Feed-forward Neural Network From Scratch

CSE472 - ML Sessional, Offline-03 Report

Md Asif Shahriar, 1805040

FNN Architecture

- Dense Layer (768, input_layer_size)
- ReLU
- Dropout Layer (Only During Training)
- Dense Layer (input_layer_size, 26)
- SoftMax Layer

Modifications for Better Result

- Adam Optimizer (during Dense layer back propagation)
- He Initializer (during weights initialization)

Variable Parameters

- Input_layer_size
- learning_rate

Fixed Parameters:

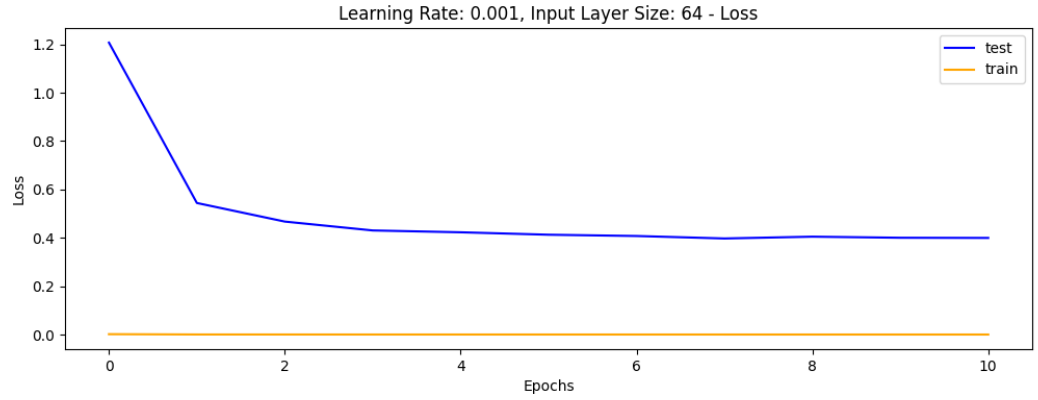
- Drop out rate: 0.1
- Minibatch size: 1024

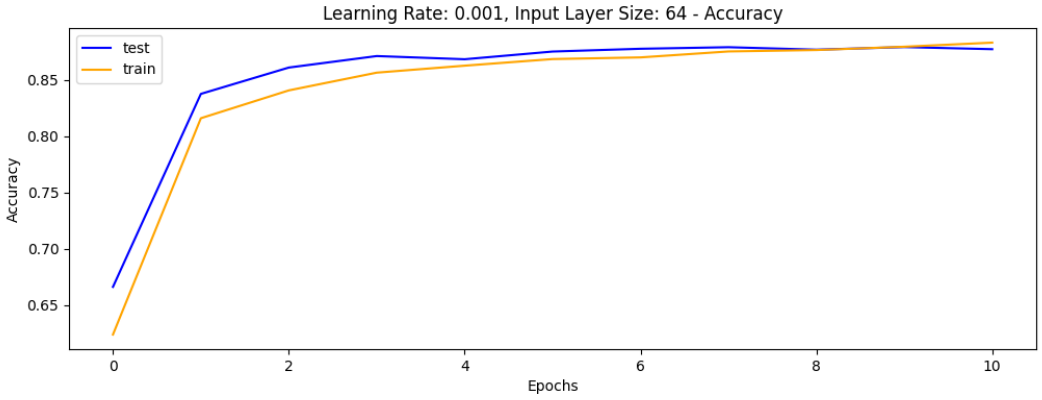
Model Summary

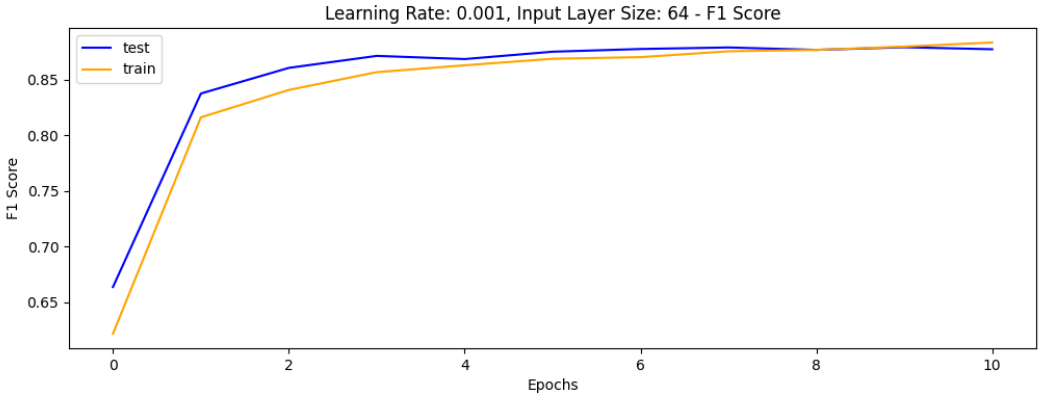
Model	Learning rate	Input layer size	training		validation	
			accuracy	f1	accuracy	f1
Model-1	0.001	64	0.8830	0.8832	0.8772	0.8771
Model-2	0.001	128	0.9299	0.93	0.8964	0.8961
Model-3	0.001	256	0.9604	0.9604	0.9056	0.9054
Model-4	0.005	256 (25)	0.9356	0.9356	0.8965	0.8964
Model-5	0.005	512 (25)	0.9519	0.9521	0.9015	0.9017
Model-6	0.0005	256 (75)	0.9446	0.9447	0.9077	0.9075
Model-7	**0.0005**	**512 (50)**	**0.9556**	**0.9556**	**0.9129**	**0.9127**

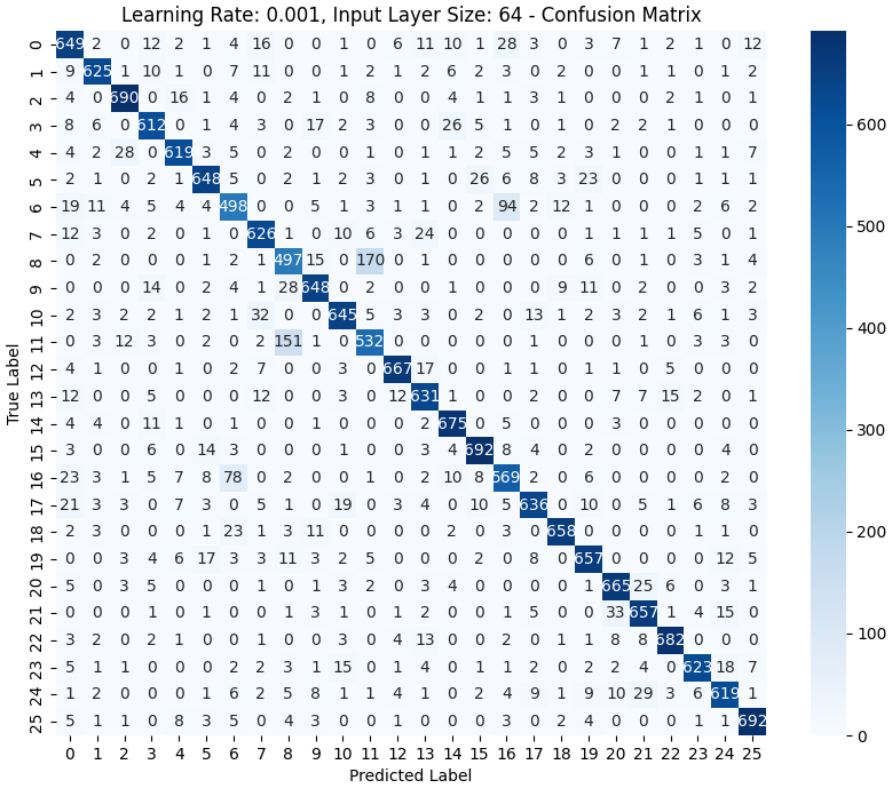
Graphs

Model 1 (learning_rate = 0.001, input_layer_size = 64)

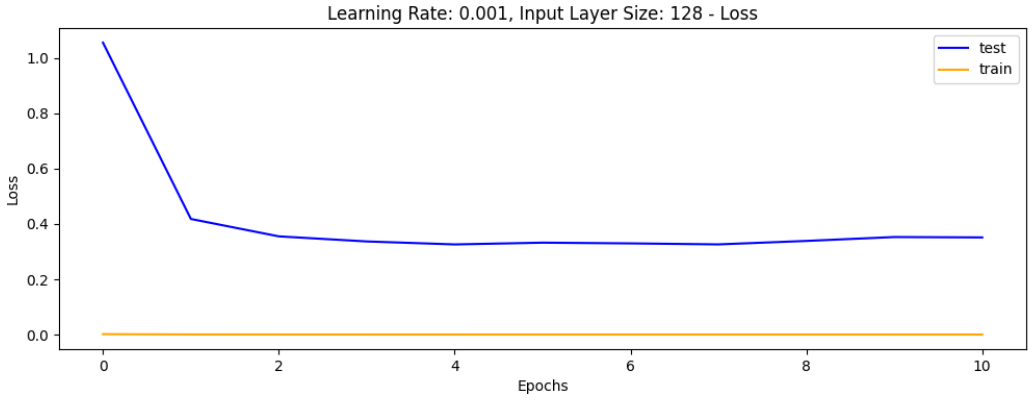


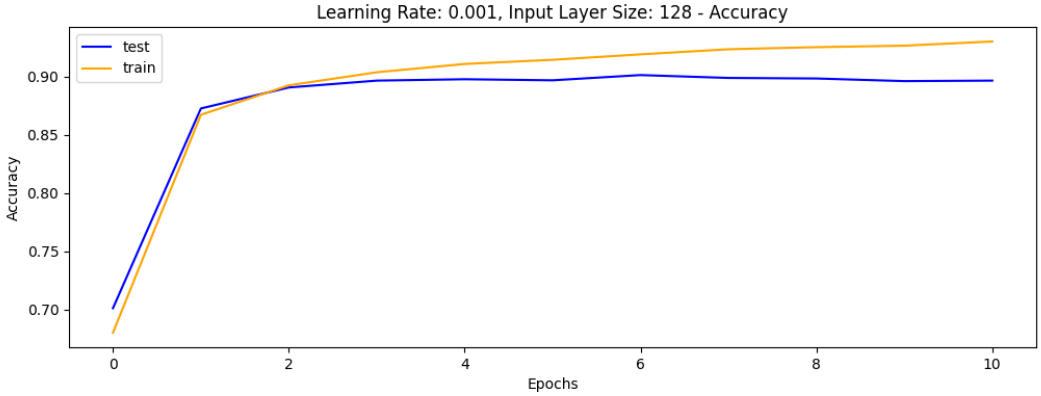


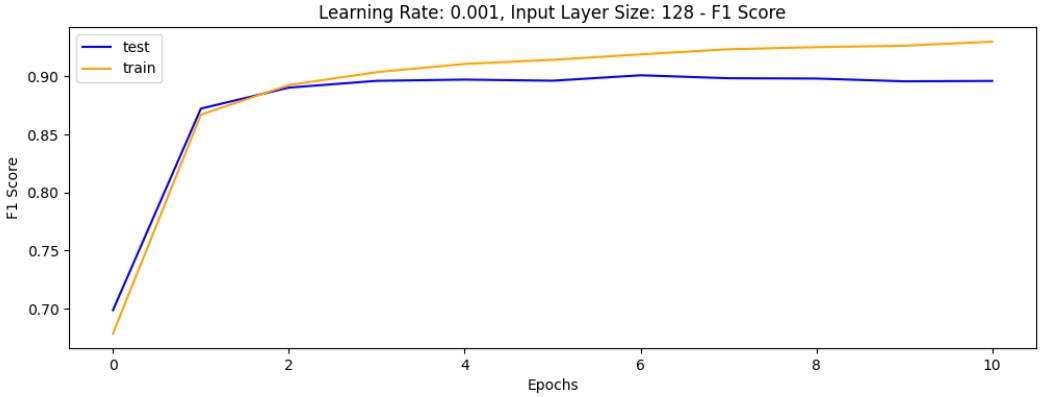


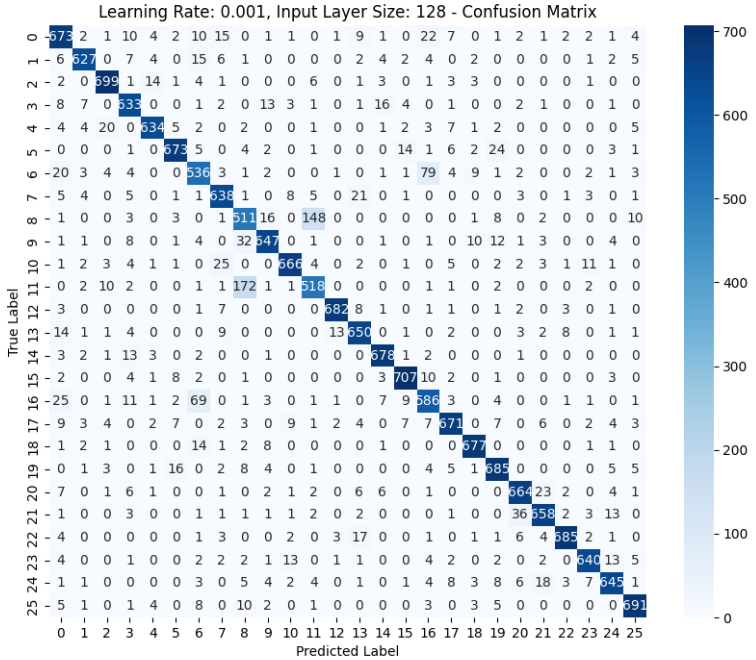


Model 2 (learning_rate = 0.001, input_layer_size = 128)

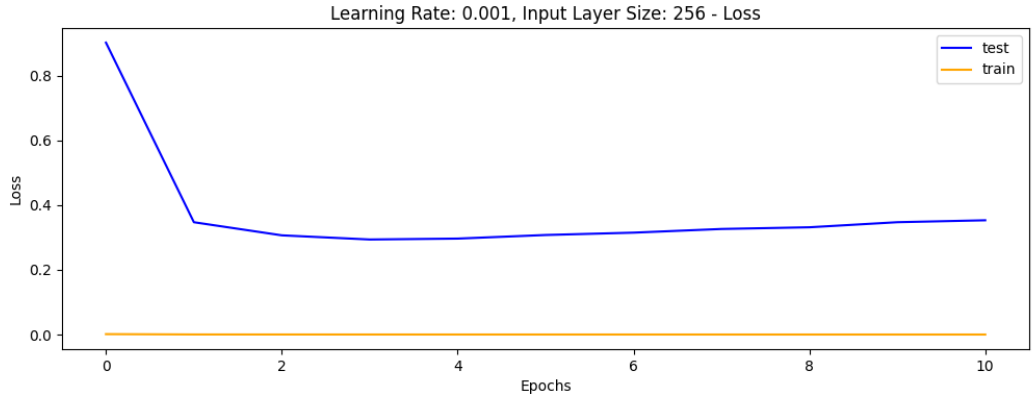


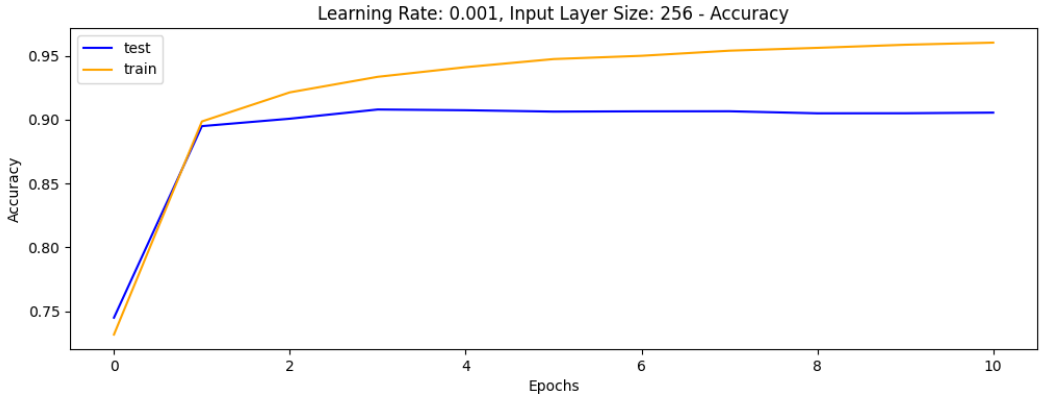

Learning Rate: 0.001, Input Layer Size: 128 - Accuracy
test
train
Accuracy
Epochs

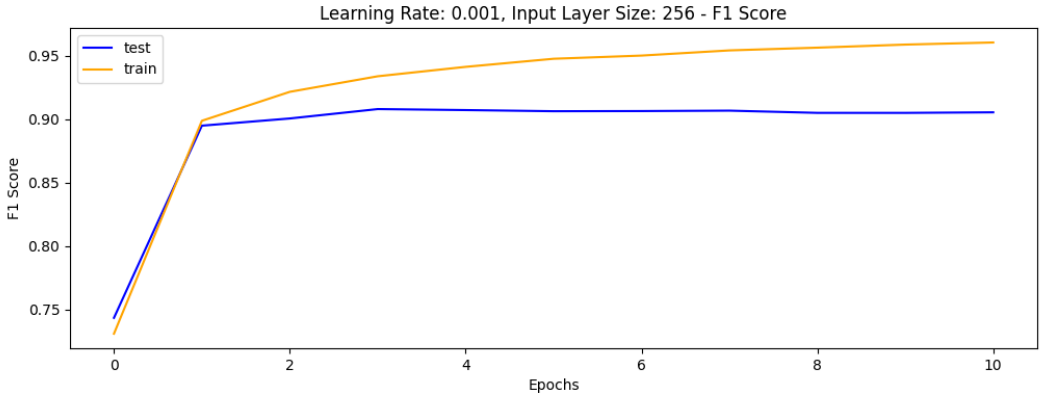

Learning Rate: 0.001, Input Layer Size: 128 - F1 Score
test
train
F1 Score
Epochs

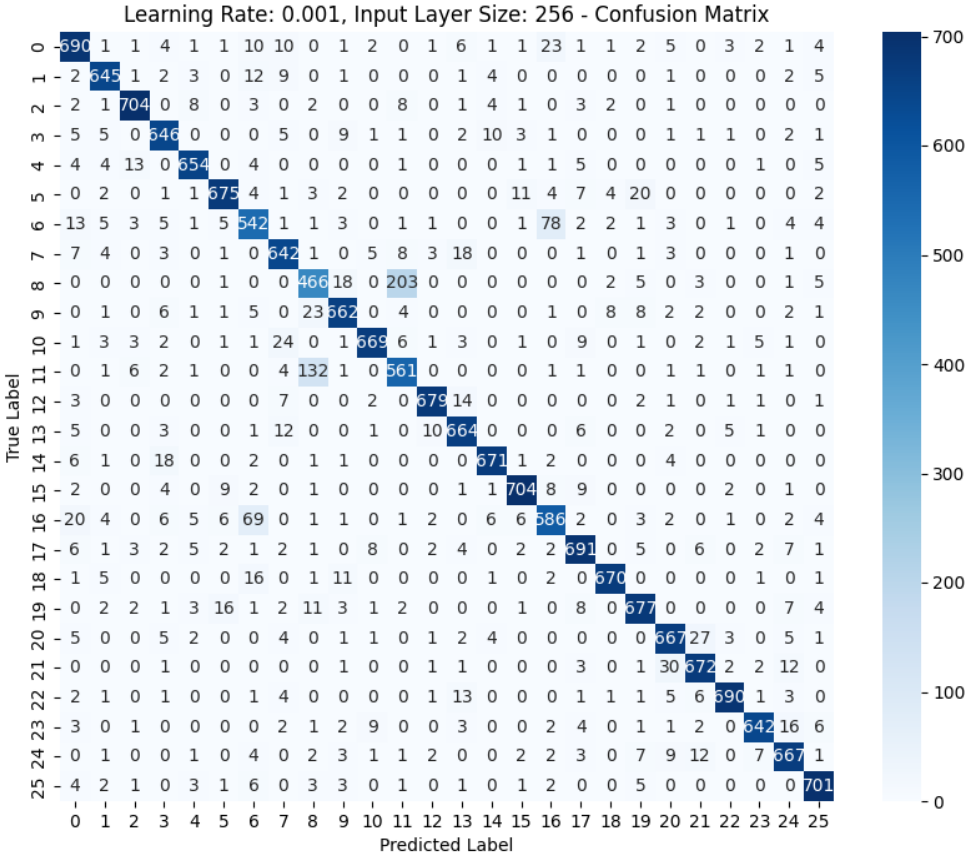

Learning Rate: 0.001, Input Layer Size: 128 - Confusion Matrix
True Label
Predicted Label

Model 3 (learning_rate = 0.001, input_layer_size = 256)

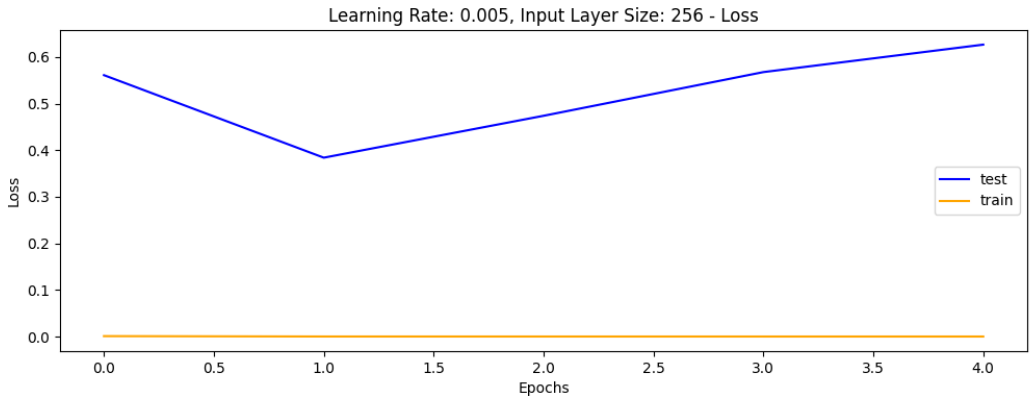


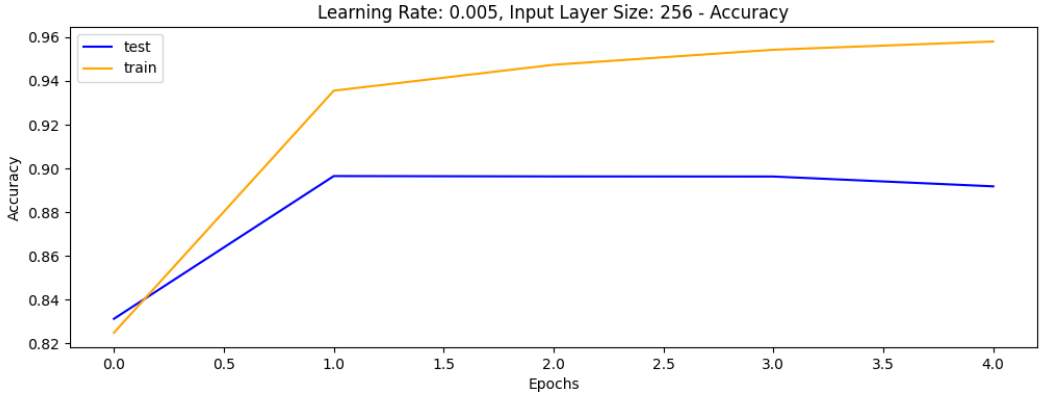


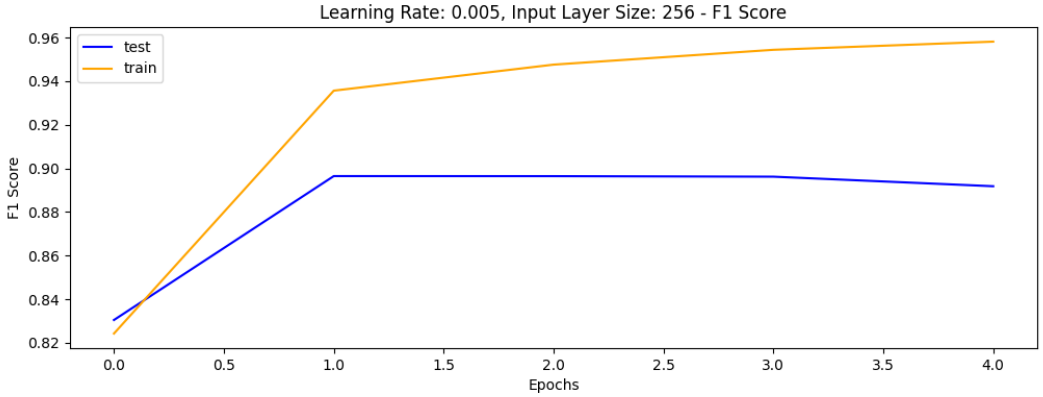


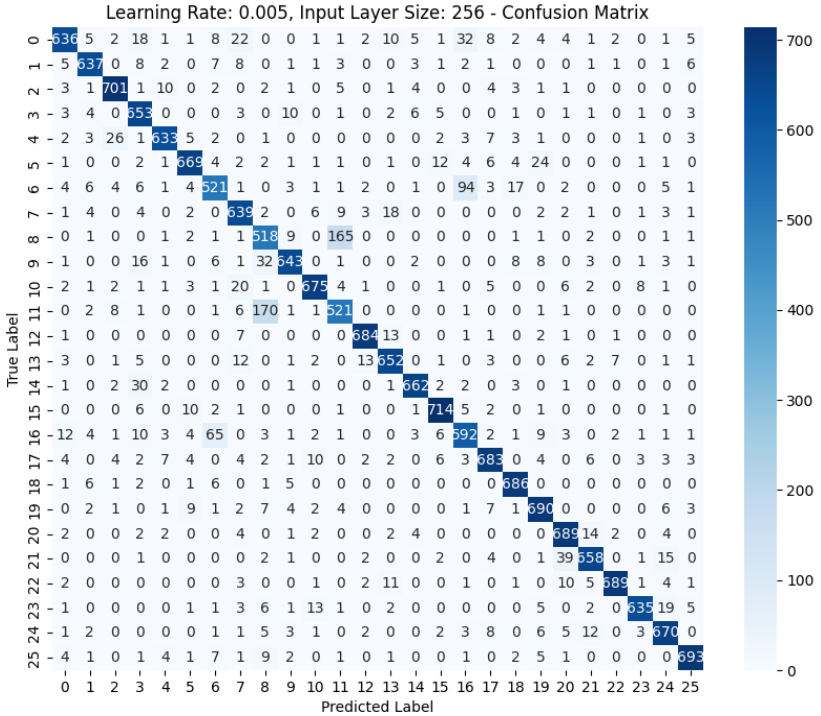


Model 4 (learning_rate = 0.005, input_layer_size = 256)

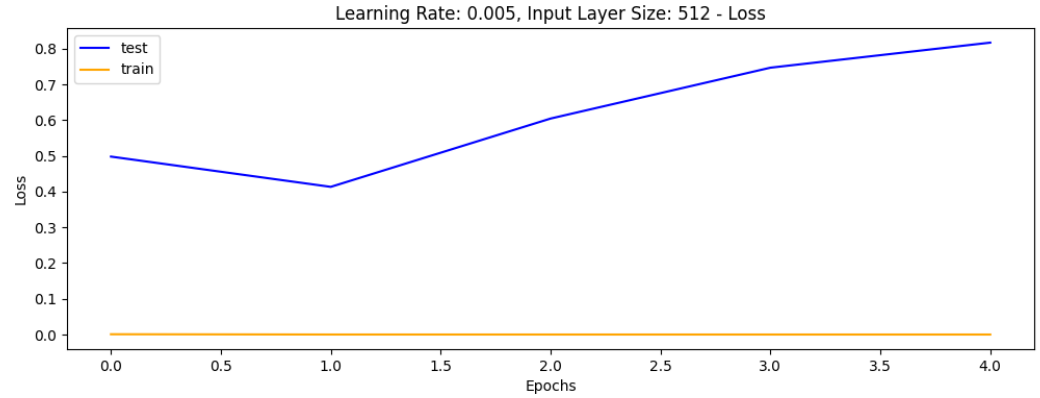


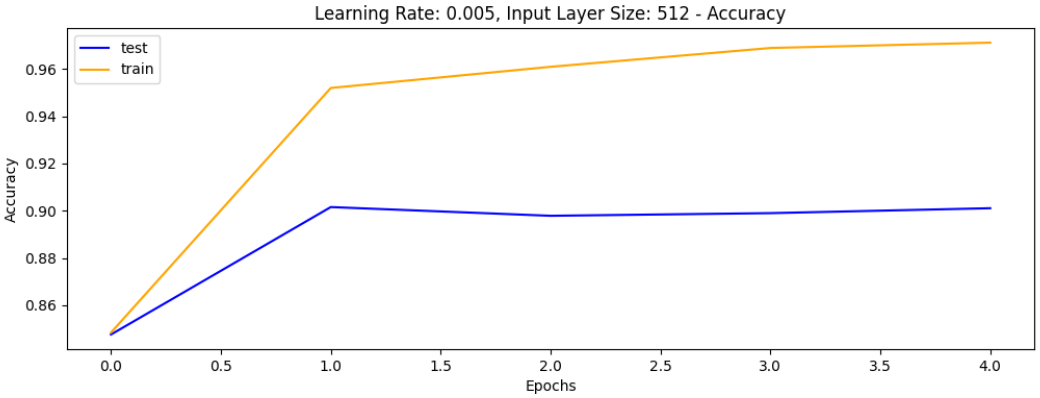

Learning Rate: 0.005, Input Layer Size: 256 - Accuracy
test
train
Accuracy
Epochs

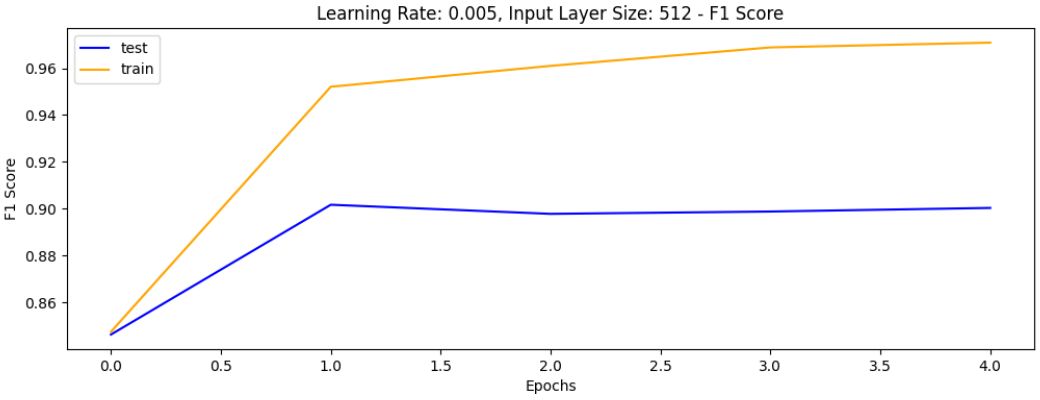

Learning Rate: 0.005, Input Layer Size: 256 - F1 Score
test
train
F1 Score
Epochs

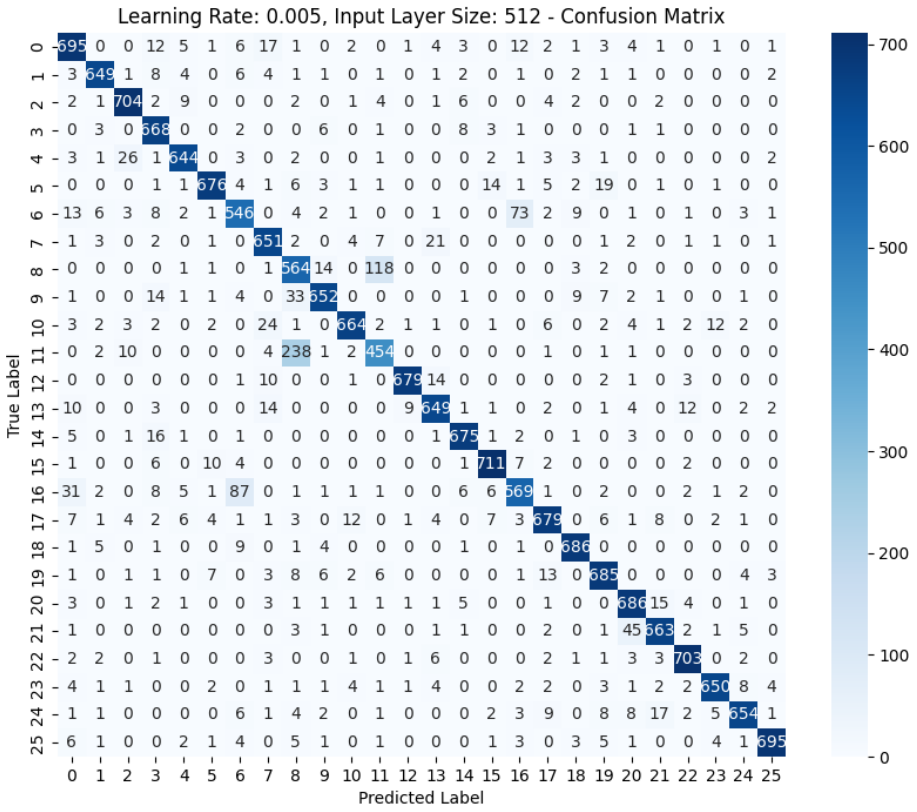

Learning Rate: 0.005, Input Layer Size: 256 - Confusion Matrix
True Label
Predicted Label

Model 5 (learning_rate = 0.005, input_layer_size = 512)

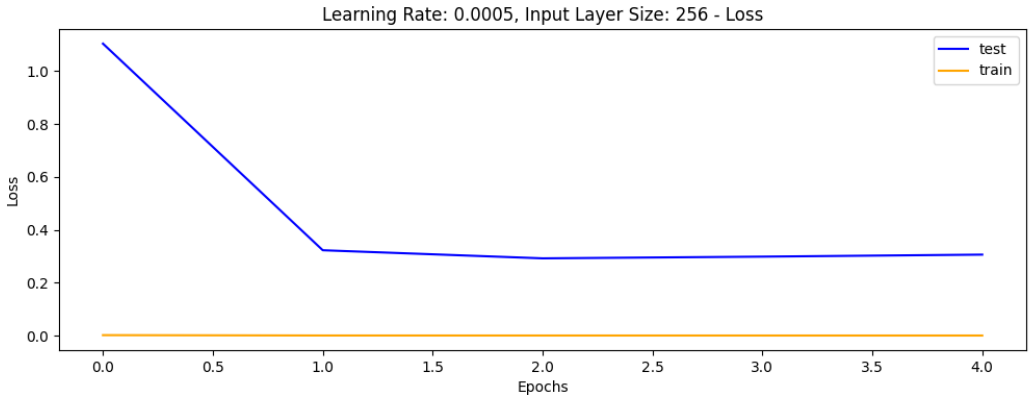


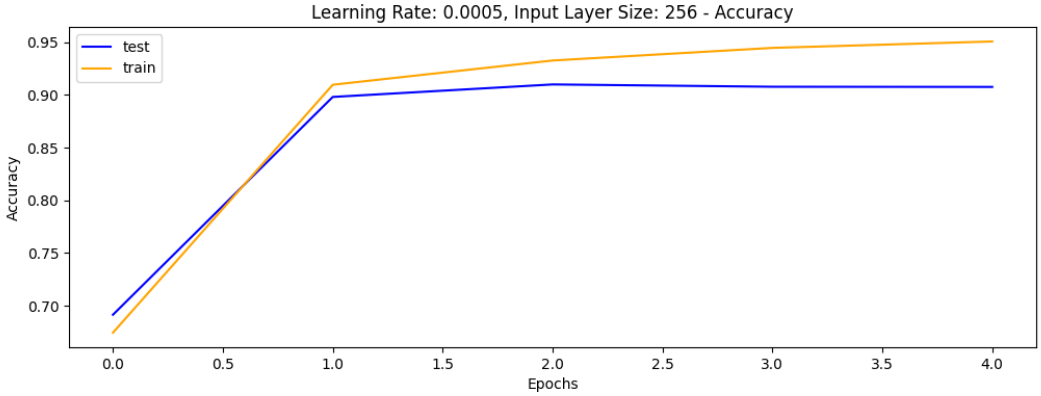


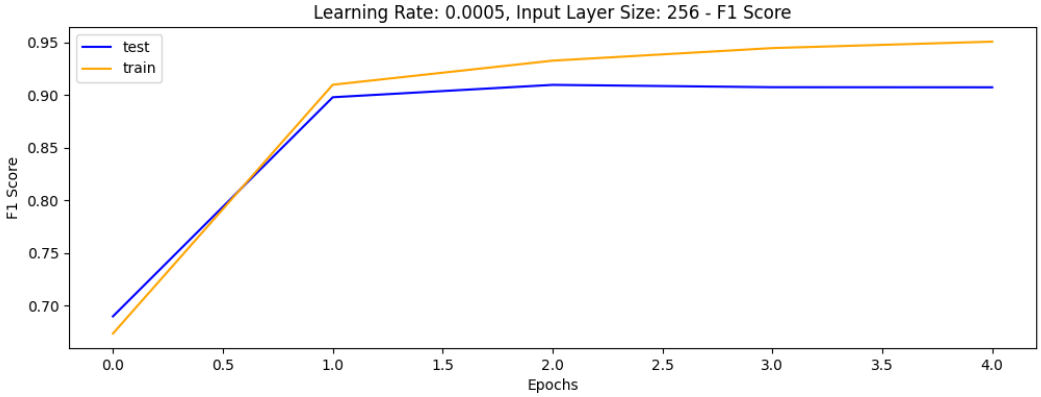


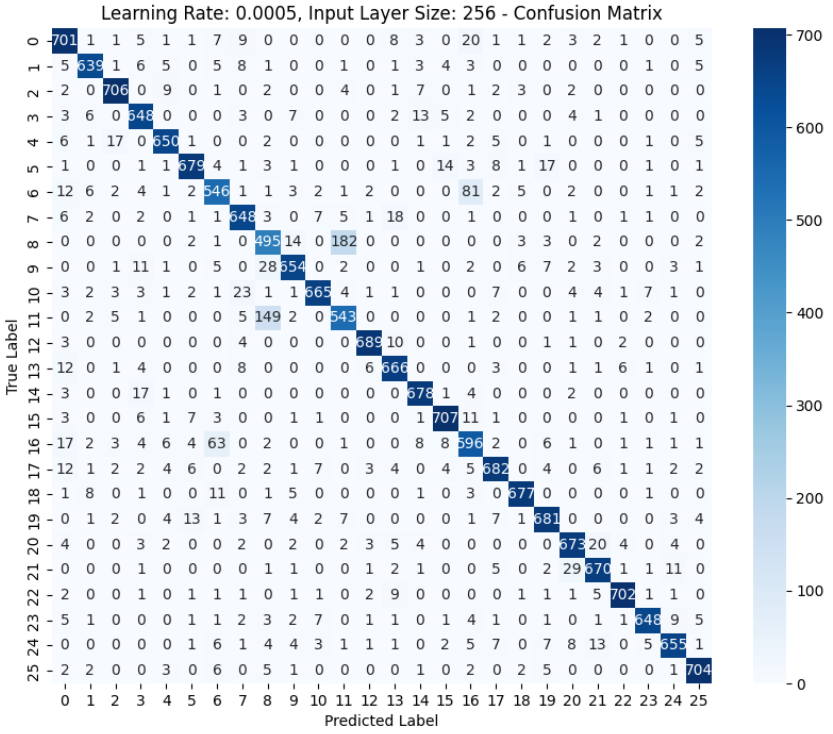


Model 6 (learning_rate = 0.0005, input_layer_size = 256)

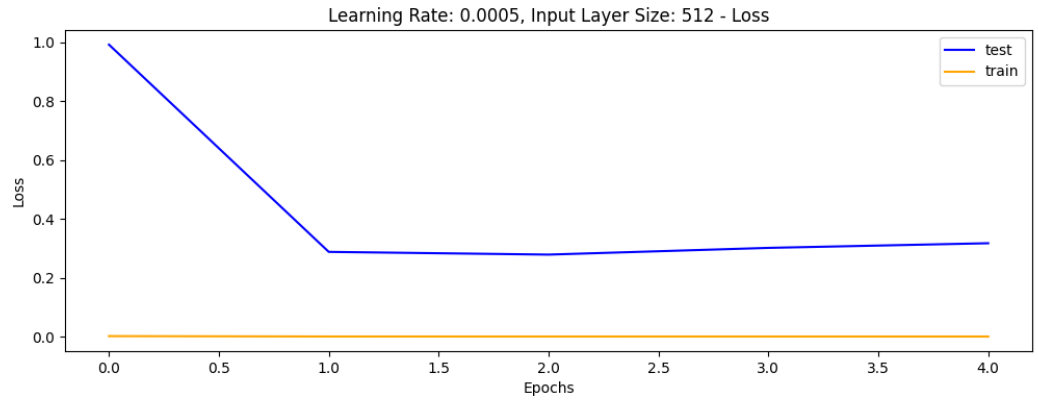


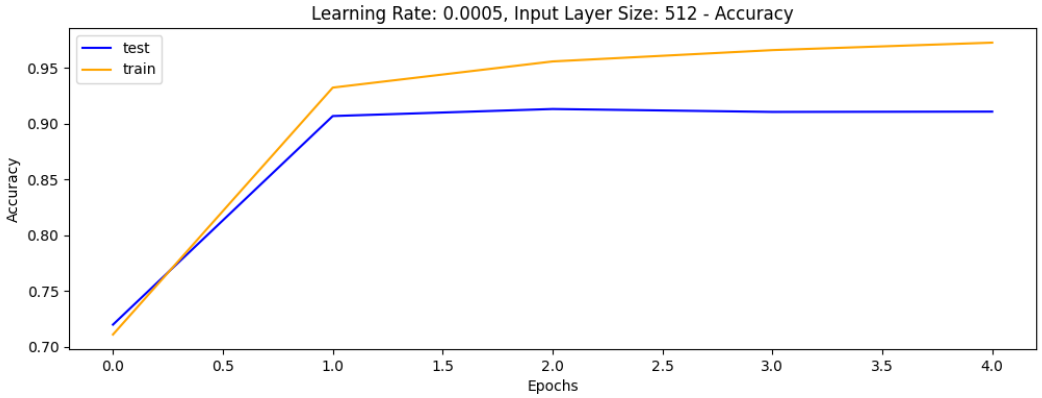

Learning Rate: 0.0005, Input Layer Size: 256 - Accuracy
test
train
Accuracy
Epochs

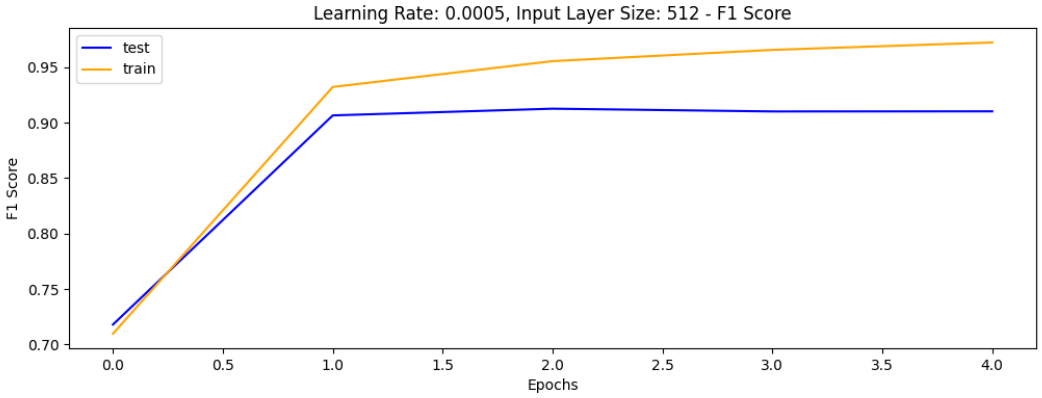

Learning Rate: 0.0005, Input Layer Size: 256 - F1 Score
test
train
F1 Score
Epochs

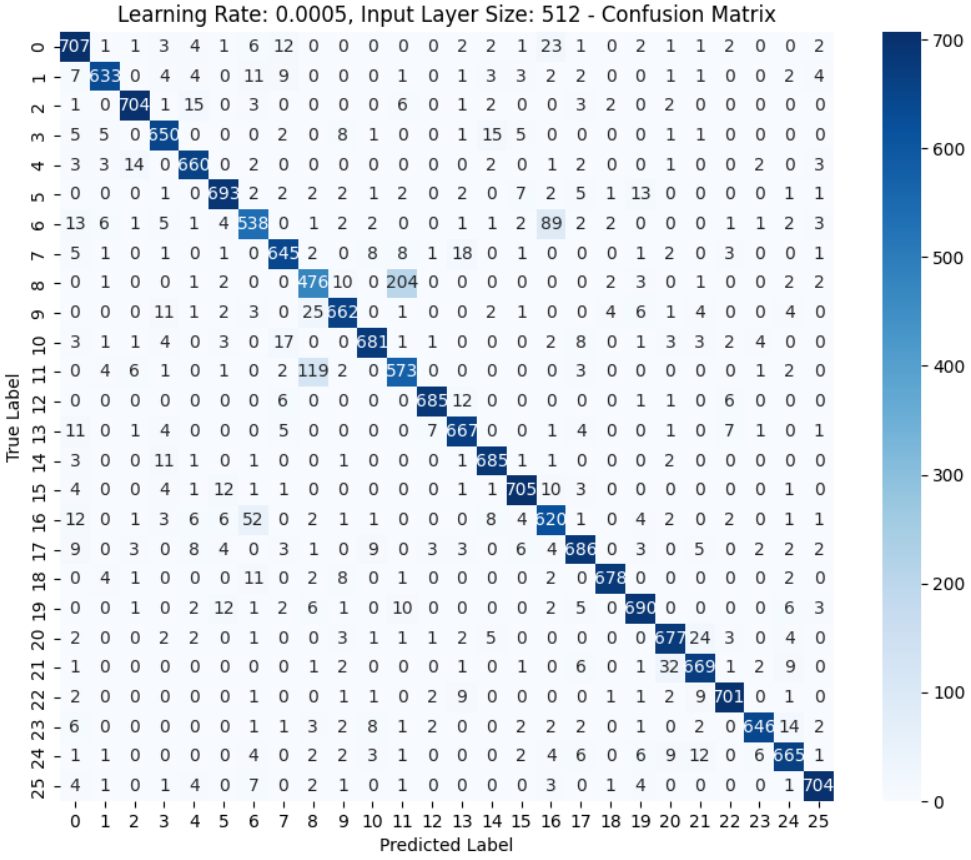

Learning Rate: 0.0005, Input Layer Size: 256 - Confusion Matrix
True Label
Predicted Label

Model 7 (learning_rate = 0.0005, input_layer_size = 512)








Learning Rate: 0.0005, Input Layer Size: 512 - Confusion Matrix
True Label
Predicted Label
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
700
600
500
400
300
200
100
0

Independent Test Performance of the Best model (Model 7)

- Loss: 0.2914
- Accuracy: 0.9129
- F1: 0.9130