# REPORT

1805061 – Noshin Nawal

## Hyperparameters

1. Batch size = 1024
2. Epoch = 15
3. Default ADAM parameters (beta1 = 0.9, beta2 = 0.999, epsilon=1e-8)
4. Xavier initialization
5. Linear learning rate scheduling (initial LR varying from 5e-3 to 5e-5, final LR 5e-6)

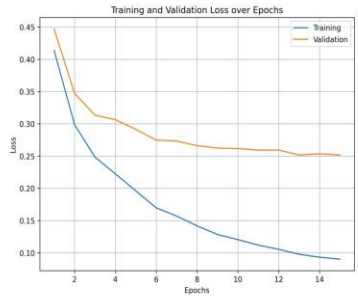
## Model 1 (Best Model)

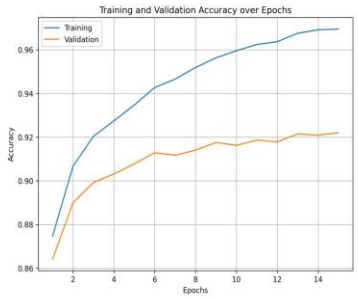
```
network = [
    Dense(28*28, 1024),
    ReLU(),
    Dropout(0.3),
    Dense(1024, 26),
    Softmax()
]
```

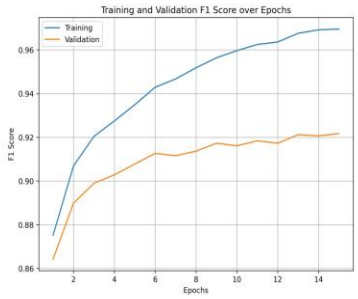
**Initial LR = 5e-3** Test Accuracy = 0.9203 Test F1 = 0.9204

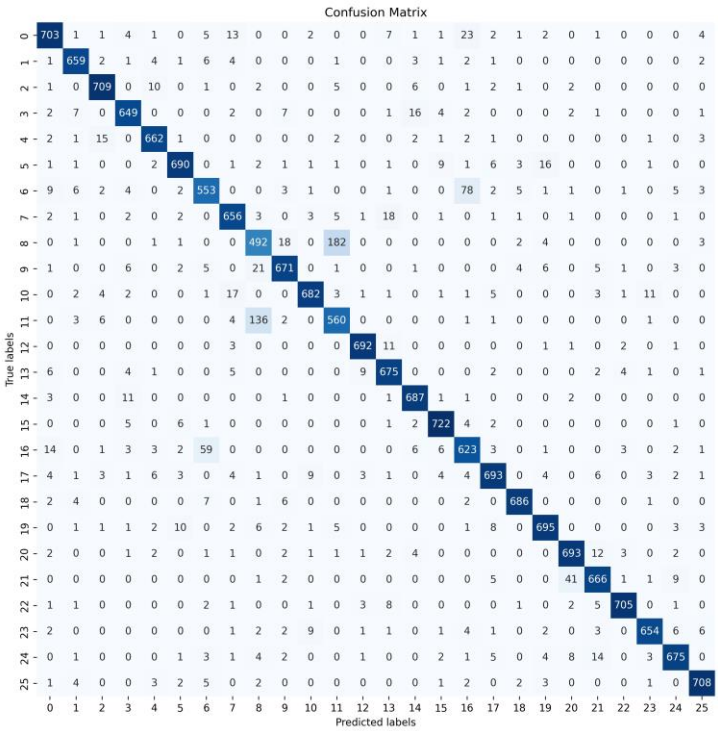
Train Loss = 0.0903 Train Accuracy = 0.9695 Train F1 = 0.9695

Val Loss = 0.2517 Val Accuracy = 0.9220 Val F1 = 0.9217

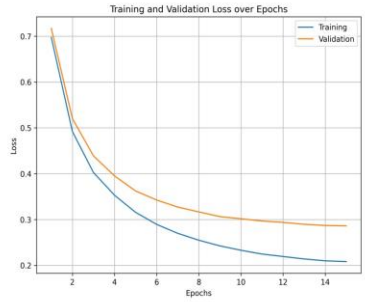


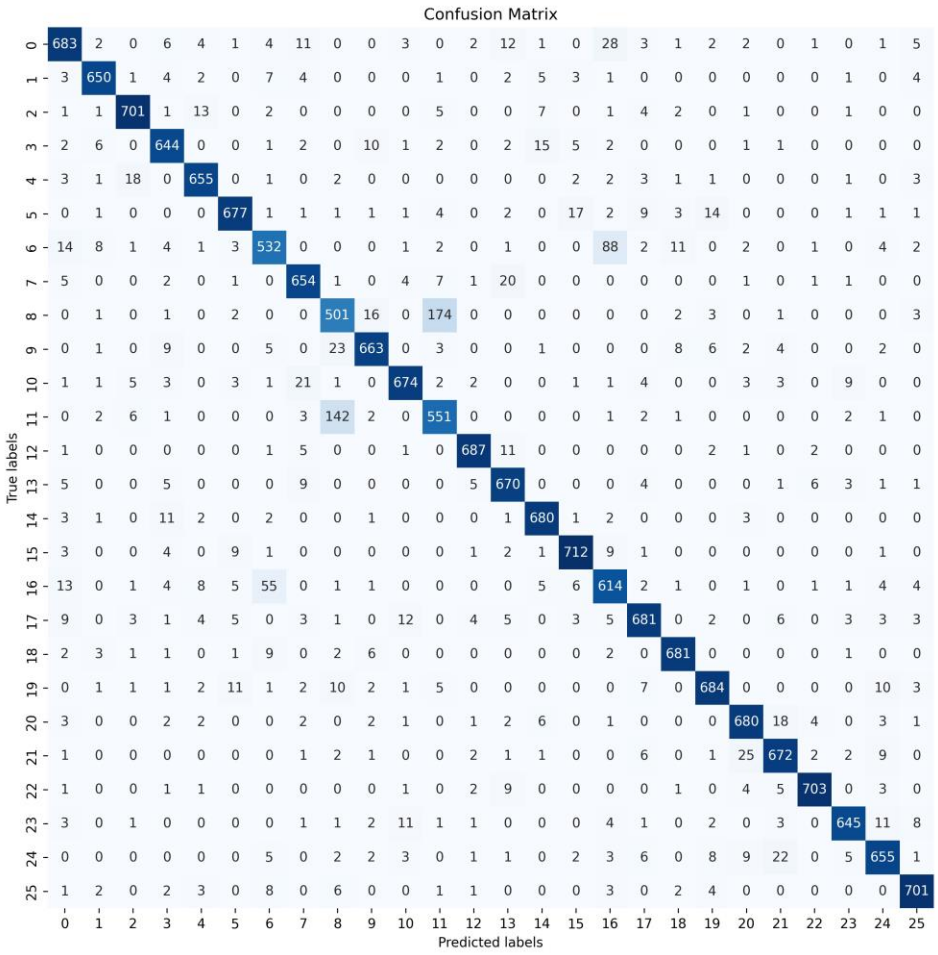



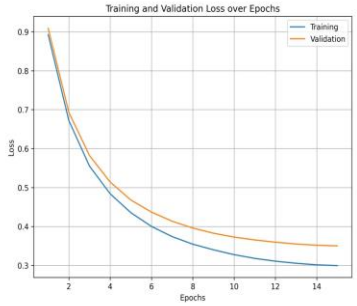
**Initial LR = 1e-3** Test Accuracy = 0.9084 Test F1 = 0.9085

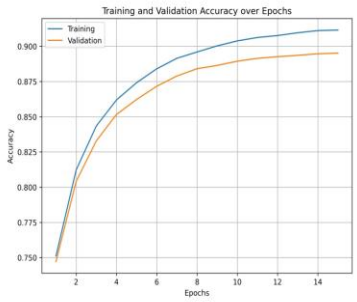


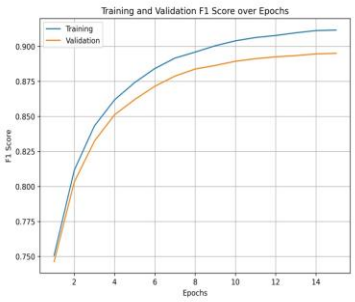


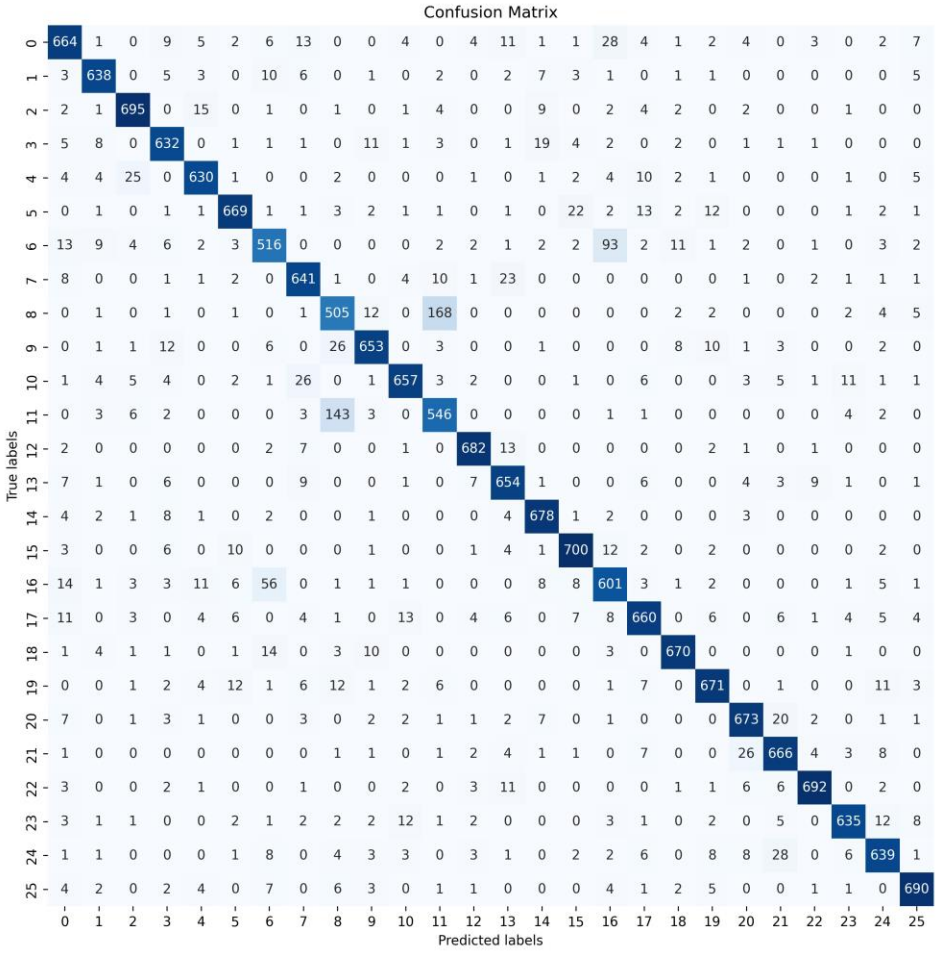



**Initial LR = 5e-4** Test Accuracy = 0.8922 Test F1 = 0.8923

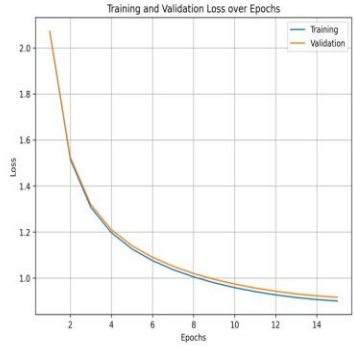


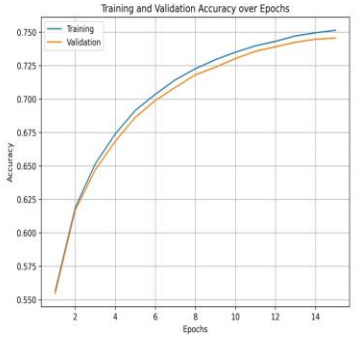



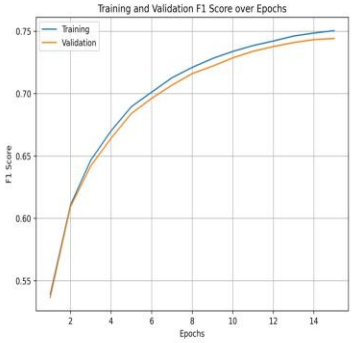



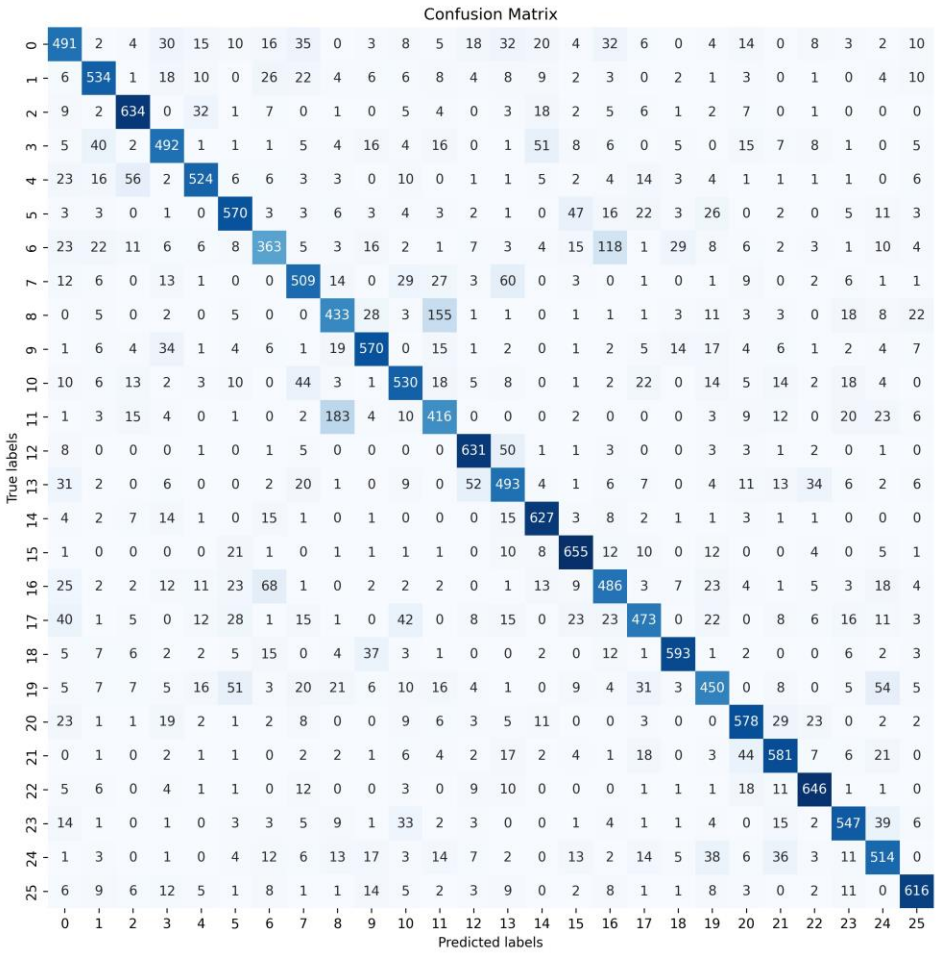



**Initial LR = 5e-5** Test Accuracy = 0.7472 Test F1 = 0.7462



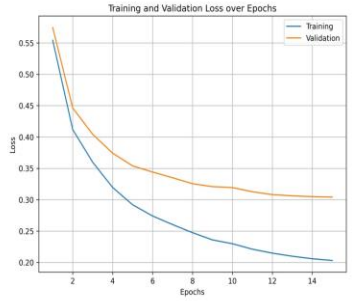



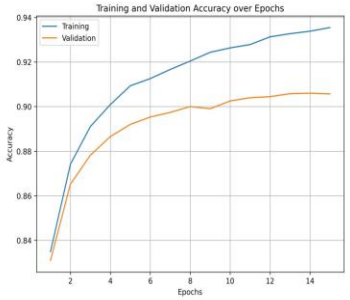



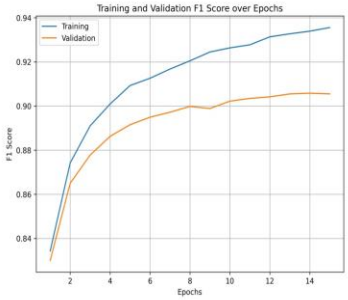
# Model 2

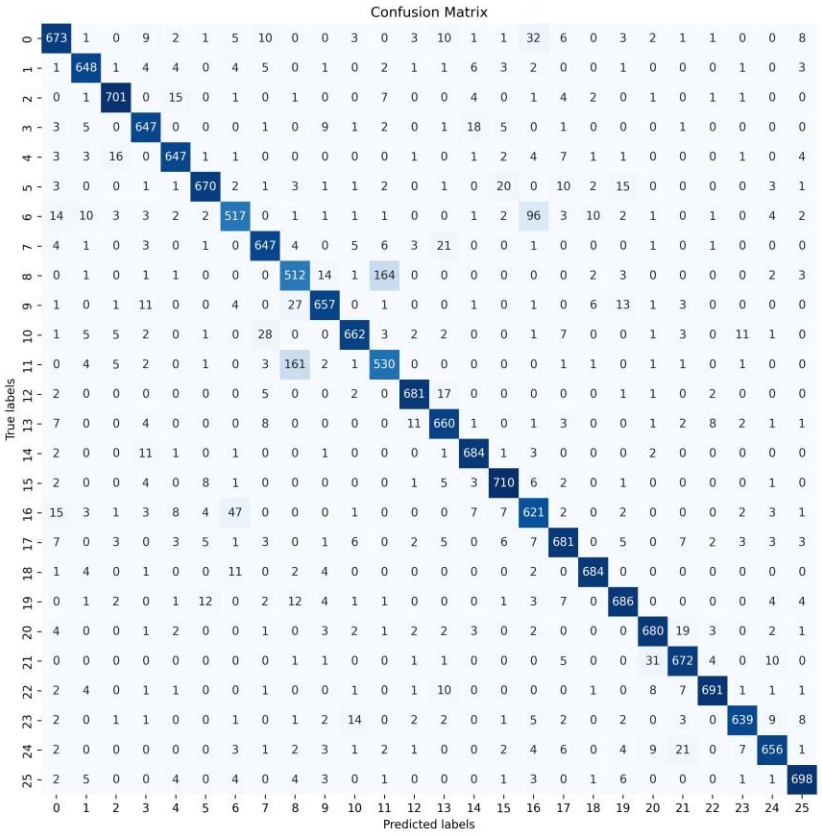
```
network = [
    Dense(28*28, 256),
    ReLU(),
    Dropout(0.3),
    Dense(256, 26),
    Softmax()
]
```

## Initial LR = 5e-3

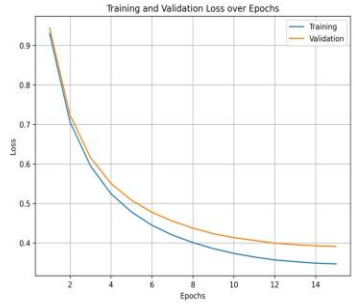


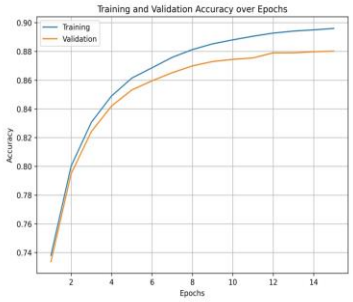



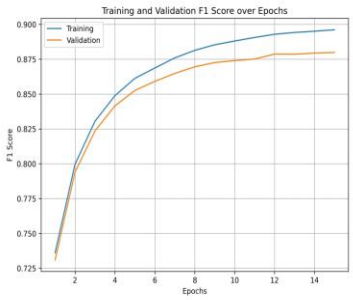



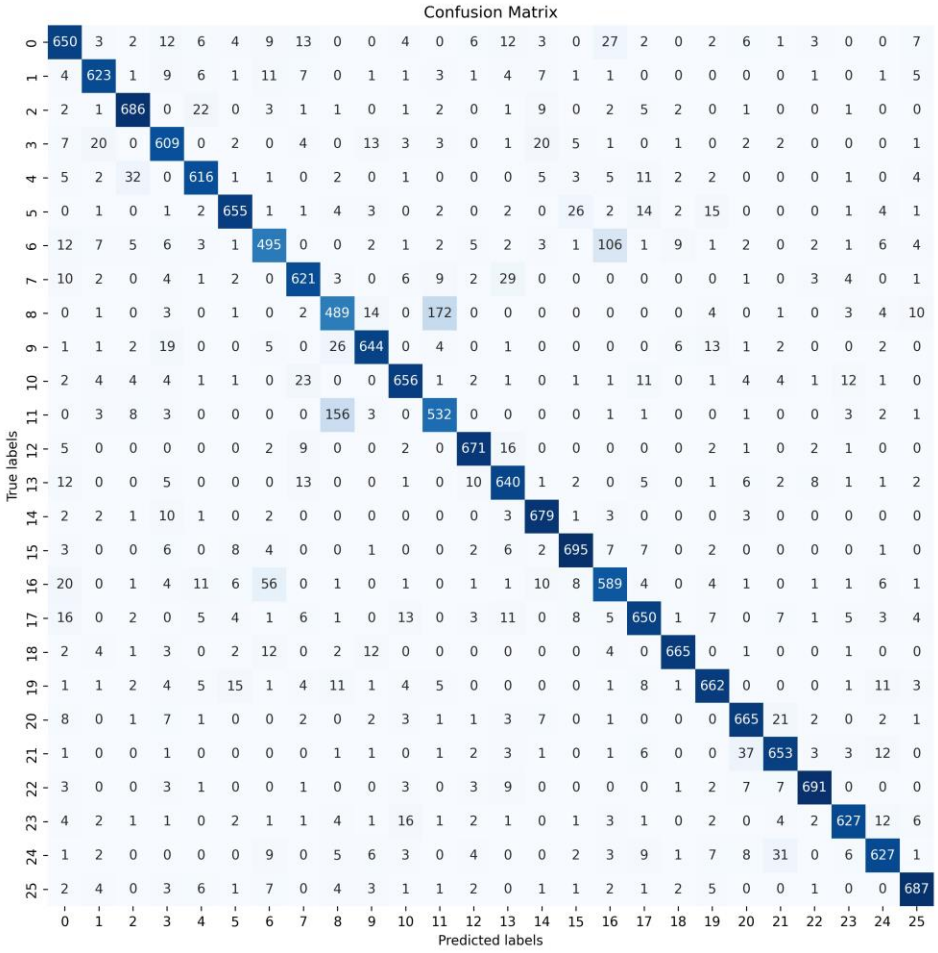



## Initial LR = 1e-3

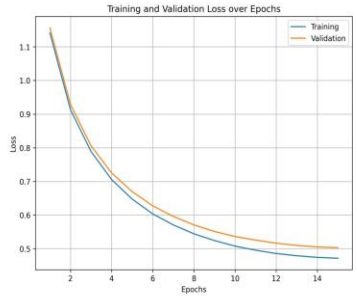


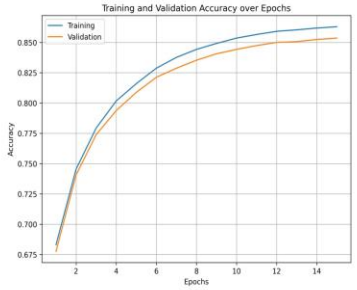



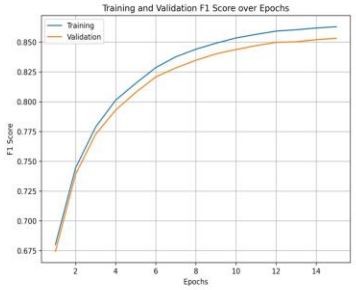



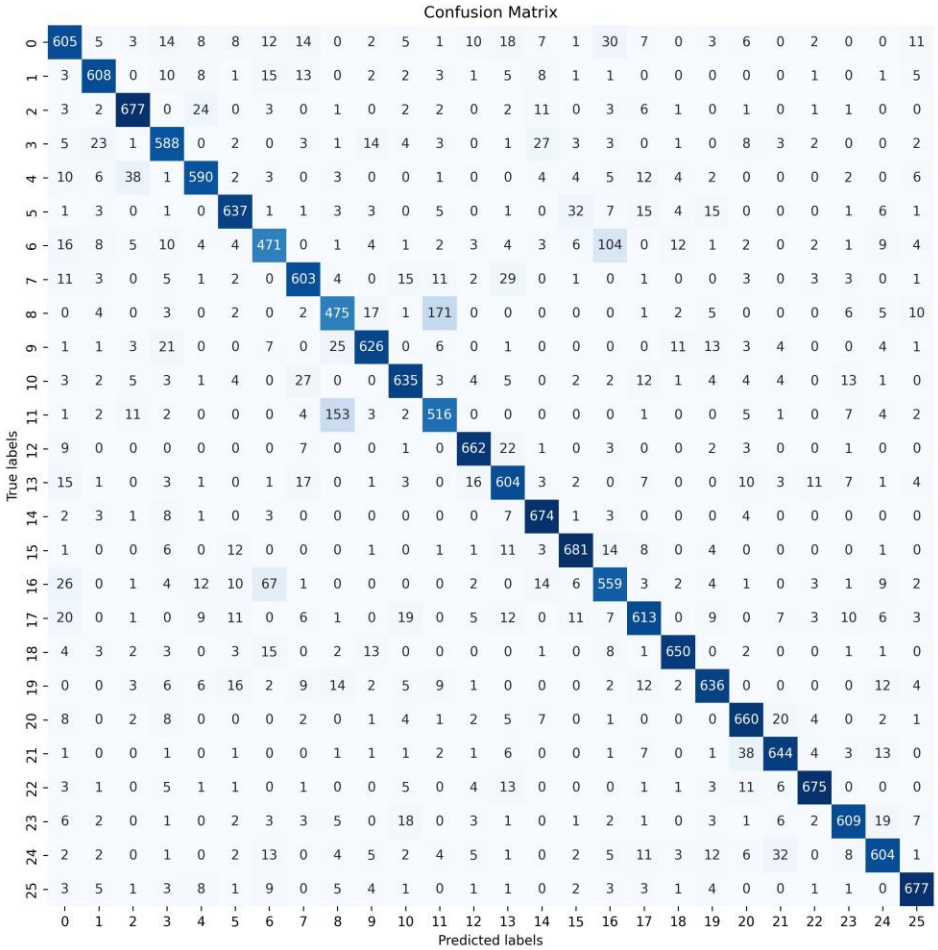



## Initial LR = 5e-4

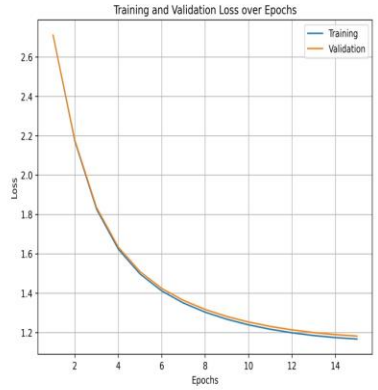


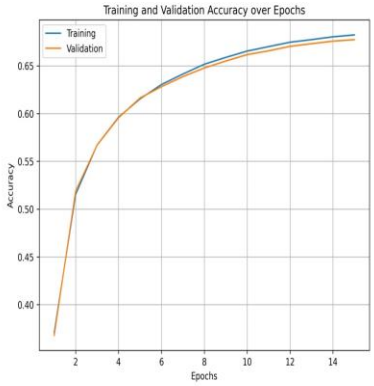



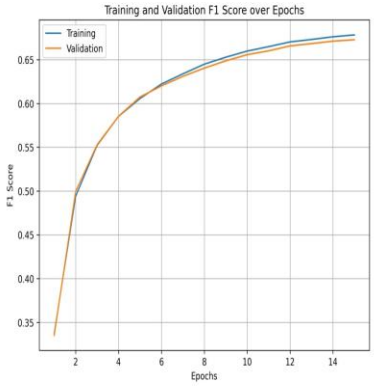



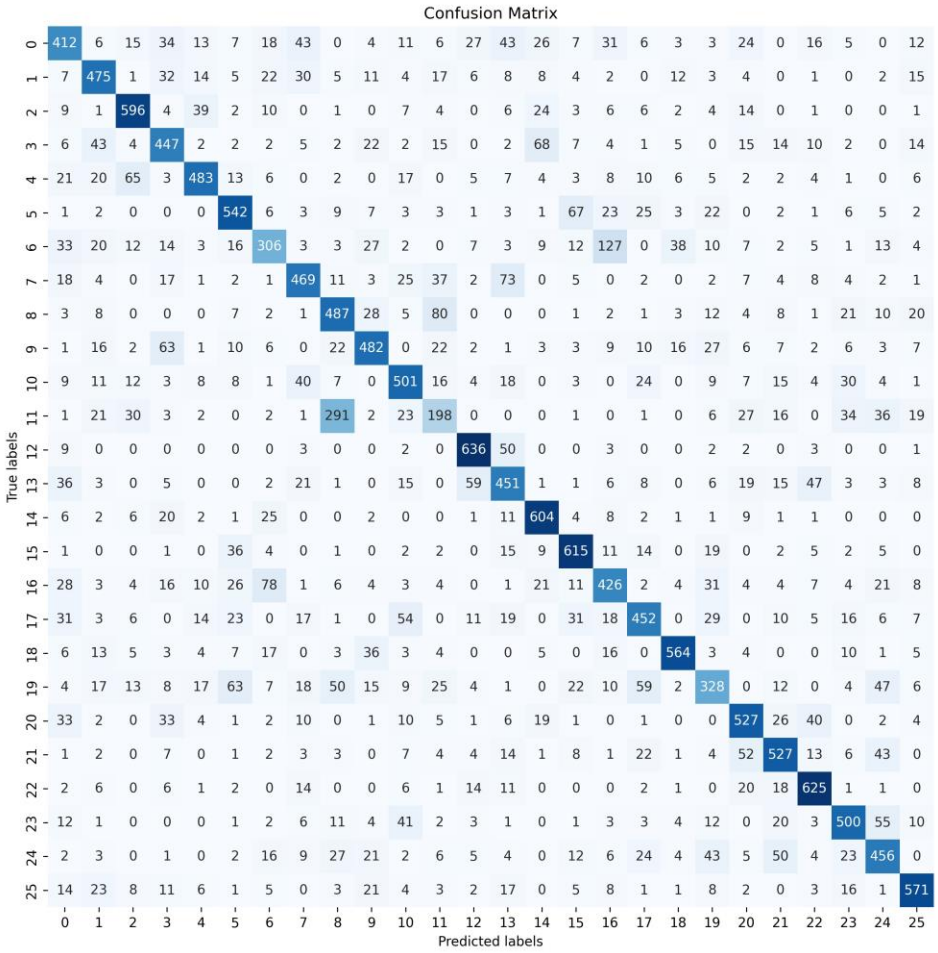



**Initial LR = 5e-5**



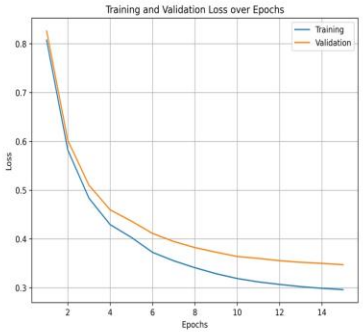



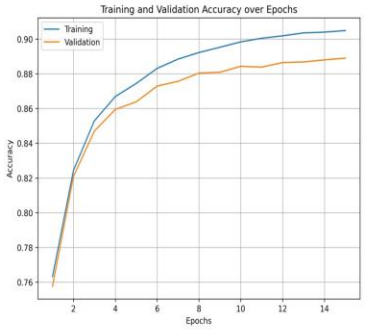



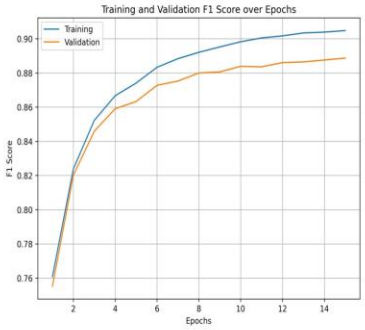
# Model 3

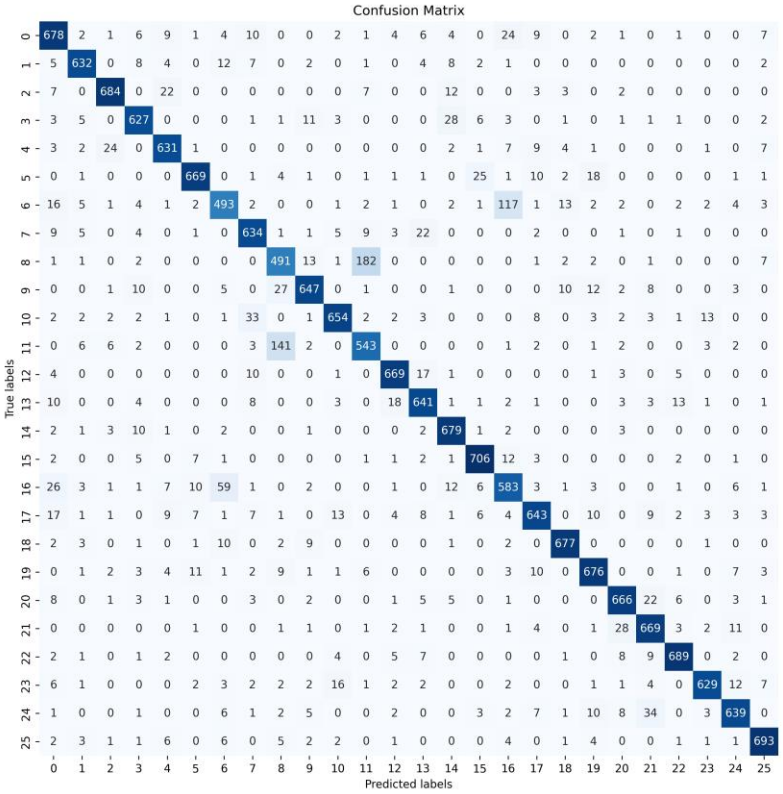
```
network = [
    Dense(28*28, 128),
    ReLU(),
    Dropout(0.2),
    Dense(128, 64),
    ReLU(),
    Dropout(0.1),
    Dense(64, 26),
    Softmax()
]
```

## Initial LR = 5e-3

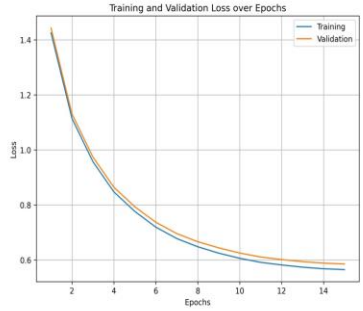


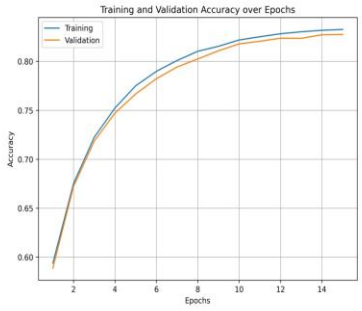



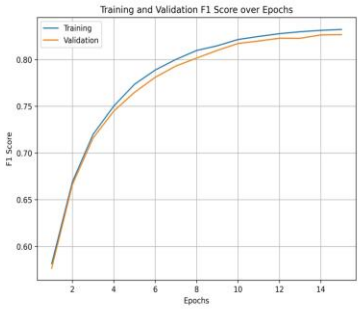



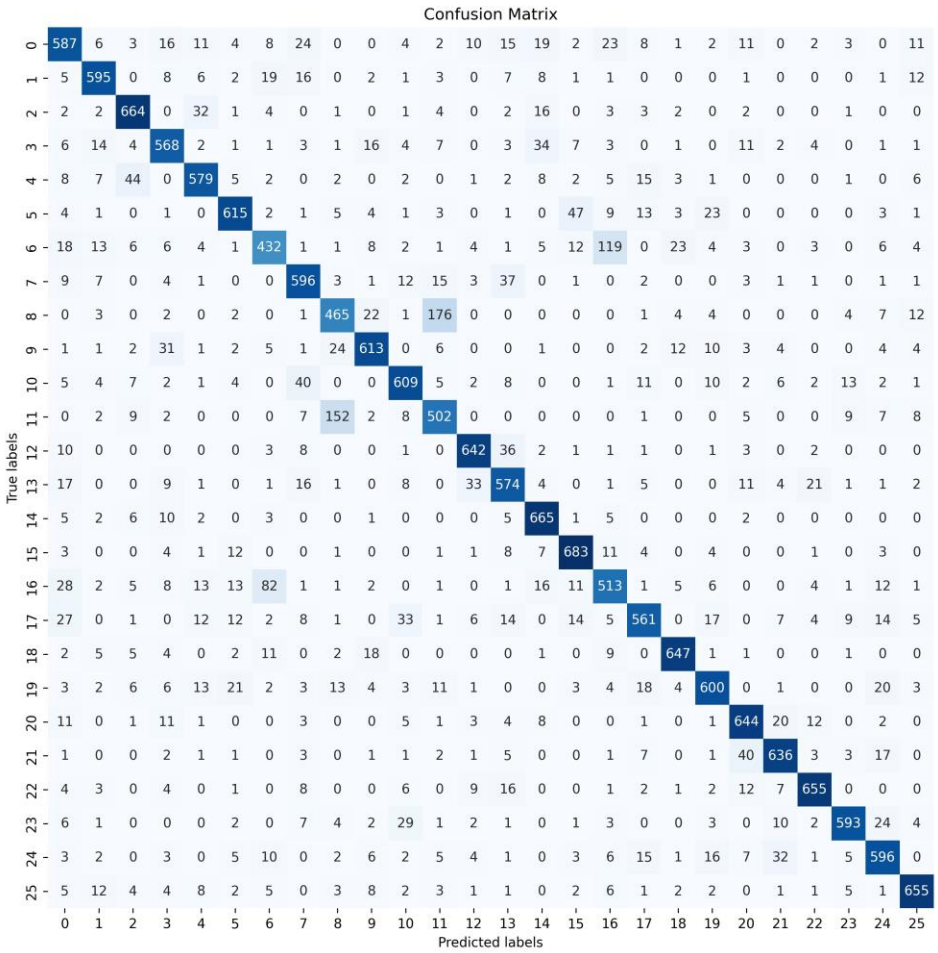



## Initial LR = 1e-3

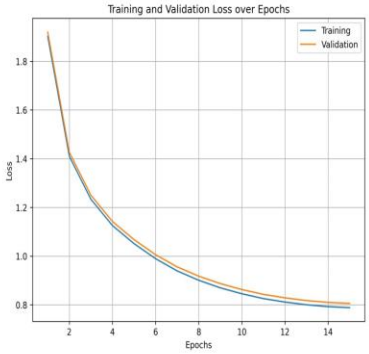


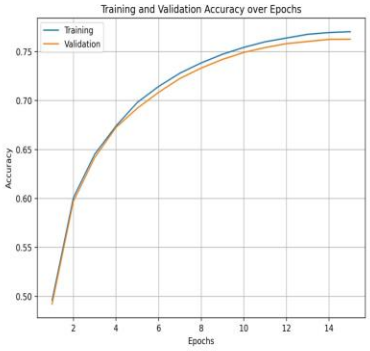



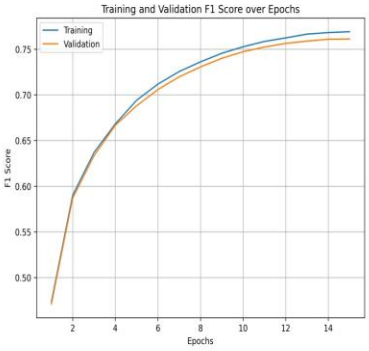



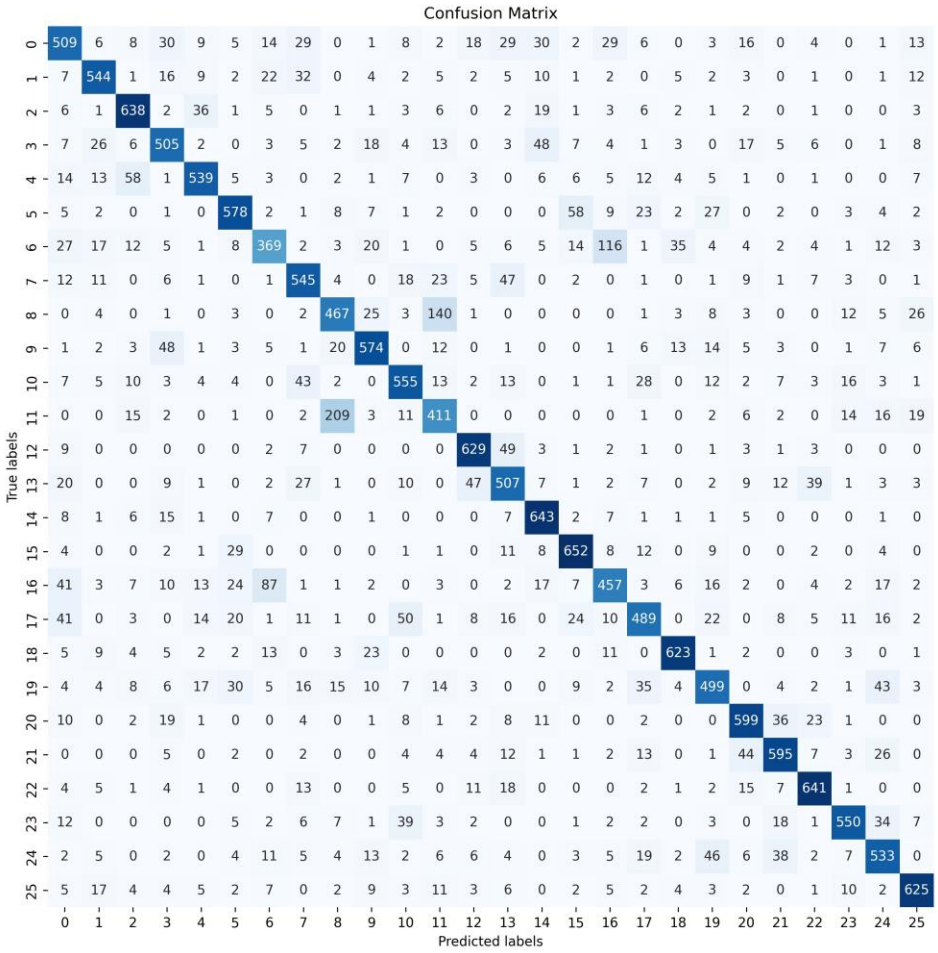



**Initial LR = 5e-4**



## Initial LR = 5e-5