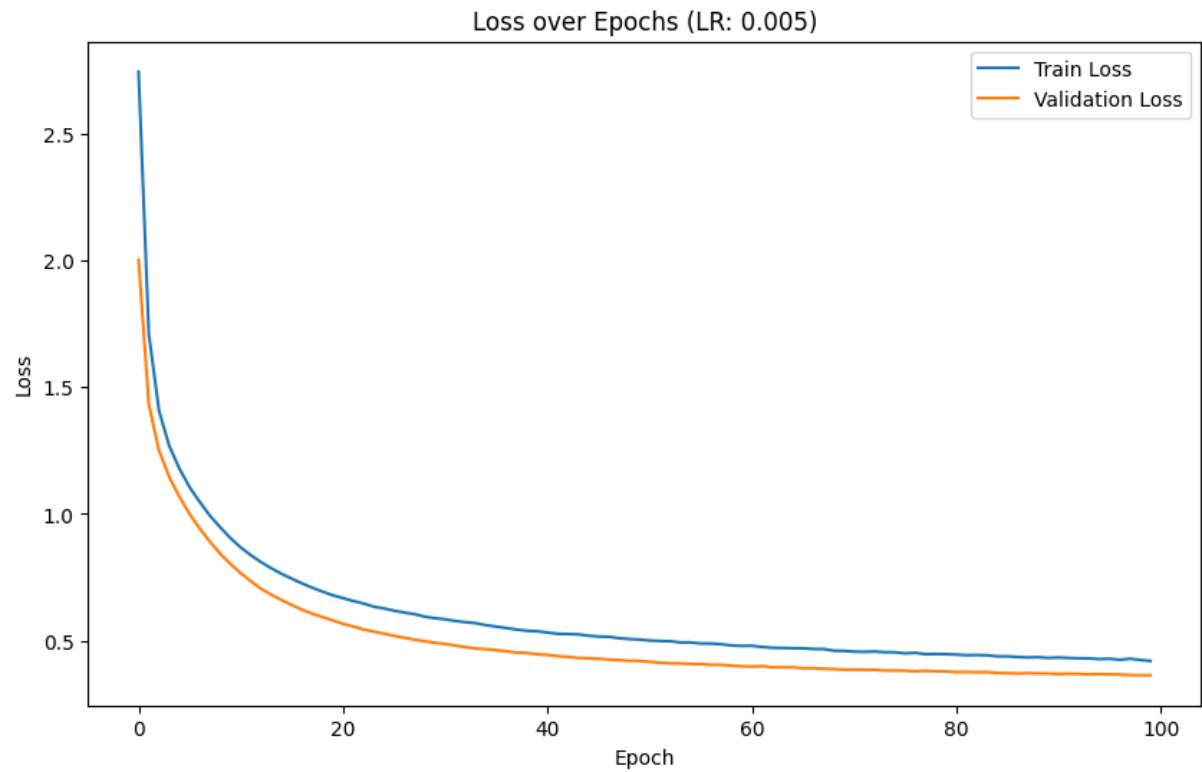
1.2.3 Graphs of Metrics:

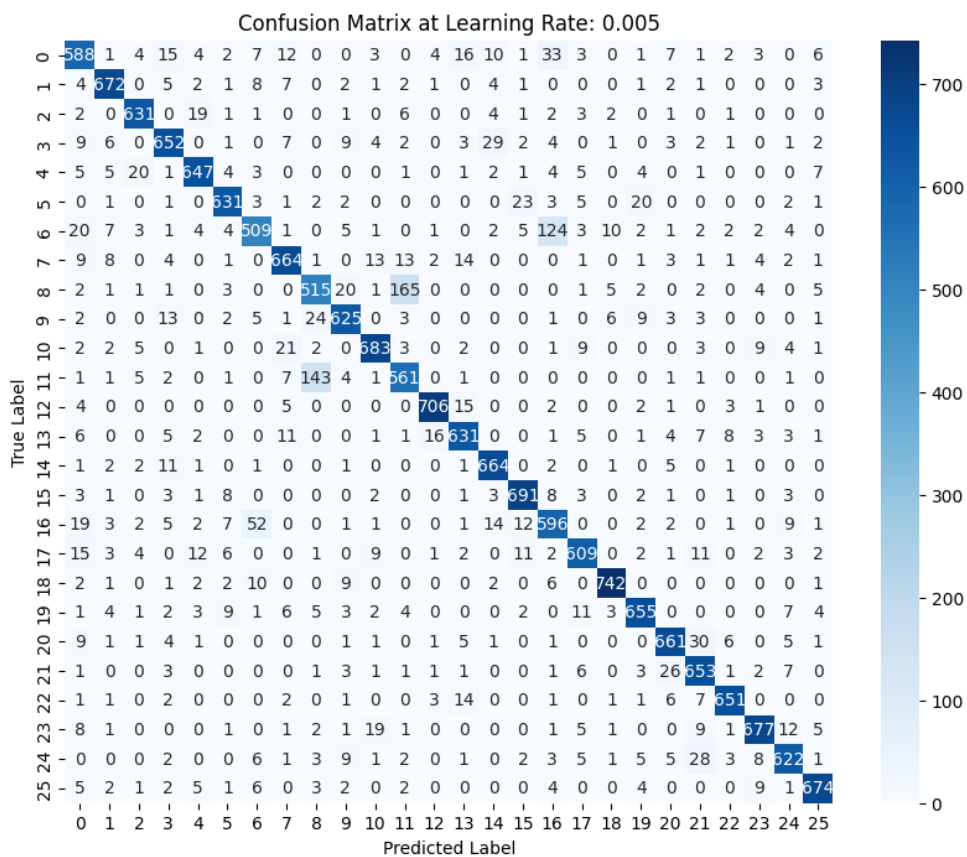
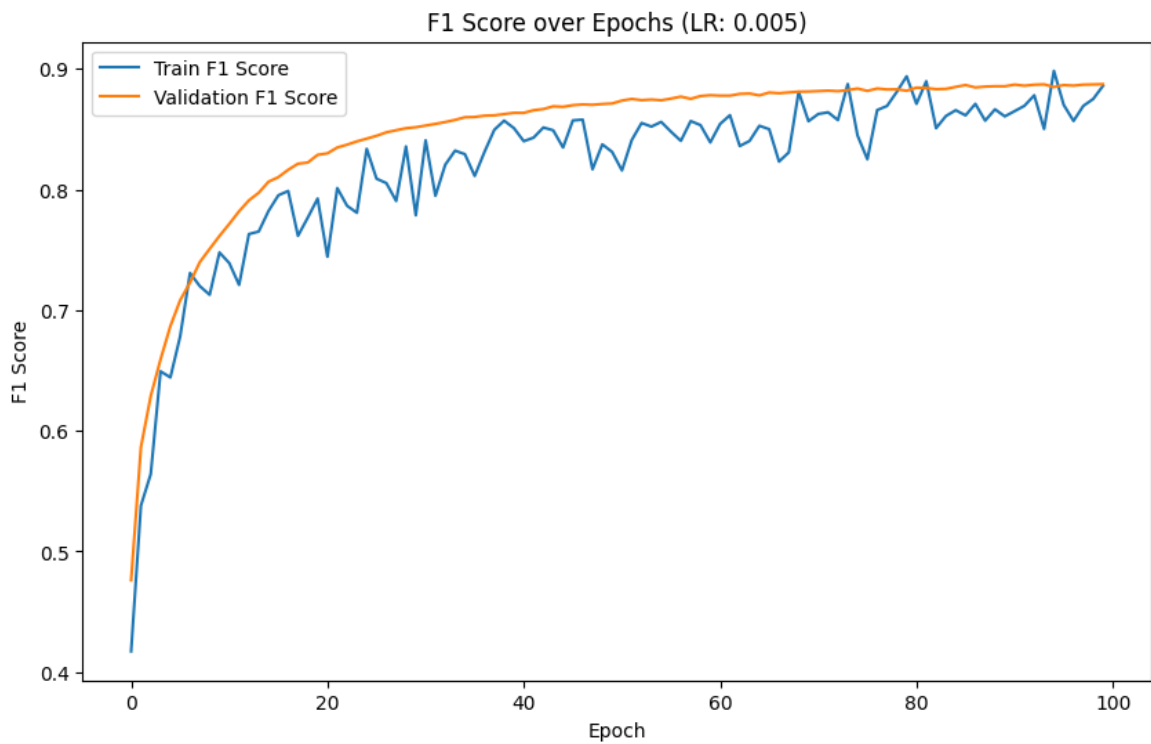
1.2.3.1: LR = 0.01



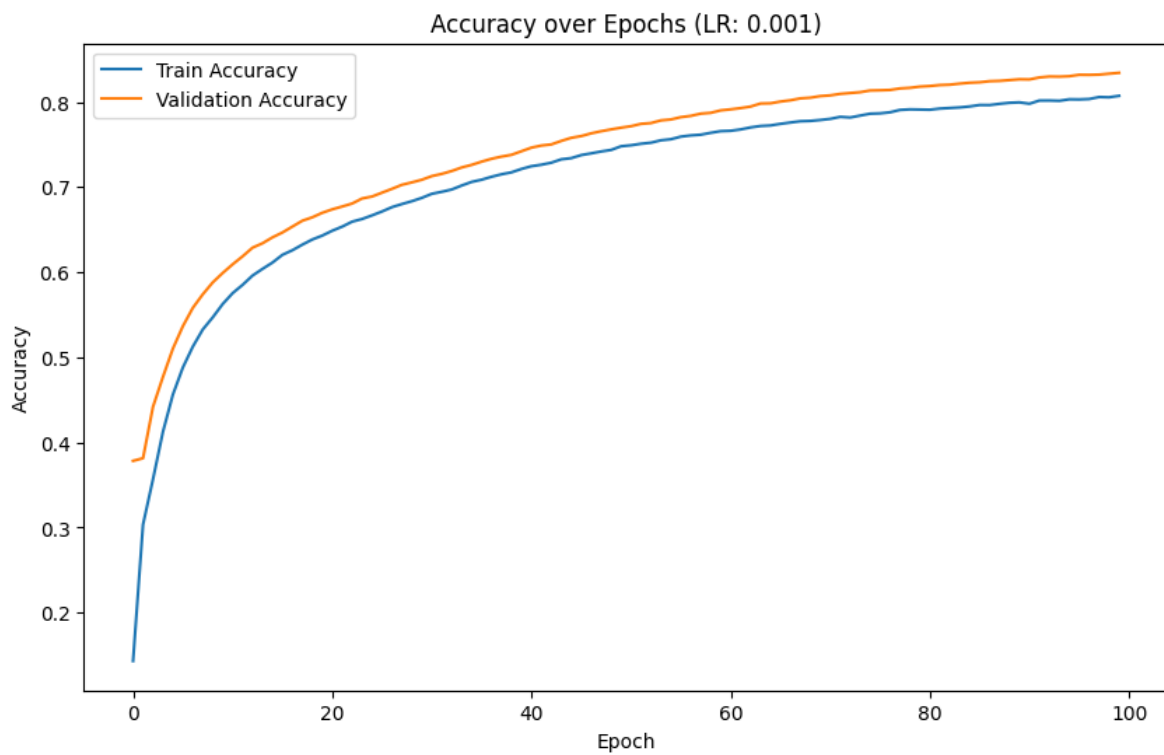
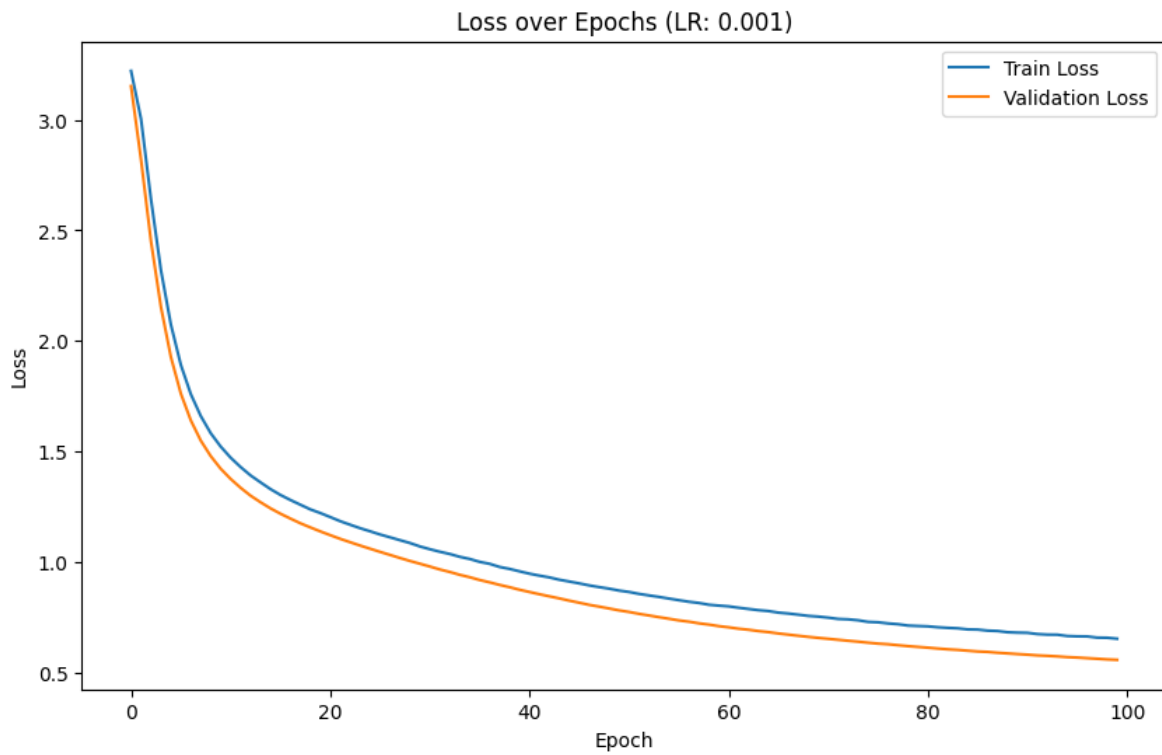


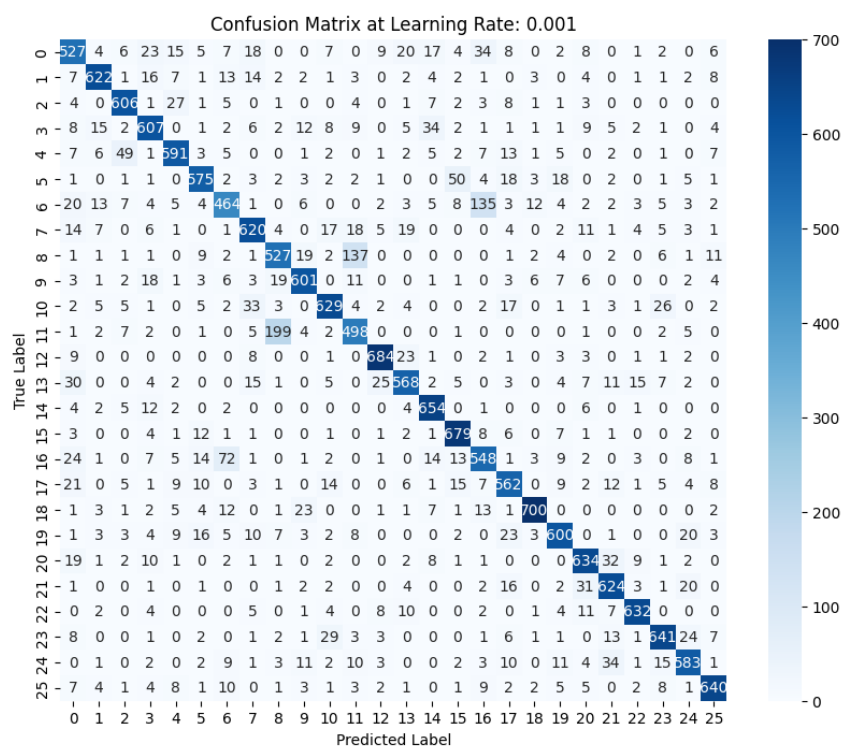
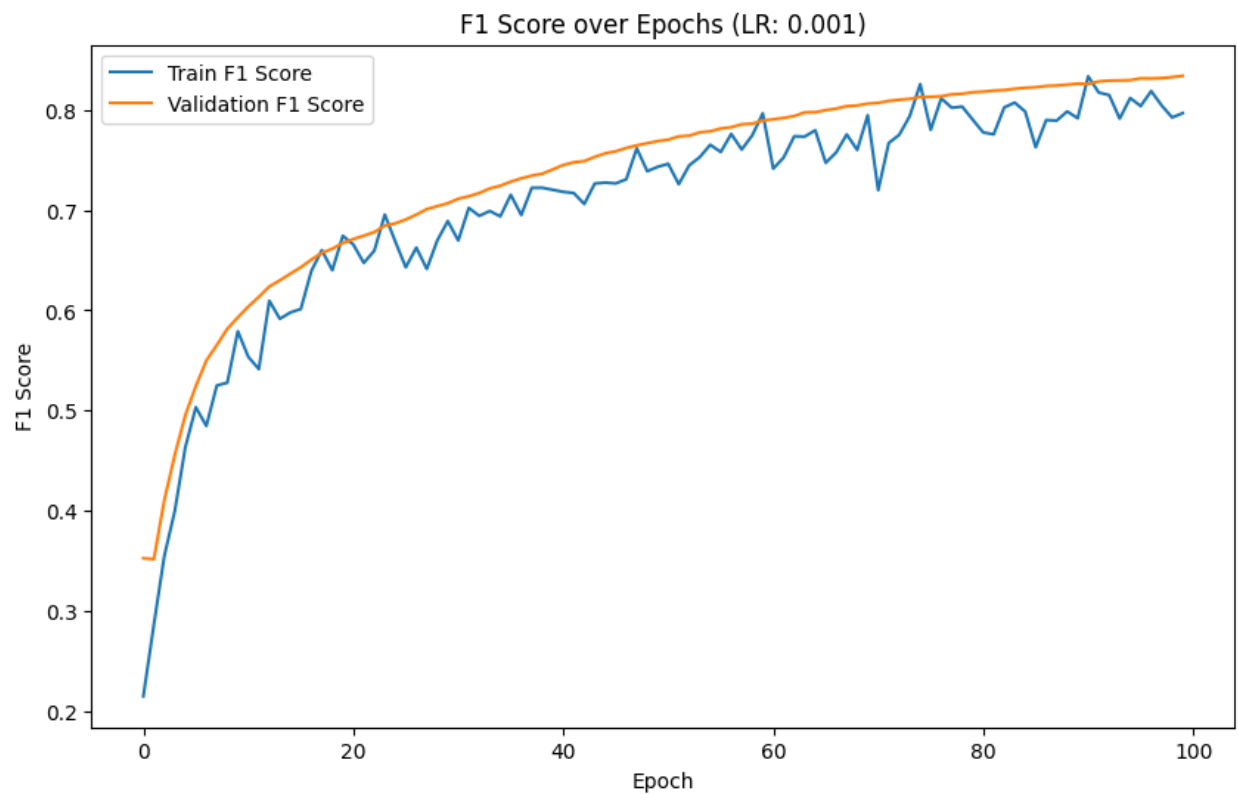
1.2.3.2 LR = 0.005



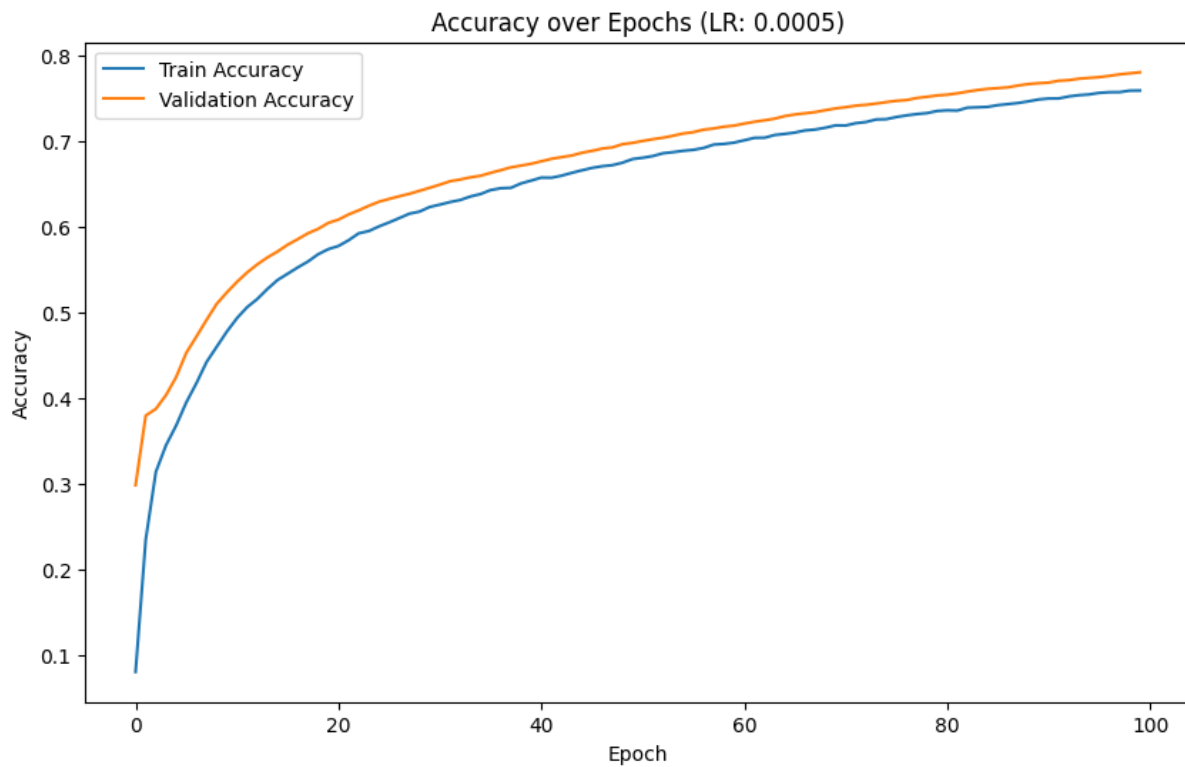
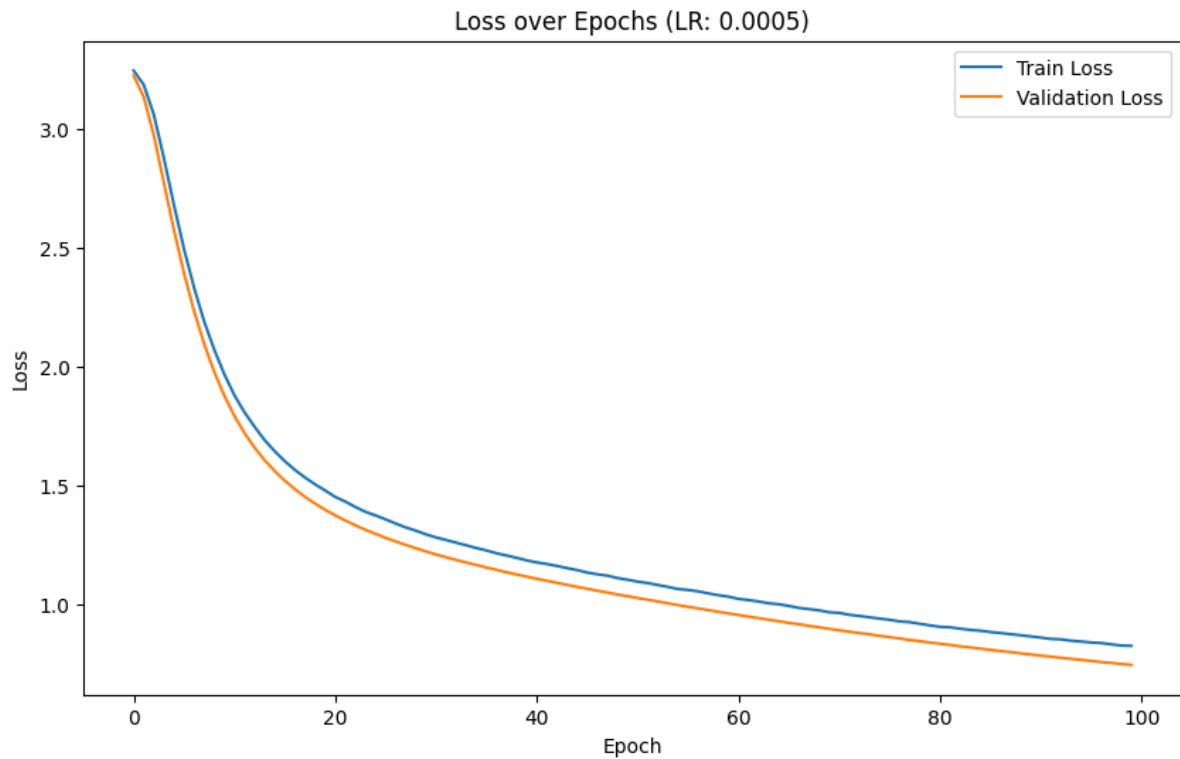


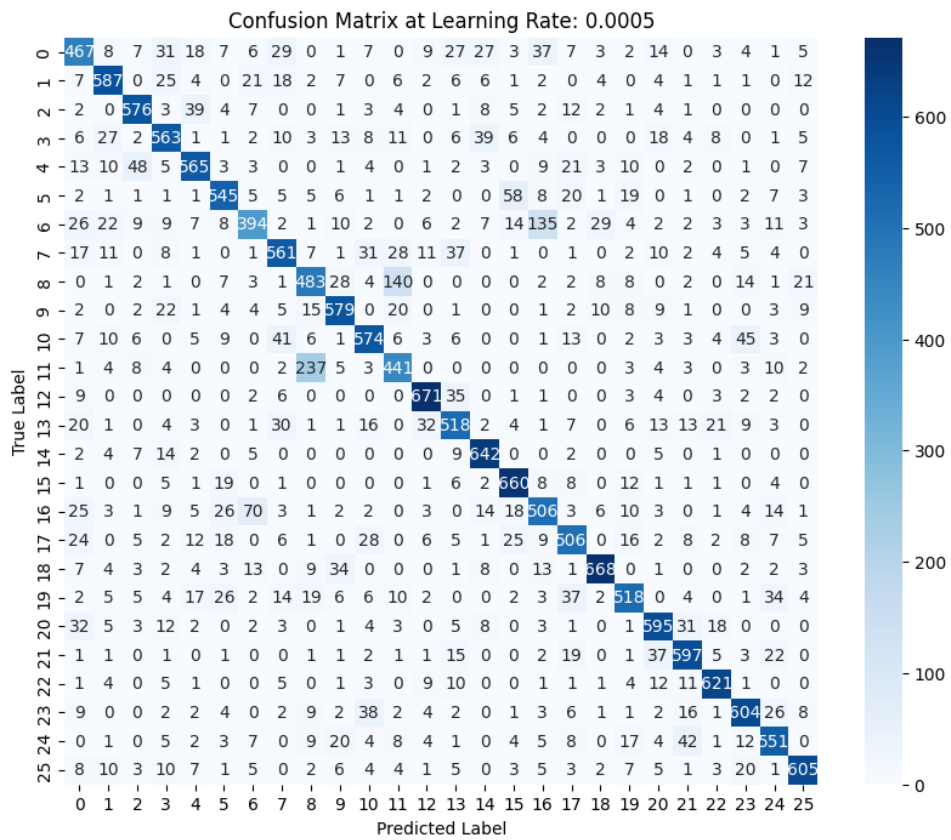
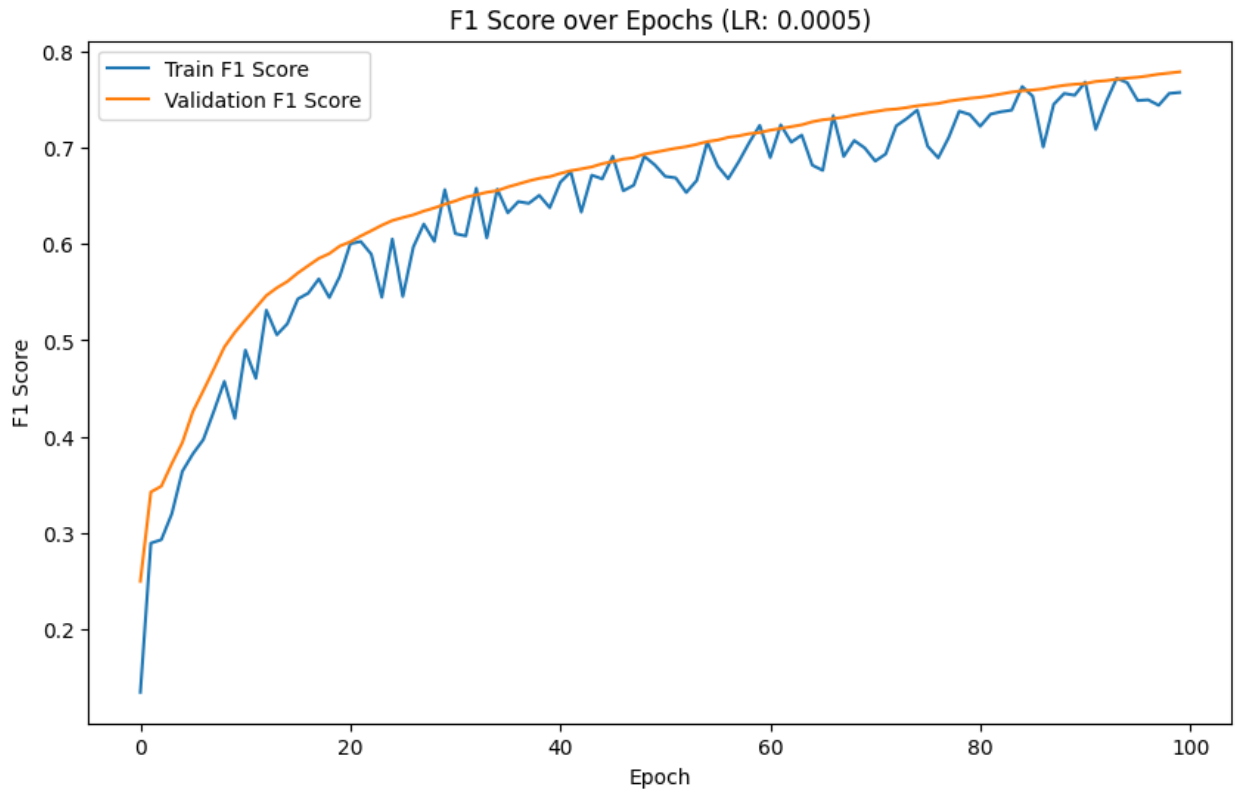
1.2.3.3 LR = 0.001





1.2.3.4 LR = 0.0005





1.3. Model 3:

1.3.1 Architecture

```
MyModel = FNNmodel()
L1 = DenseLayer(784, 256)
L1.init_weights_he()
MyModel.addLayer(L1)
# CHANGE
L2 = ReLUActivationLayer()
MyModel.addLayer(L2)

L2_5 = DropoutLayer(0.3)
MyModel.addLayer(L2_5)

L3 = DenseLayer(256, 128)
L3.init_weights_he()
MyModel.addLayer(L3)
# CHANGE
L4 = ReLUActivationLayer()
MyModel.addLayer(L4)

L4_5 = DropoutLayer(0.3)
MyModel.addLayer(L4_5)

L5 = DenseLayer(128, 26)
L5.init_weights_he()
MyModel.addLayer(L5)

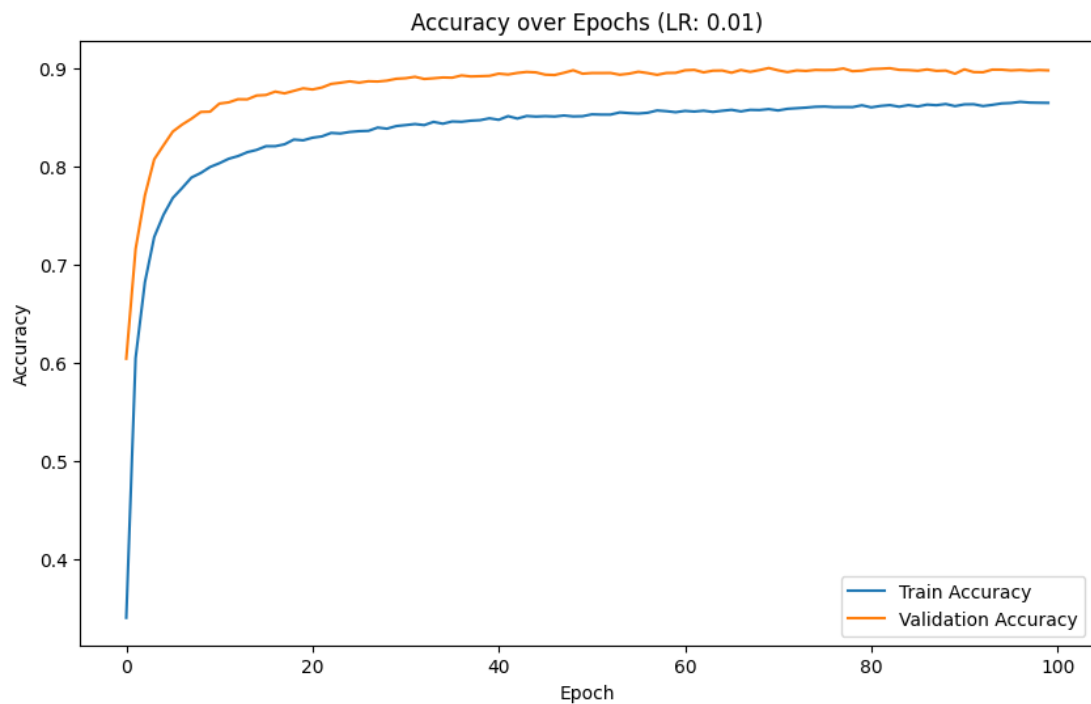
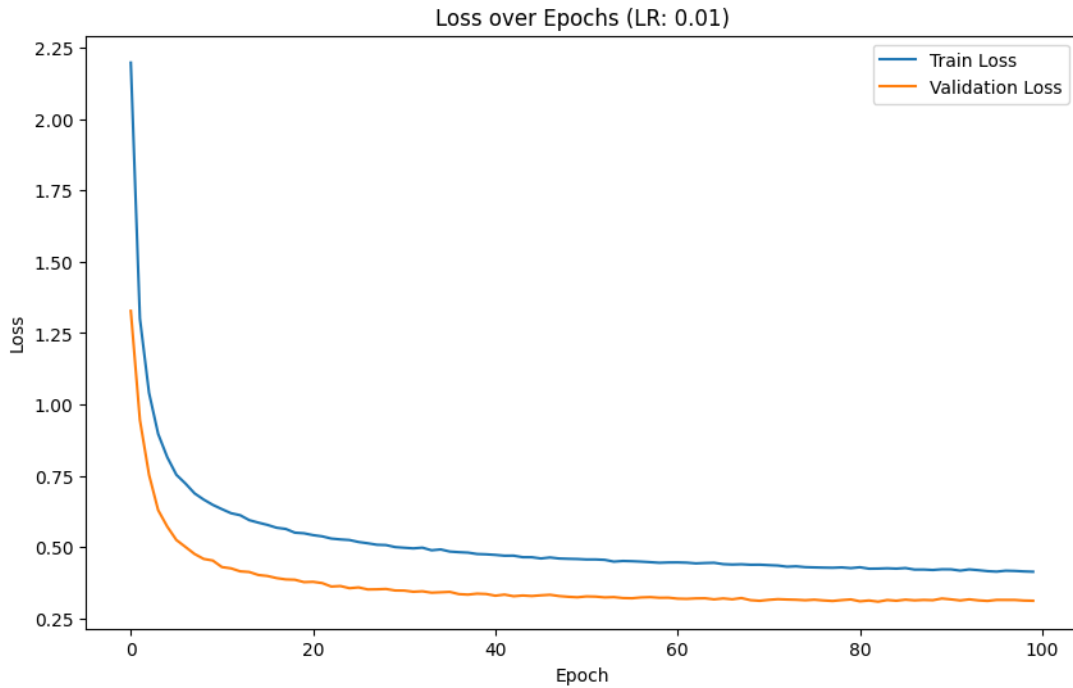
MyModel.addFinalLayer(SoftmaxCrossEntropyLayer())
```

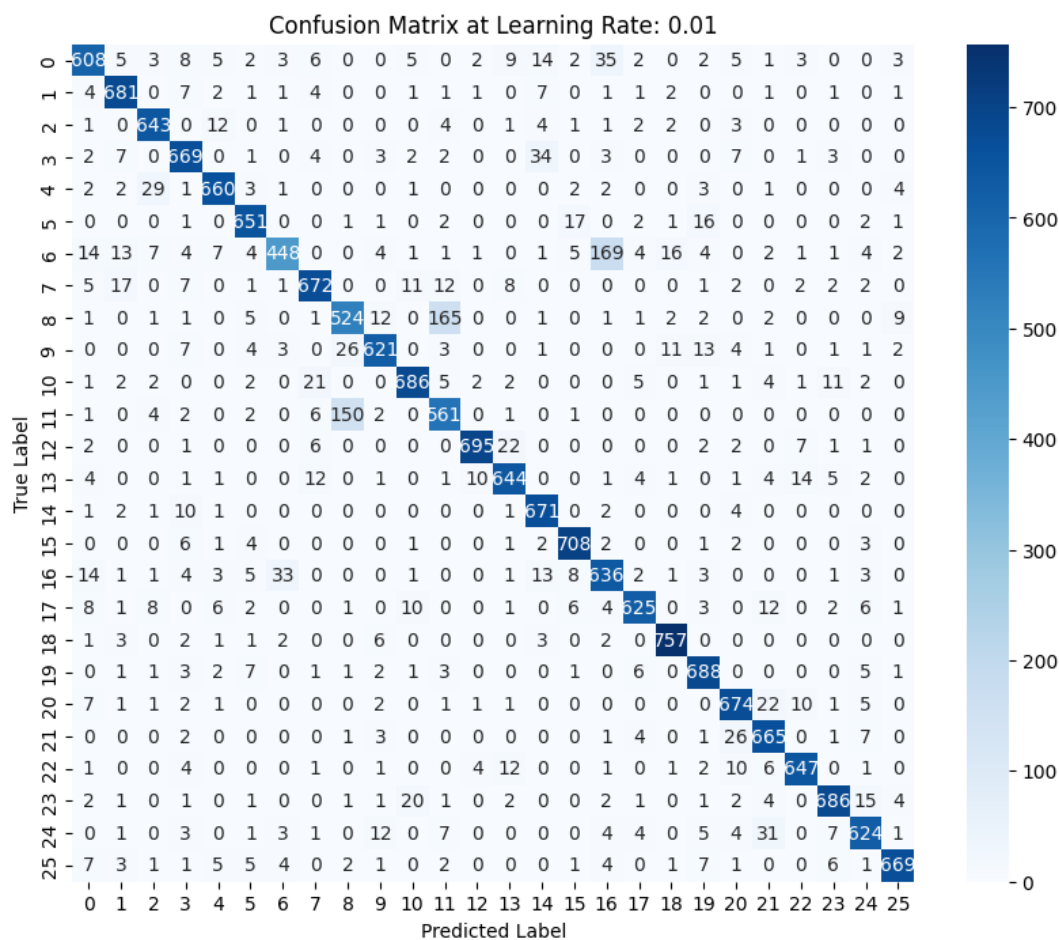
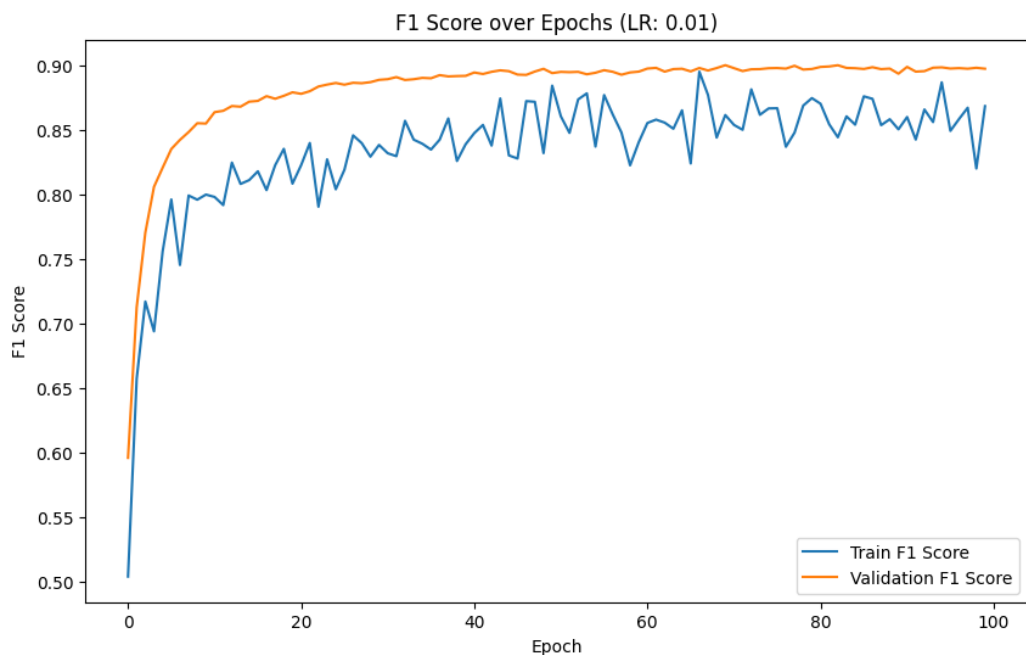
1.3.2 Best Metrics:

learning rate	loss		accuracy		macro_f1	
	training	validation	training	validation	training	validation
0.01	0.439	0.313	0.859	0.901	0.862	0.900
0.005	0.400	0.302	0.870	0.904	0.862	0.904
0.001	0.475	0.350	0.850	0.888	0.855	0.888
0.0005	0.586	0.439	0.819	0.862	0.817	0.862

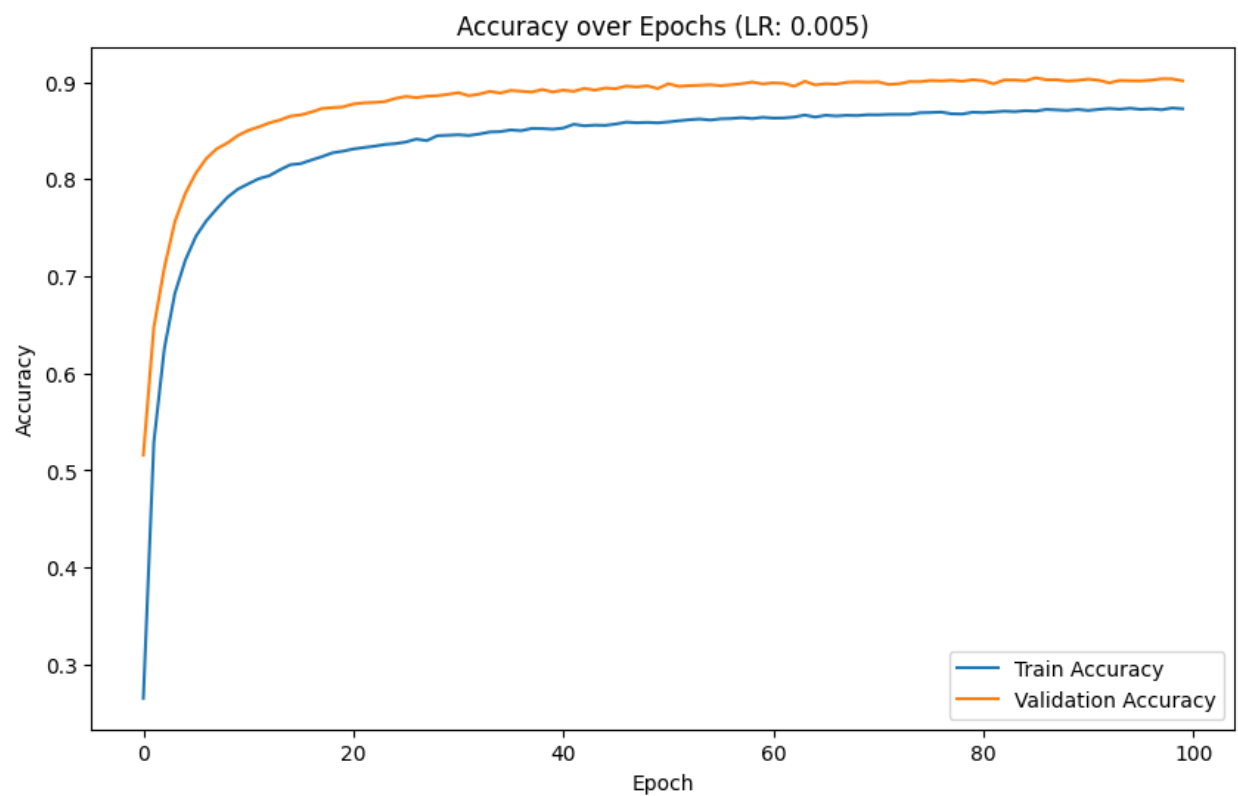
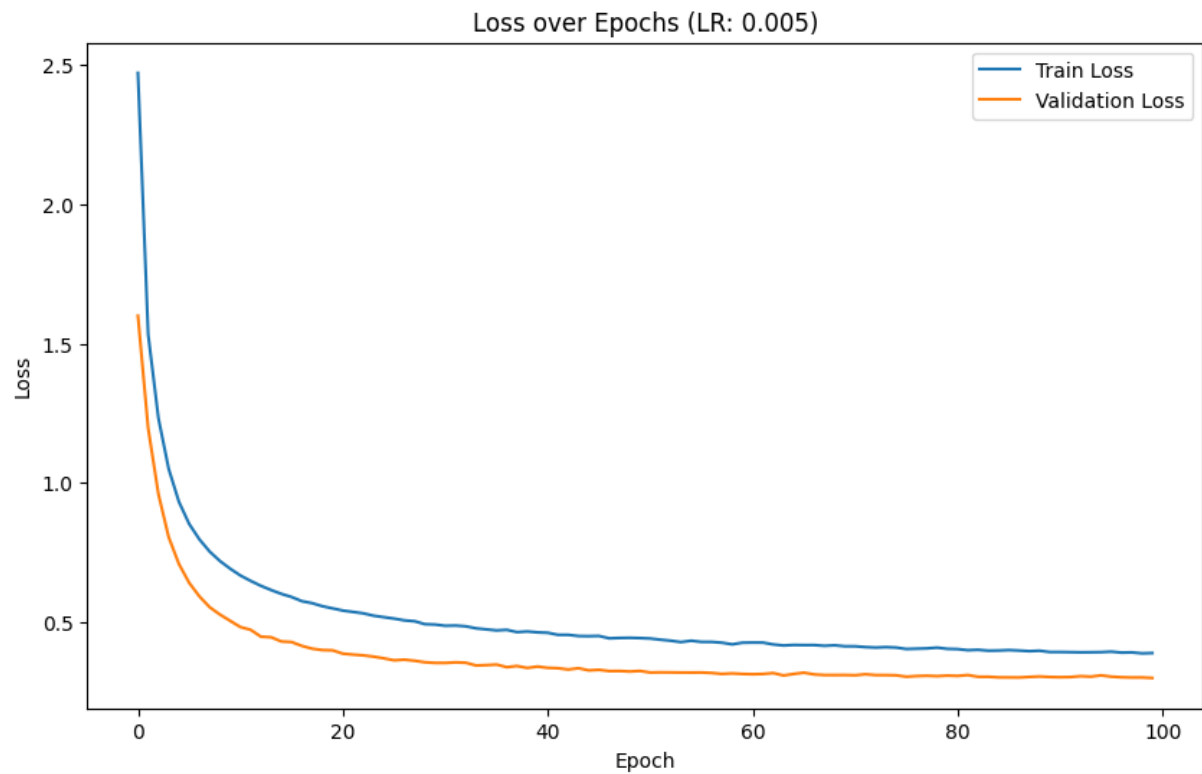
1.3.3 Graphs of Metrics:

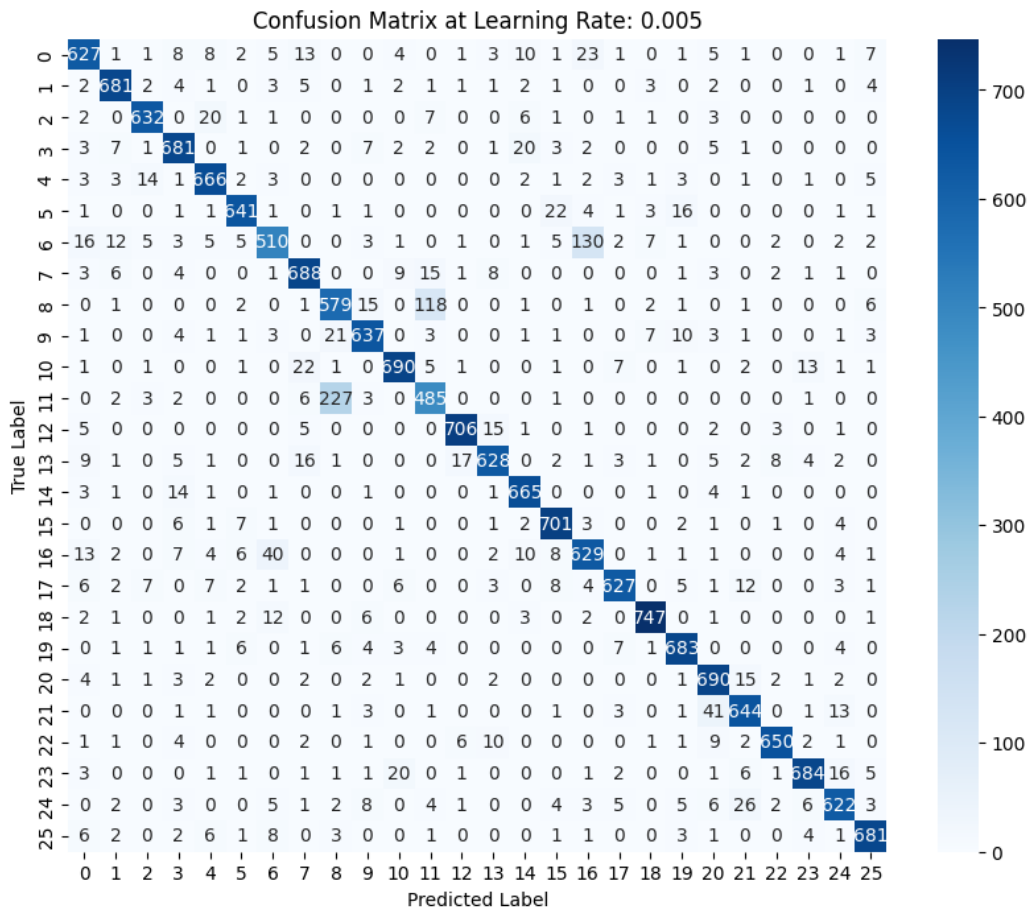
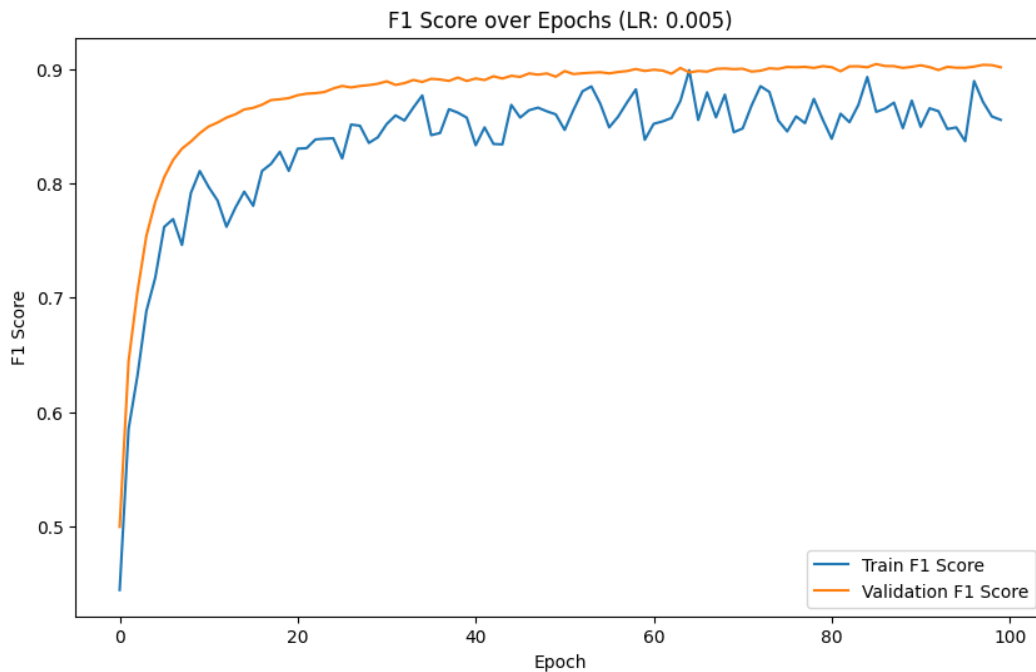
1.3.3.1 LR = 0.01



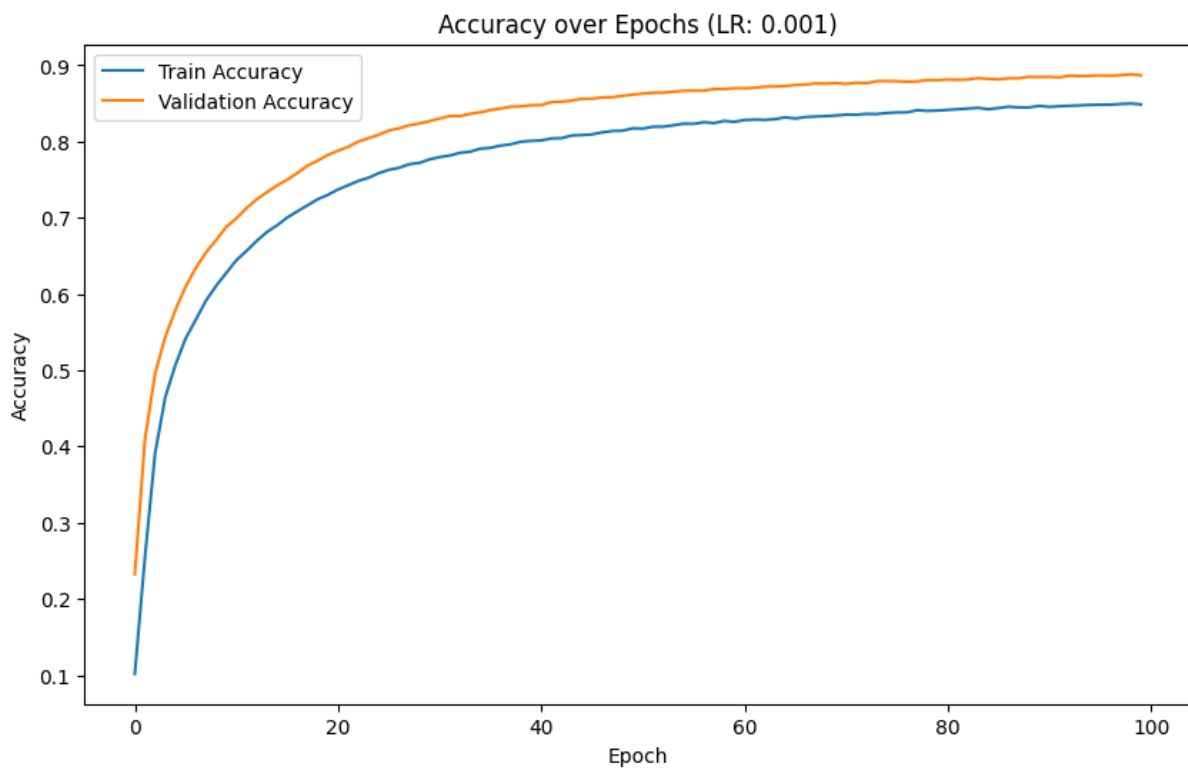
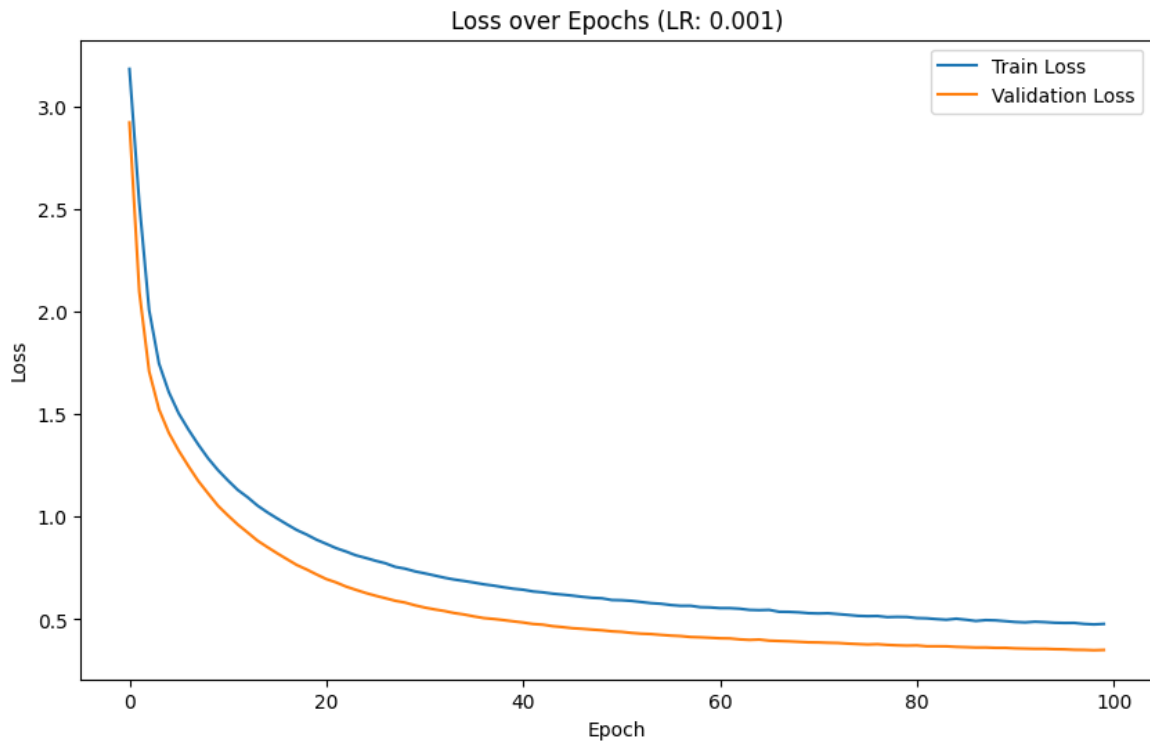


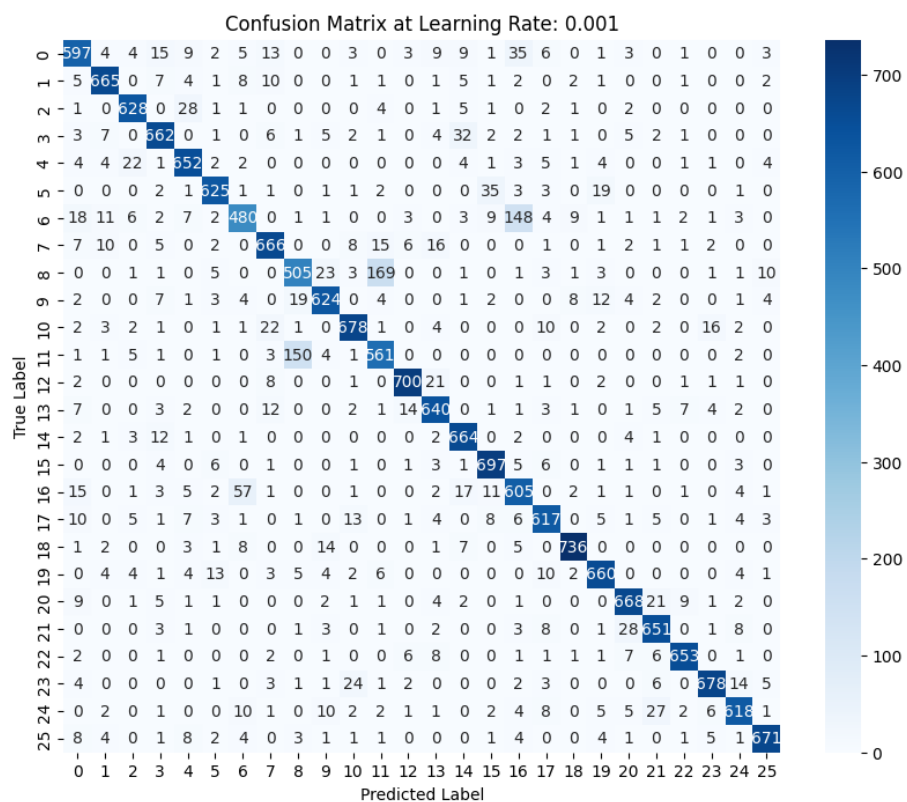
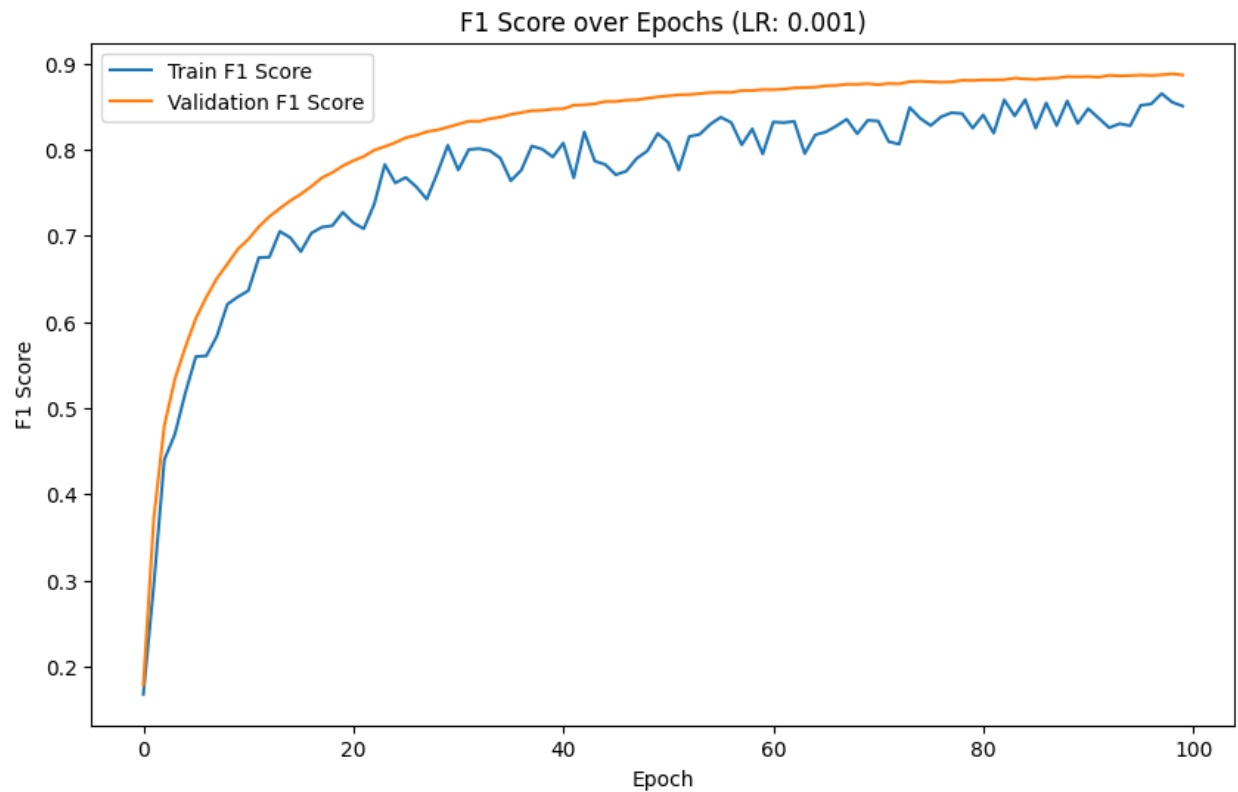
1.3.3.2 LR = 0.005



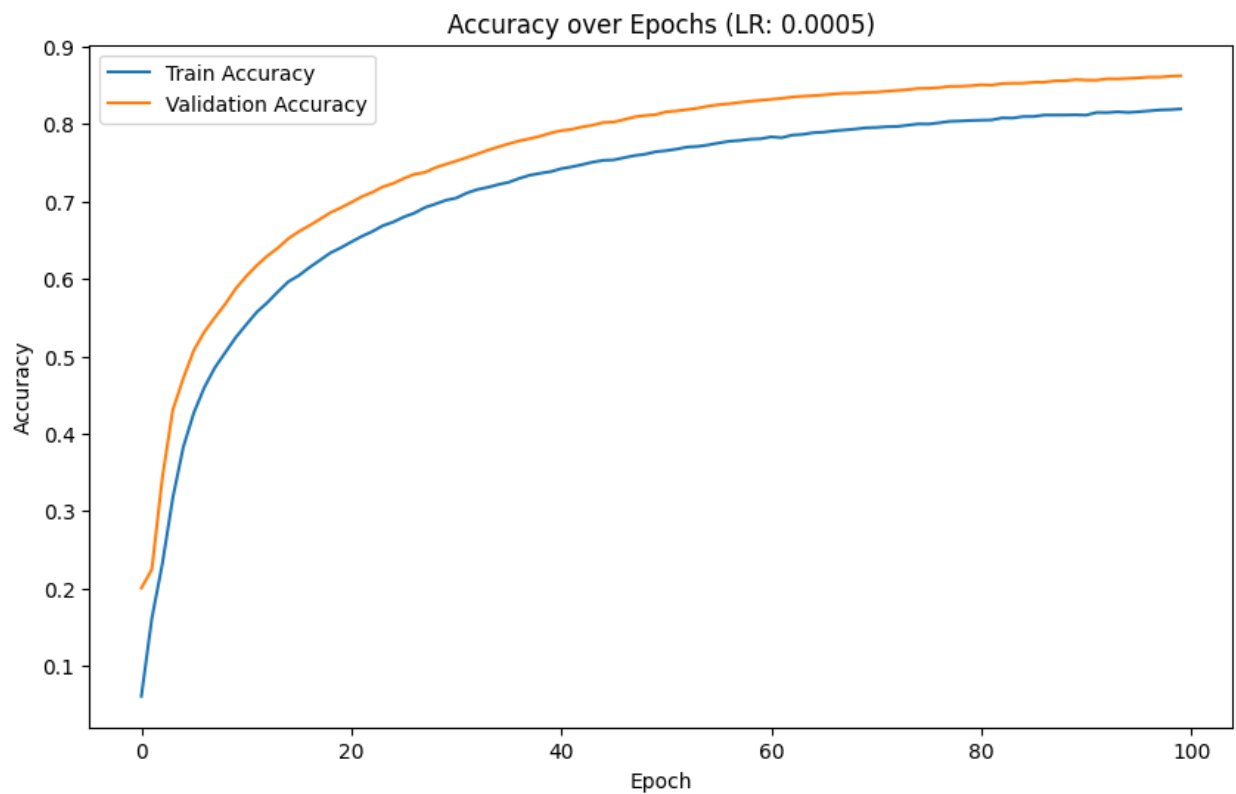
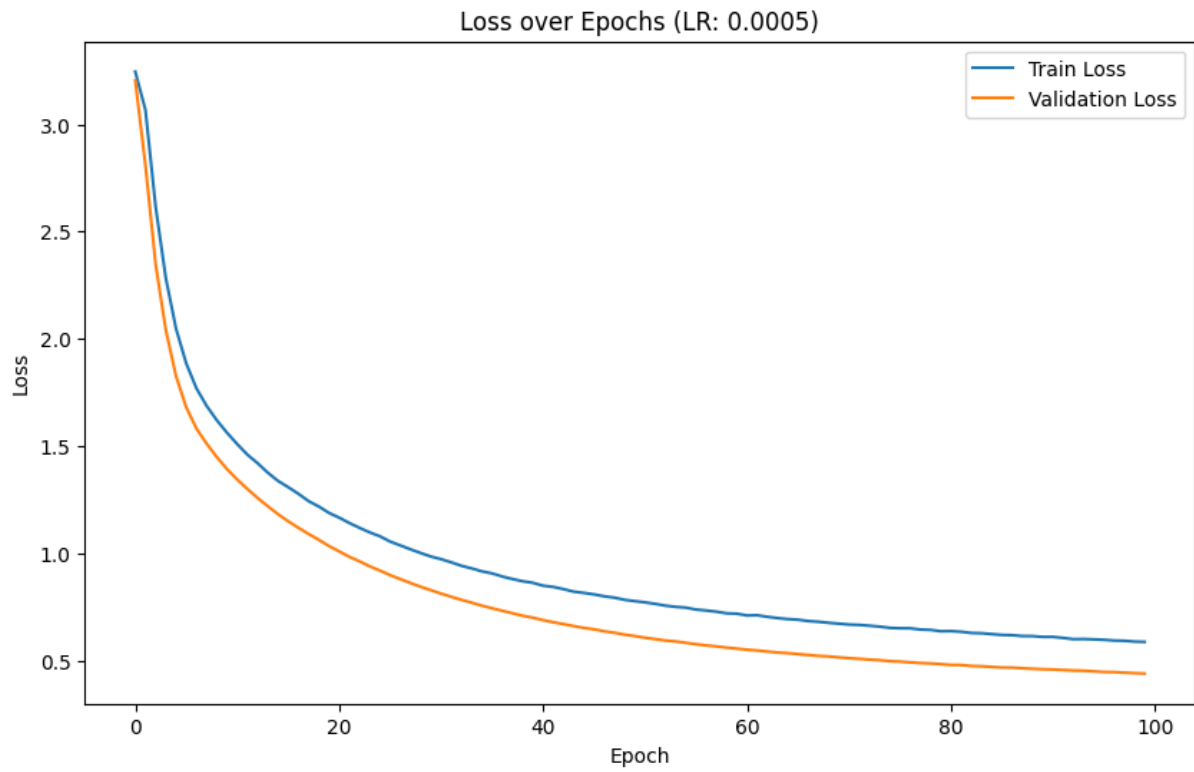


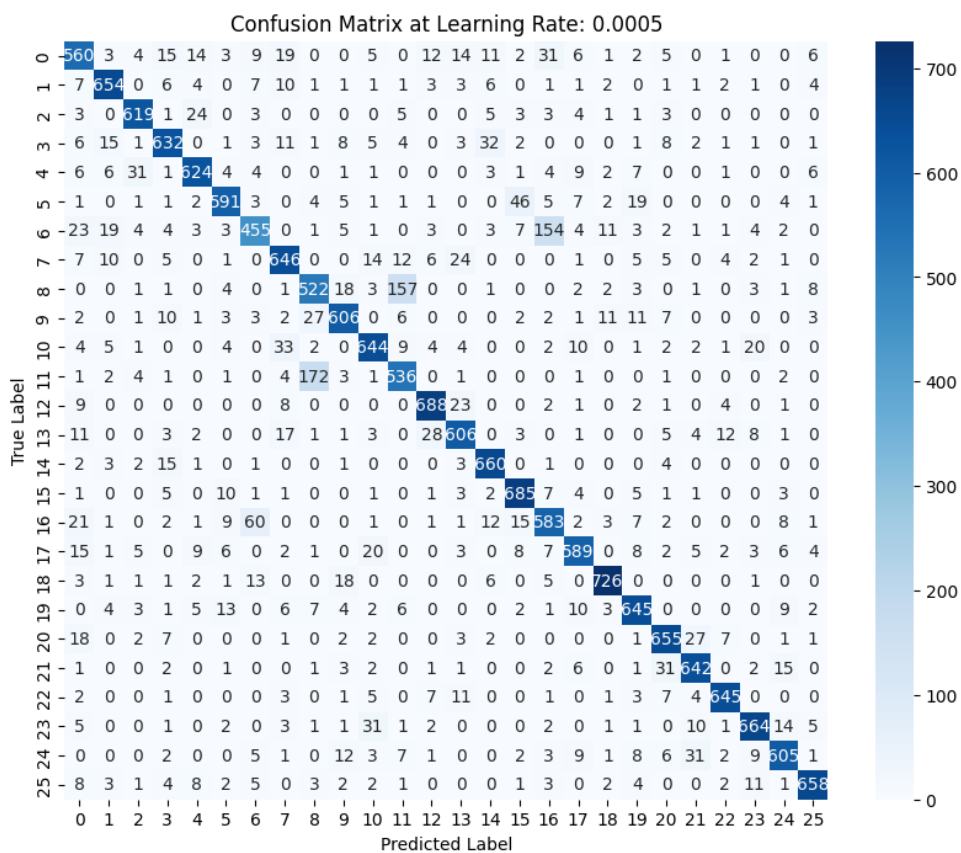
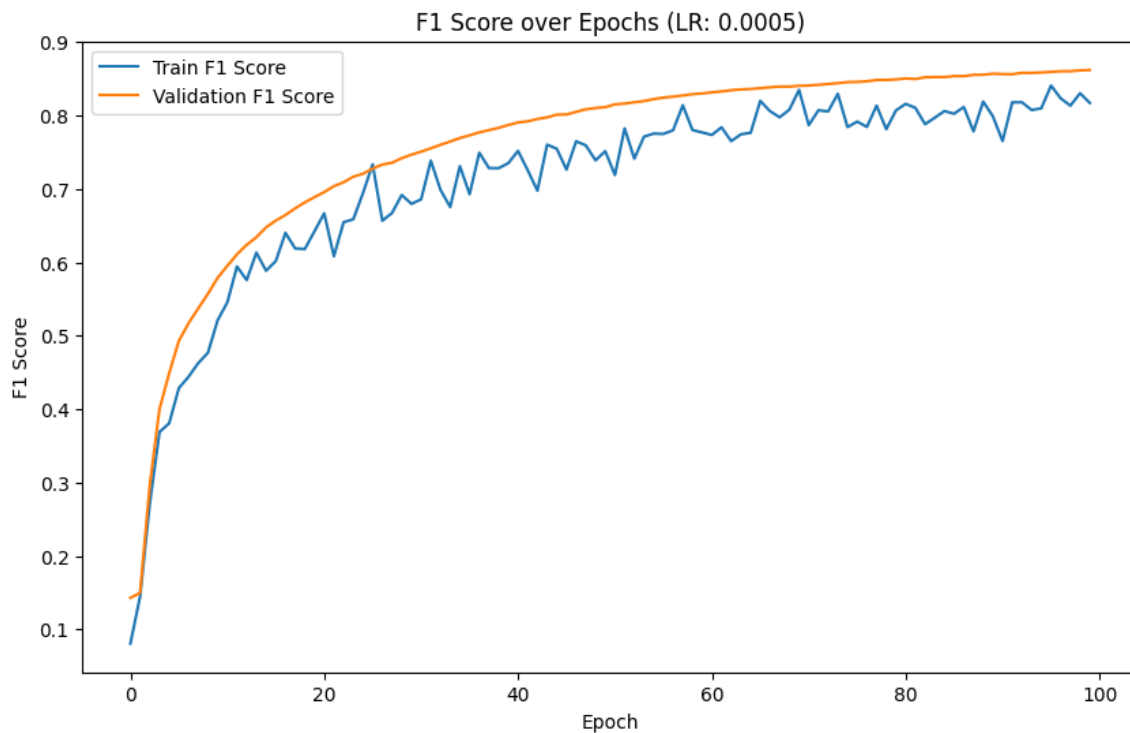
1.3.3.3 LR = 0.001





1.3.3.4 LR = 0.0005

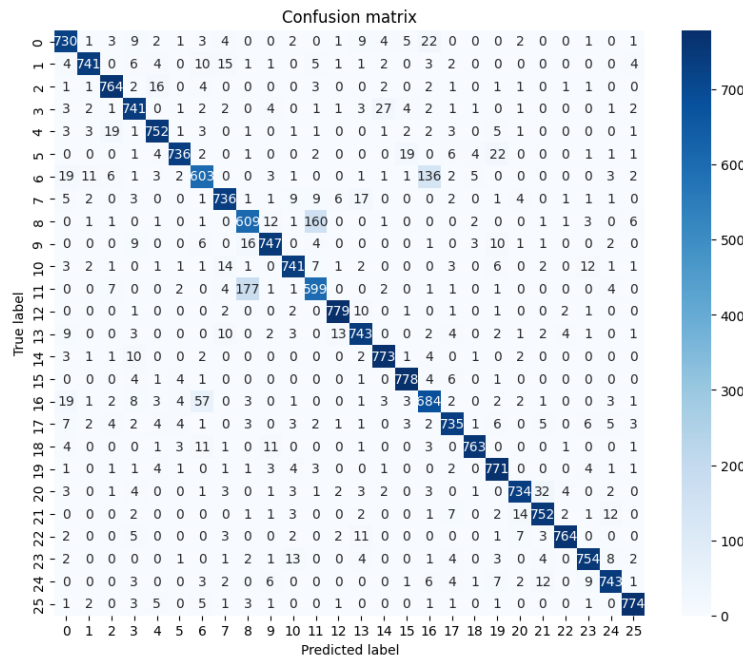




2. Independent Test Performance

(Model 1, LR = 0.01, epochs = 100):

- Test Loss: 0.293
- Test Accuracy: 0.916
- Test Macro f1: 0.916
- Confusion Matrix:



3. Observations:

- 100 epochs were used in each case due to resource constraints. For learning rate 5e-5, the model was very far from converging within this restriction. So after model 1, learning rate of 0.01 was used for all 3 models instead of 5e-5
- Models 1 and 2 could not converge within 100 epoch for lower, as seen from the rising accuracy, and best performance at LR = 0.01. These models both had 2 dense layers
- Model 3 had 3 Dense Layers and converged much quicker. Thus LR = 0.005 had a better accuracy than LR = 0.01
- The train macro f₁ was very unstable since small batches of train data were used to predict this.
- Almost all models had the most difficulty in predicting the letters ("L"(11), "I"(8)) and the letters ("G"(6), "Q"(16))