CSE 472

Offline 3

Md Nafiu Rahman

1905077

Model 0

Architecture

DenseLayer(784, 256),

ReLULayer(),

BatchNormLayer(256),

DropoutLayer(0.5),

DenseLayer(256, 128),

ReLULayer(),

BatchNormLayer(128),

DropoutLayer(0.5),

DenseLayer(128, 64),

ReLULayer(),

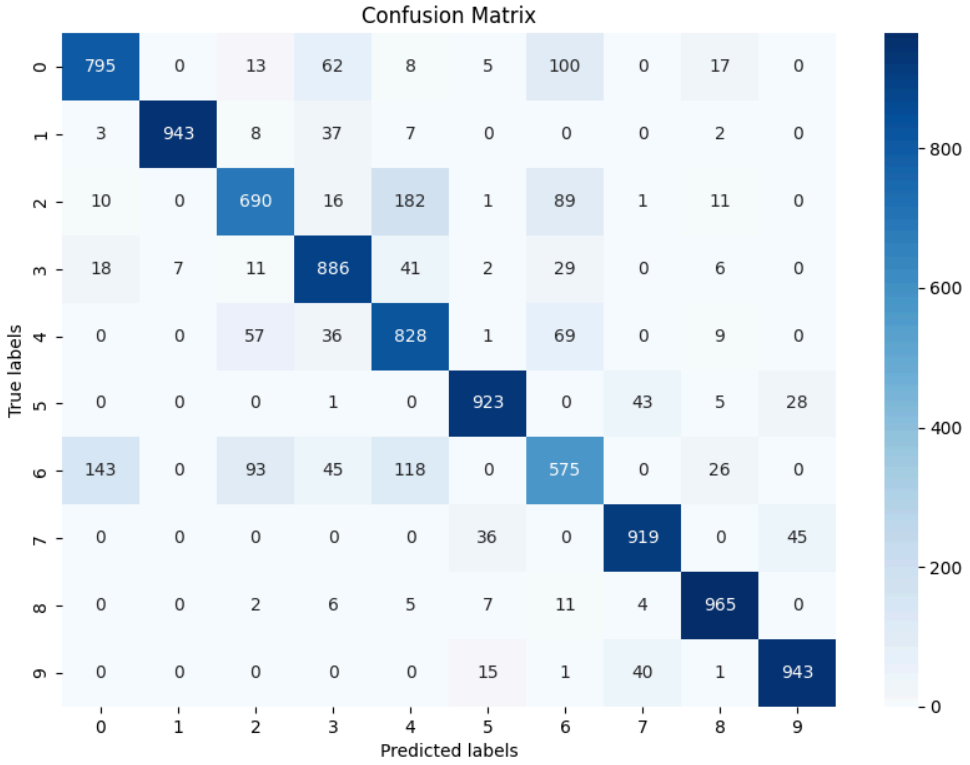
BatchNormLayer(64),

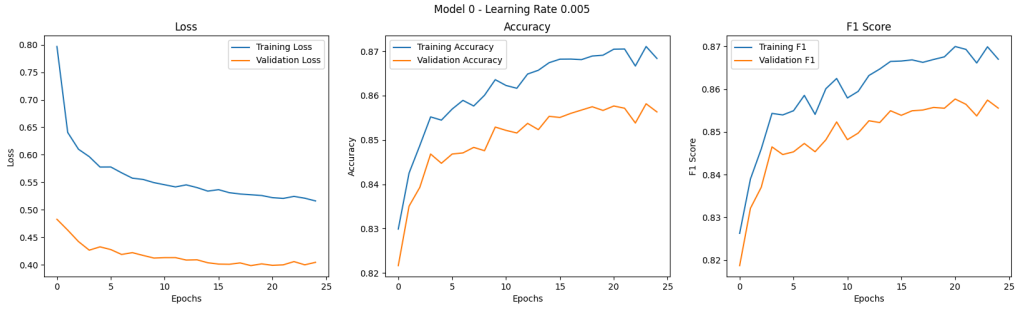
DropoutLayer(0.5),

DenseLayer(64, 10),

SoftmaxLayer()

Learning Rate 0.005





Training Loss: 0.5159877103750761

Validation Loss: 0.4043047502936237

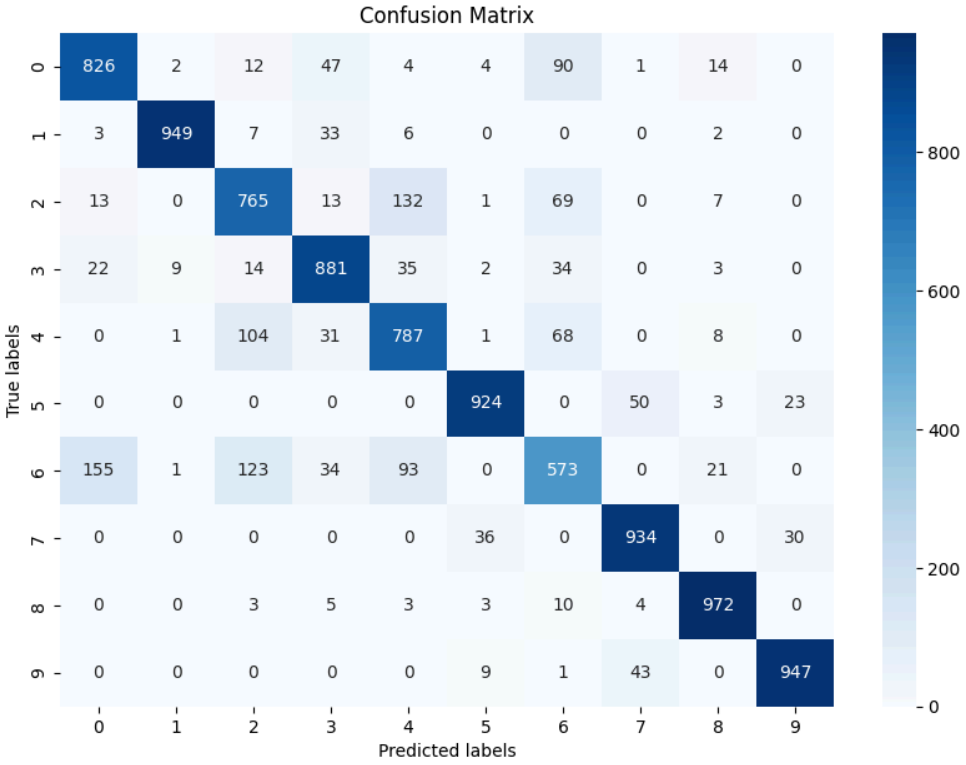
Training Accuracy: 0.8683958333333334

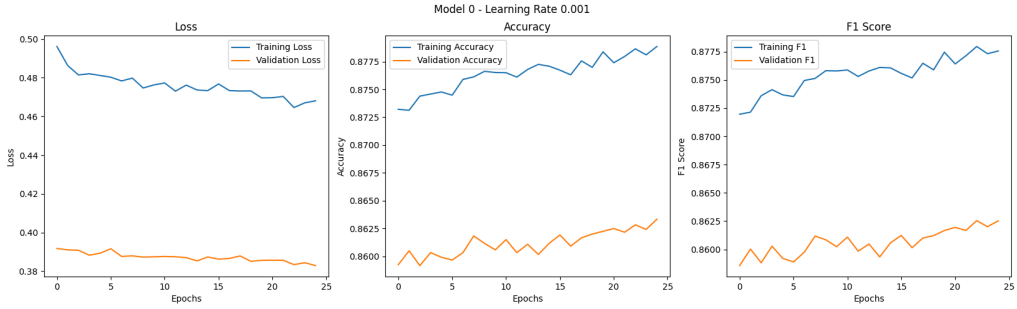
Validation Accuracy: 0.8563333333333333

Training F1: 0.8669872325268748

Validation F1: 0.8555505745497765

Learning Rate 0.001





Training Loss: 0.4680918799434667

Validation Loss: 0.38289580956765407

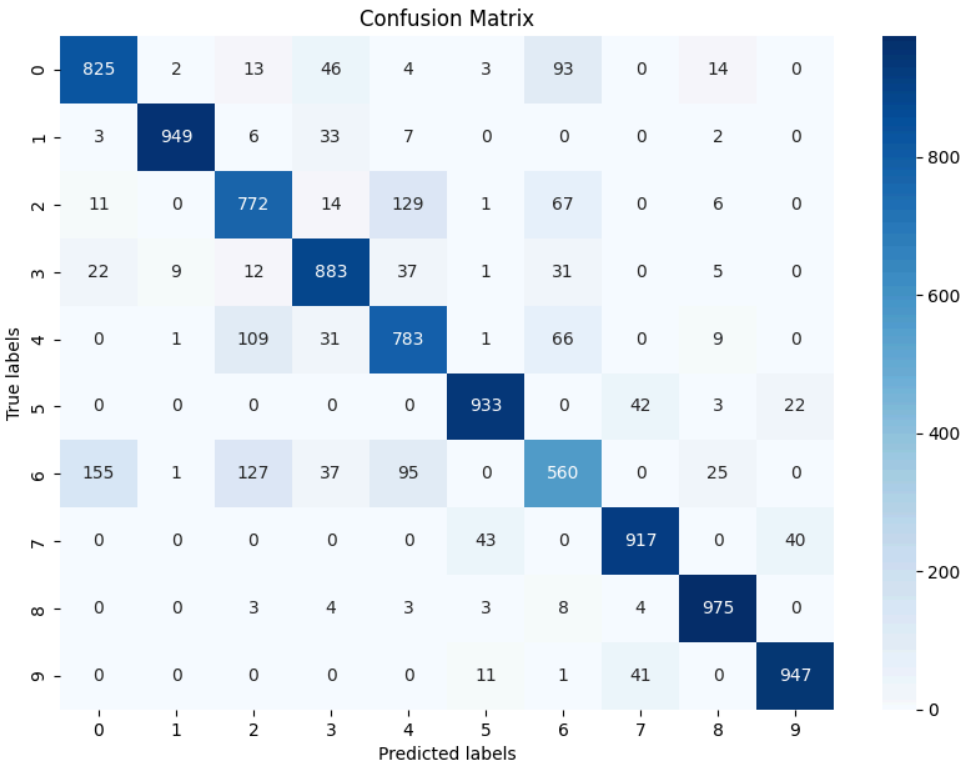
Training Accuracy: 0.8788541666666667

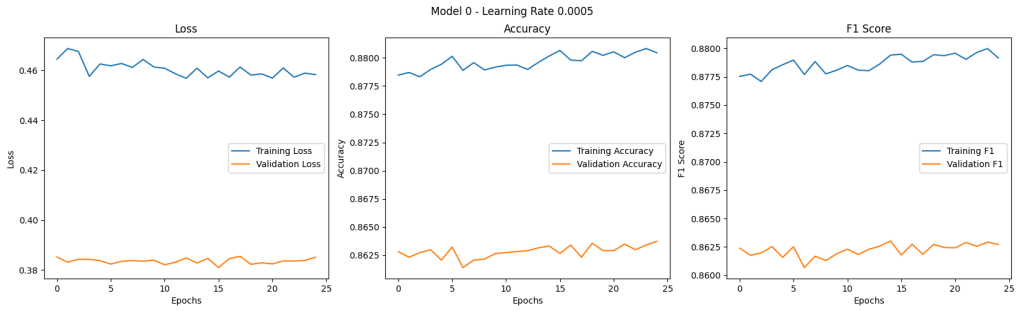
Validation Accuracy: 0.8633333333333333

Training F1: 0.877552488753915

Validation F1: 0.8625258324923383

Learning Rate 0.0005





Training Loss: 0.4582507614096216

Validation Loss: 0.38514712412206353

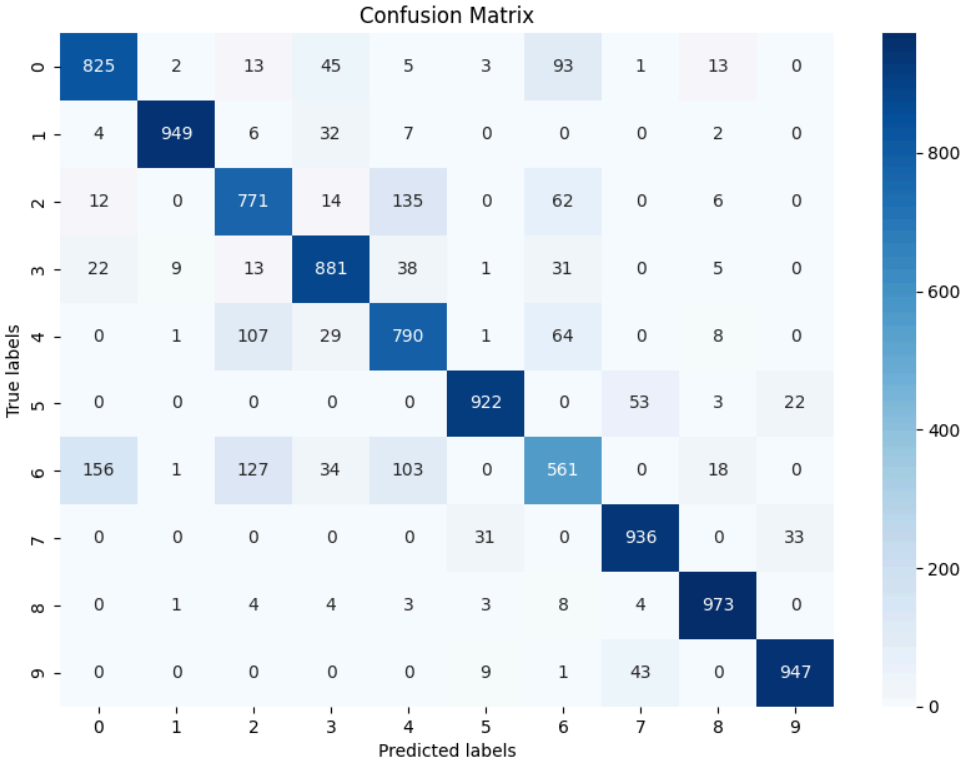
Training Accuracy: 0.8804375

Validation Accuracy: 0.86375

Training F1: 0.879159240789438

Validation F1: 0.8627243368310259

Learning Rate 0.0001



Training Loss: 0.45160981523928967

Validation Loss: 0.3815737642439396

Training Accuracy: 0.8811666666666667

Validation Accuracy: 0.86375

Training F1: 0.8799067338467477

Validation F1: 0.8628542815013482

Model 1

Architecture

DenseLayer(784, 512),

ReLULayer(),

BatchNormLayer(512),

DropoutLayer(0.5),

DenseLayer(512, 256),

ReLULayer(),

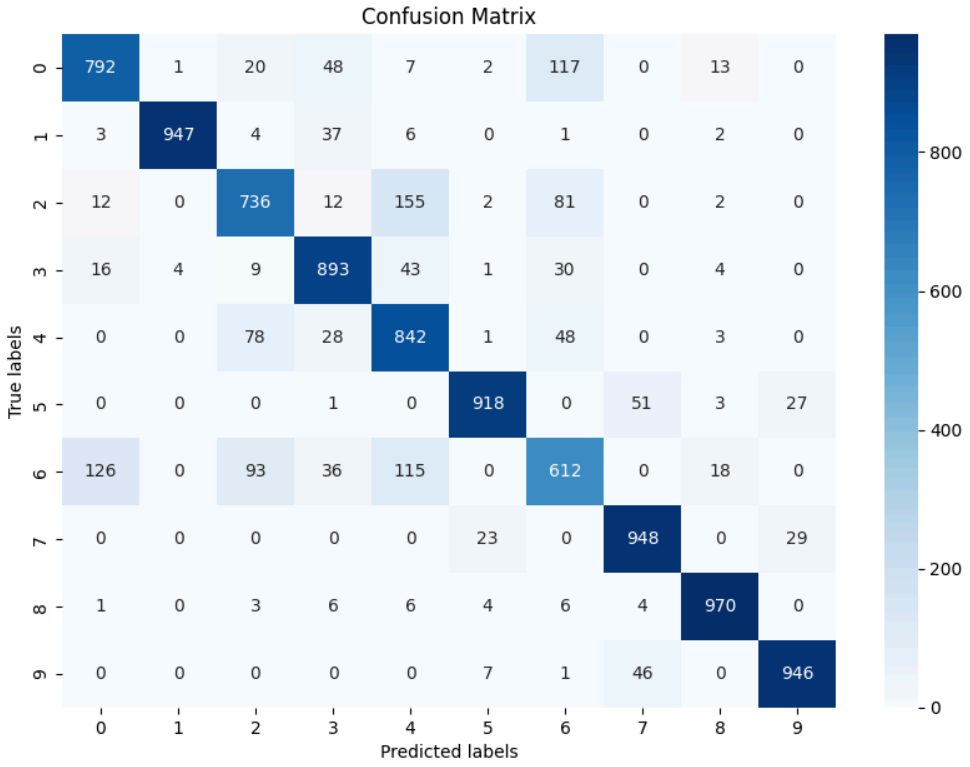
BatchNormLayer(256),

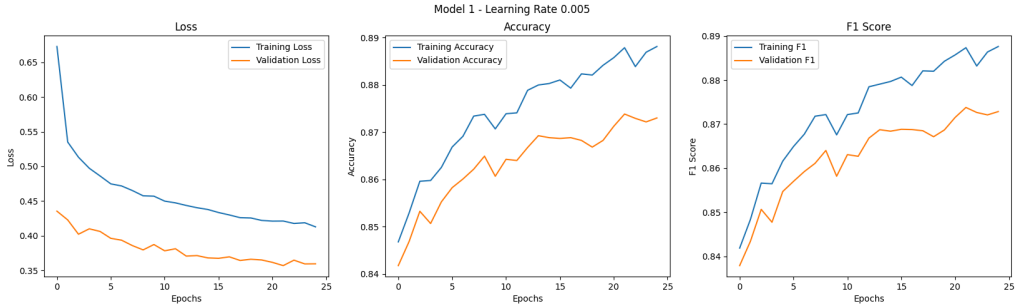
DropoutLayer(0.5),

DenseLayer(256, 10),

SoftmaxLayer()

Learning Rate 0.005





Training Loss: 0.41286020004029766

Validation Loss: 0.35958334586687096

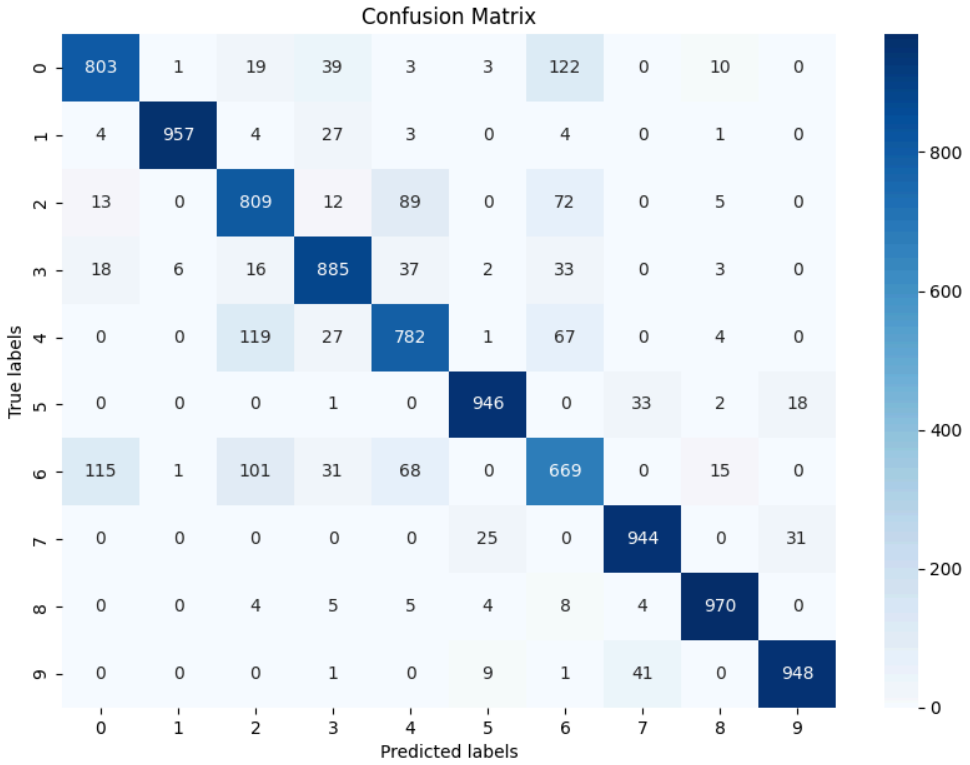
Training Accuracy: 0.888125

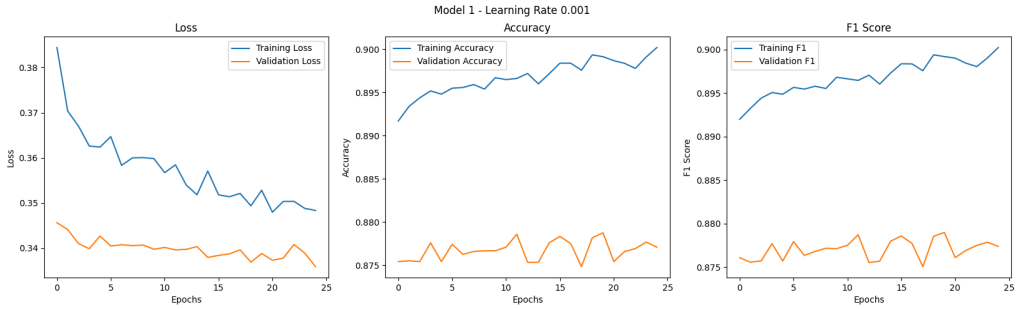
Validation Accuracy: 0.873

Training F1: 0.8876170537289816

Validation F1: 0.8728465935427389

Learning Rate 0.001





Training Loss: 0.3483368428991516

Validation Loss: 0.3359070225282773

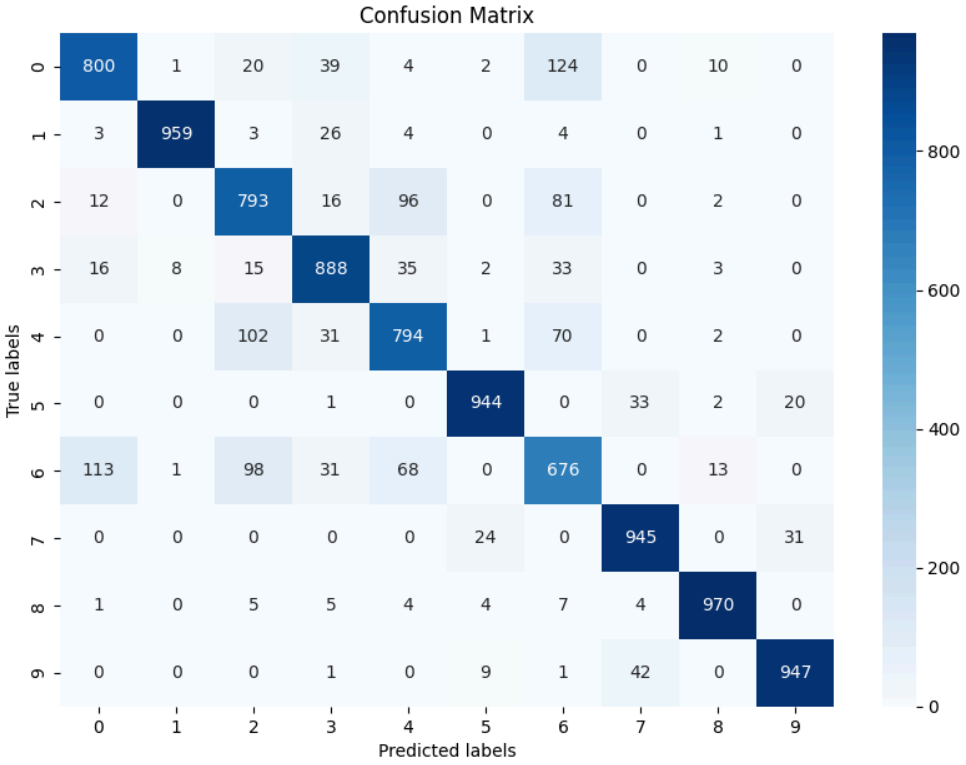
Training Accuracy: 0.9001875

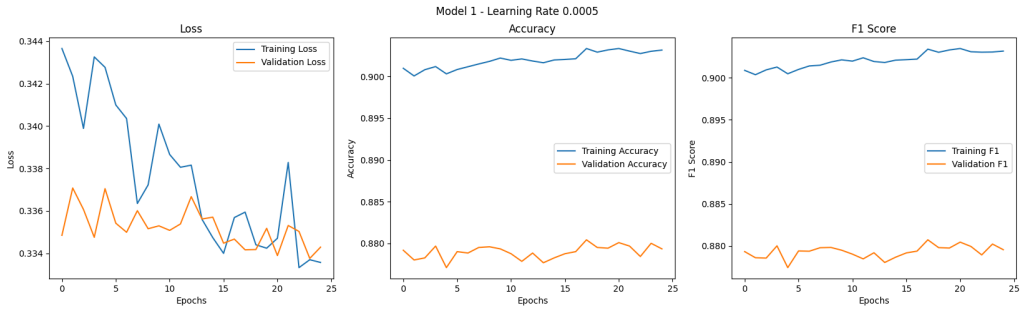
Validation Accuracy: 0.8770833333333333

Training F1: 0.9002409555430141

Validation F1: 0.8773648484945029

Learning Rate 0.0005





Training Loss: 0.33357079179765436

Validation Loss: 0.3342913607913285

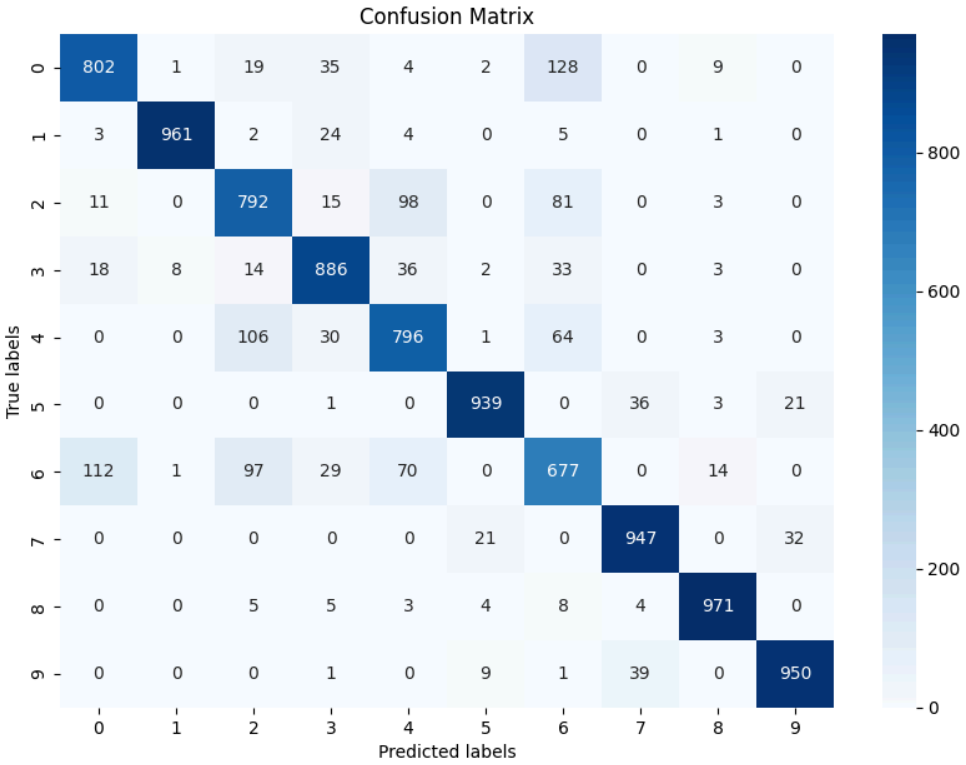
Training Accuracy: 0.9031875

Validation Accuracy: 0.8793333333333333

Training F1: 0.9031754679940983

Validation F1: 0.8795451291664376

Learning Rate 0.0001



Training Loss: 0.32700759702063903

Validation Loss: 0.3318912094043512

Training Accuracy: 0.9043958333333333

Validation Accuracy: 0.8793333333333333

Training F1: 0.9044814049760141

Validation F1: 0.8796492320438816

Model 2

Architecture

DenseLayer(784, 512),

ReLULayer(),

BatchNormLayer(512),

DropoutLayer(0.5),

DenseLayer(512, 256),

ReLULayer(),

BatchNormLayer(256),

DropoutLayer(0.5),

DenseLayer(256, 128),

ReLULayer(),

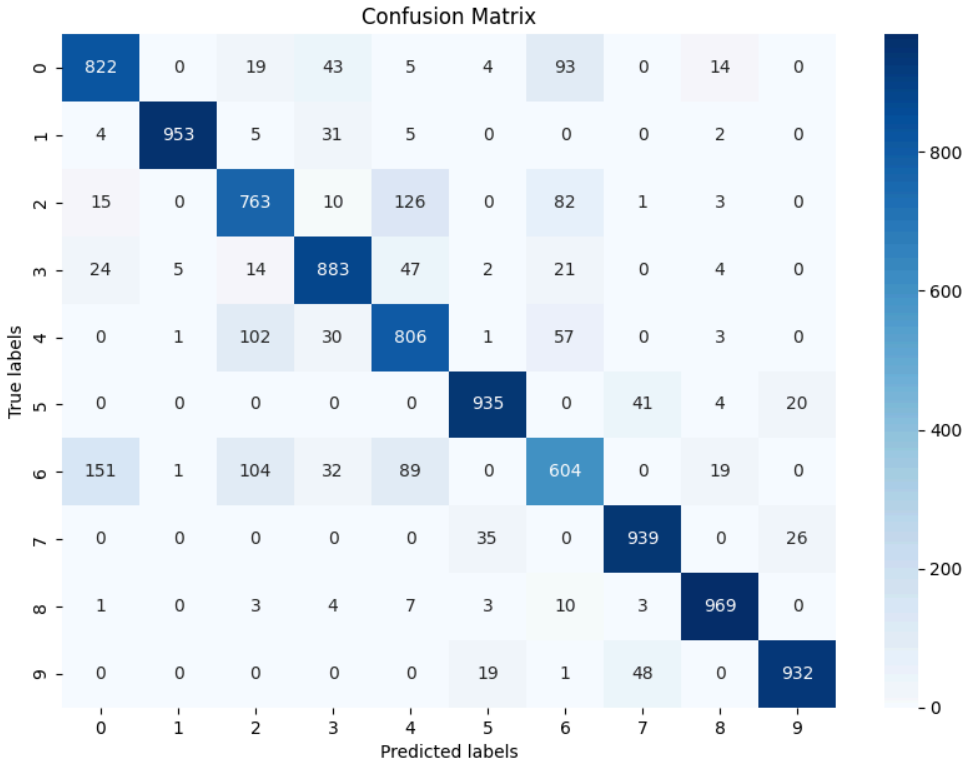
BatchNormLayer(128),

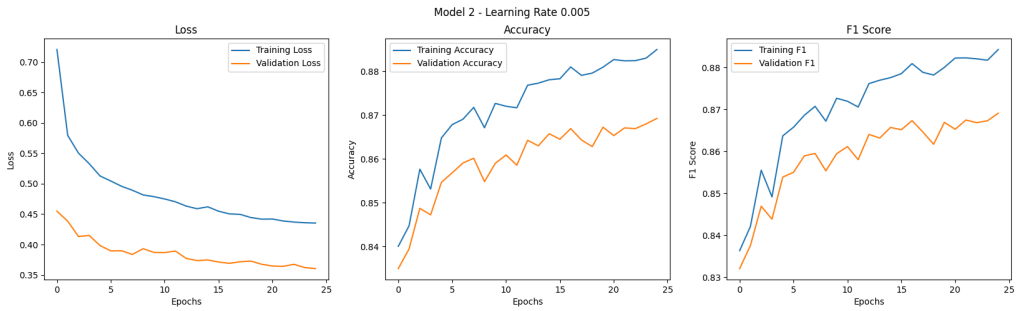
DropoutLayer(0.5),

DenseLayer(128, 10),

SoftmaxLayer()

Learning Rate 0.005





Training Loss: 0.43536108719799277

Validation Loss: 0.36044824775649653

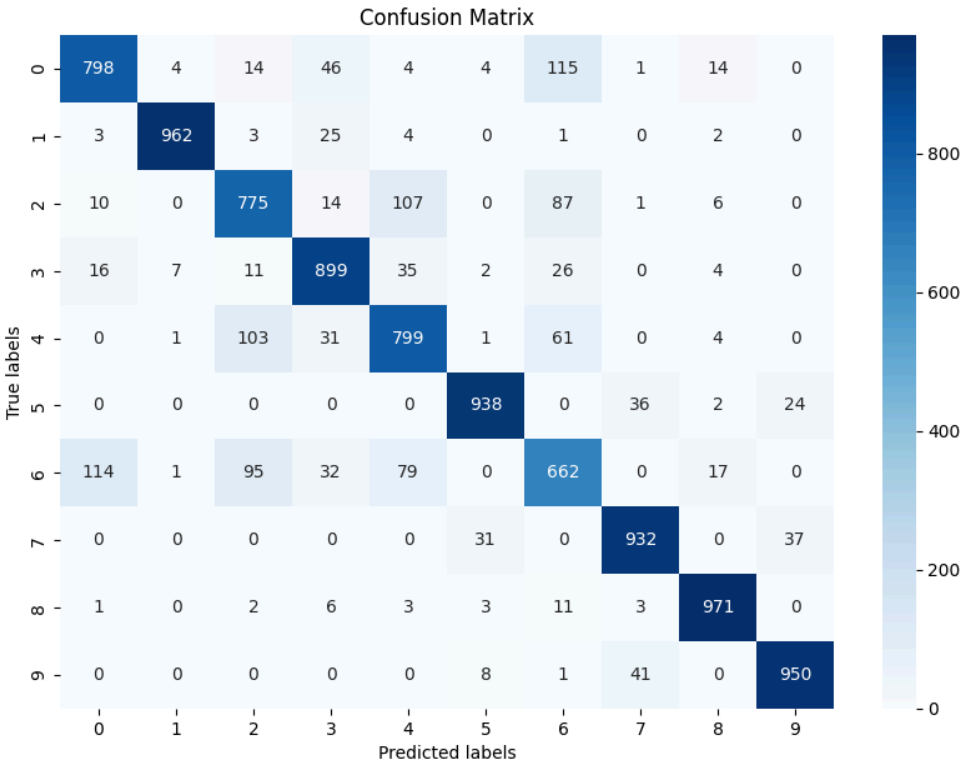
Training Accuracy: 0.8849583333333333

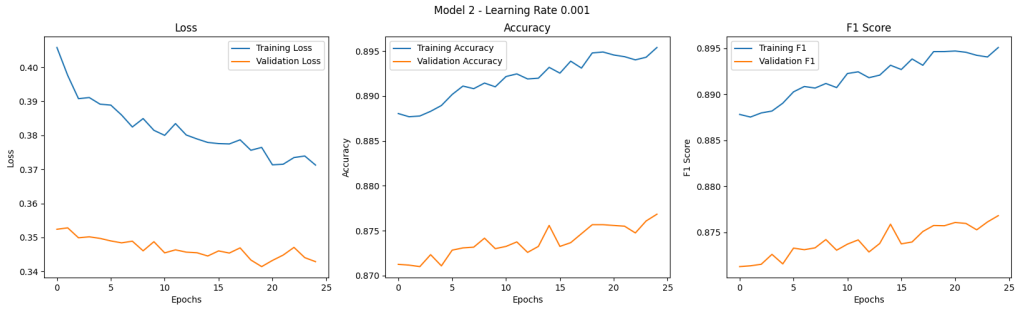
Validation Accuracy: 0.86925

Training F1: 0.8843235963656596

Validation F1: 0.8691404739013802

Learning Rate 0.001





Training Loss: 0.37127027036676946

Validation Loss: 0.34290332779589144

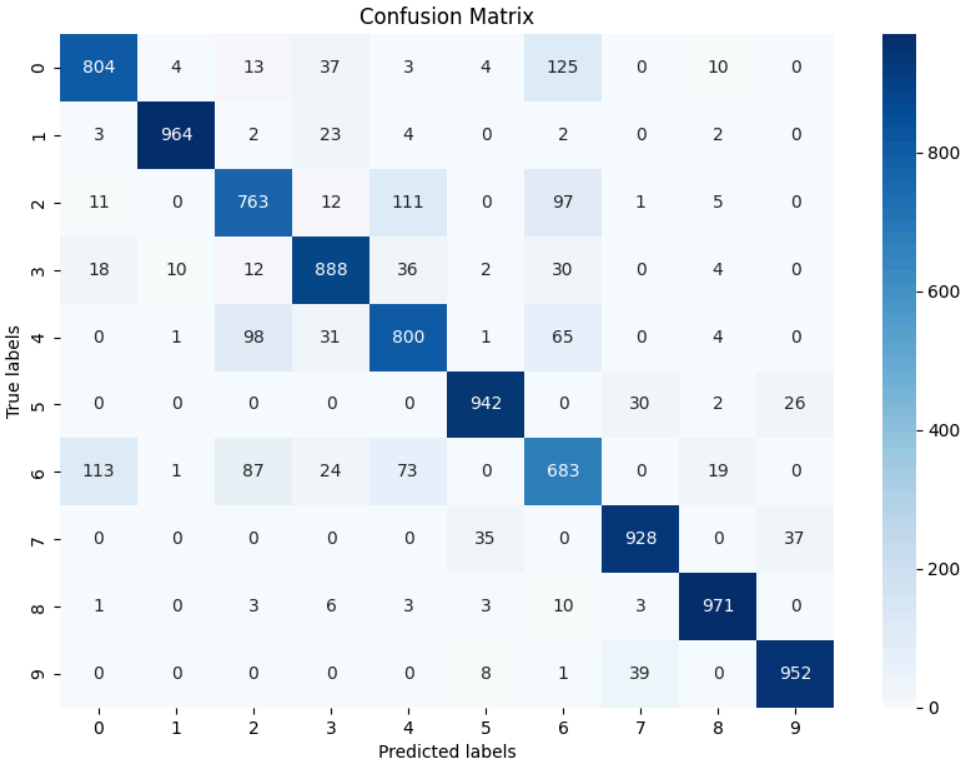
Training Accuracy: 0.8953958333333333

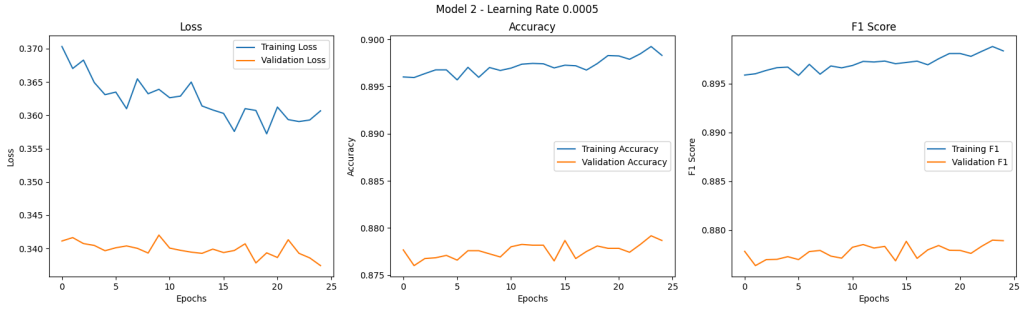
Validation Accuracy: 0.8768333333333334

Training F1: 0.8950796554207093

Validation F1: 0.8768136851375423

Learning Rate 0.0005





Training Loss: 0.3606686248700324

Validation Loss: 0.33742318049397896

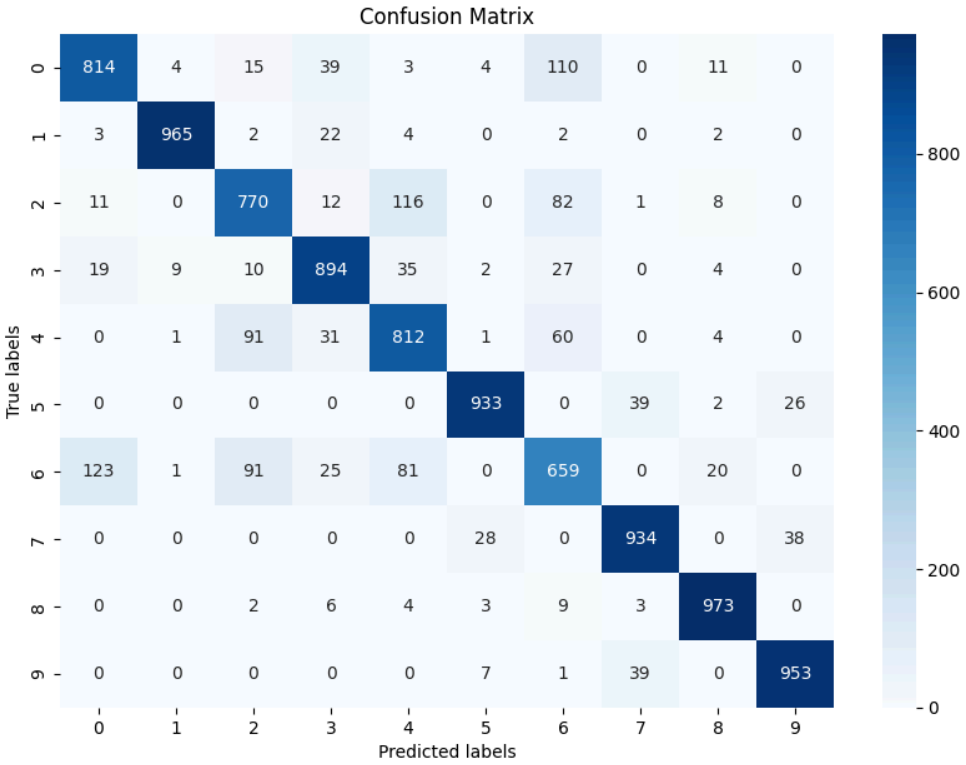
Training Accuracy: 0.8983125

Validation Accuracy: 0.8786666666666667

Training F1: 0.8983559913117485

Validation F1: 0.8789000339315974

Learning Rate 0.0001



Training Loss: 0.353342297290866

Validation Loss: 0.3360935978403882

Training Accuracy: 0.9000833333333333

Validation Accuracy: 0.87975

Training F1: 0.8997182026414512

Validation F1: 0.8796802739589559

Here, we observe that among the 12 combinations of 3 models and 4 learning rates, Model 2 with a learning rate of 0.0001 has the highest validation F1 score, with a value of 0.8796802739589559. Therefore, we conclude that Model 2 with a learning rate of 0.0001 is the best model.
On test data, this model gives Accuracy: 0.8707 and F1: 0.8702