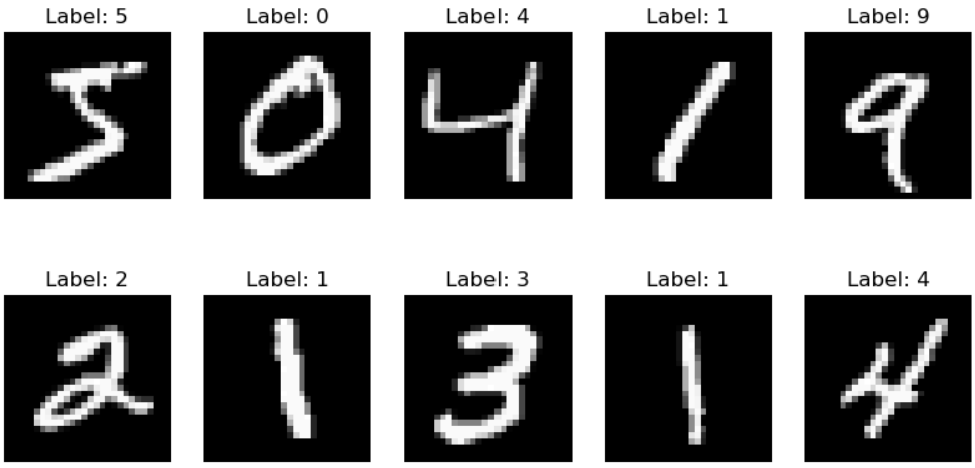
CAP 4611 Programming Assignment 5: Neural Network Report

By: Kashyap Bathina (ka964126) Section 1

Part 2: Building and Training a Neural Network Model

Results, graphs, reports:

```
(60000, 28, 28)
Training data shape: (60000, 784)
Test data shape: (10000, 784)
```

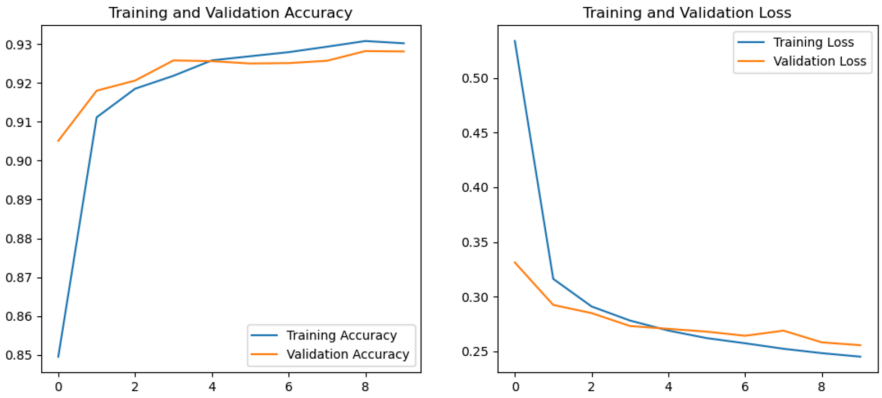


Model: "sequential_2"

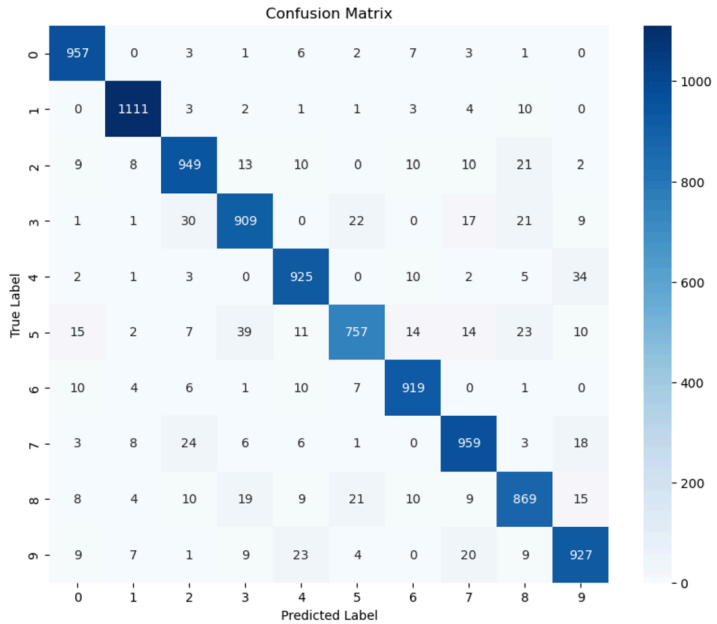
Layer (type)	Output Shape	Param #
dense_4 (Dense)	(None, 8)	6,280
dense_5 (Dense)	(None, 10)	90

Total params: 6,370 (24.88 KB)
Trainable params: 6,370 (24.88 KB)
Non-trainable params: 0 (0.00 B)

```
Epoch 1/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 889us/step - accuracy: 0.7443 - loss: 0.8646 - val_accuracy: 0.9051 - val_loss: 0.3312
Epoch 2/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 804us/step - accuracy: 0.9088 - loss: 0.3265 - val_accuracy: 0.9180 - val_loss: 0.2924
Epoch 3/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 837us/step - accuracy: 0.9170 - loss: 0.2972 - val_accuracy: 0.9206 - val_loss: 0.2849
Epoch 4/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 806us/step - accuracy: 0.9229 - loss: 0.2766 - val_accuracy: 0.9258 - val_loss: 0.2731
Epoch 5/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 797us/step - accuracy: 0.9256 - loss: 0.2694 - val_accuracy: 0.9256 - val_loss: 0.2705
Epoch 6/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 810us/step - accuracy: 0.9280 - loss: 0.2591 - val_accuracy: 0.9250 - val_loss: 0.2679
Epoch 7/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 865us/step - accuracy: 0.9287 - loss: 0.2544 - val_accuracy: 0.9251 - val_loss: 0.2642
Epoch 8/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 858us/step - accuracy: 0.9305 - loss: 0.2517 - val_accuracy: 0.9257 - val_loss: 0.2688
Epoch 9/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 798us/step - accuracy: 0.9319 - loss: 0.2421 - val_accuracy: 0.9282 - val_loss: 0.2581
Epoch 10/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 831us/step - accuracy: 0.9301 - loss: 0.2446 - val_accuracy: 0.9281 - val_loss: 0.2555
```




```
313/313 ━━━━━━━━━━━━━━━━━━━━ 0s 856us/step - accuracy: 0.9190 - loss: 0.2929
Test Accuracy: 0.9282
313/313 ━━━━━━━━━━━━━━━━━━━━ 0s 707us/step
              precision    recall  f1-score   support

           0       0.94      0.98      0.96       980
           1       0.97      0.98      0.97      1135
           2       0.92      0.92      0.92      1032
           3       0.91      0.90      0.90      1010
           4       0.92      0.94      0.93       982
           5       0.93      0.85      0.89       892
           6       0.94      0.96      0.95       958
           7       0.92      0.93      0.93      1028
           8       0.90      0.89      0.90       974
           9       0.91      0.92      0.92      1009

    accuracy                           0.93     10000
   macro avg       0.93      0.93      0.93     10000
weighted avg       0.93      0.93      0.93     10000
```



Part 3: Experimenting with a Larger Hidden Layer

Results, graphs, reports:

Model: "sequential_5"

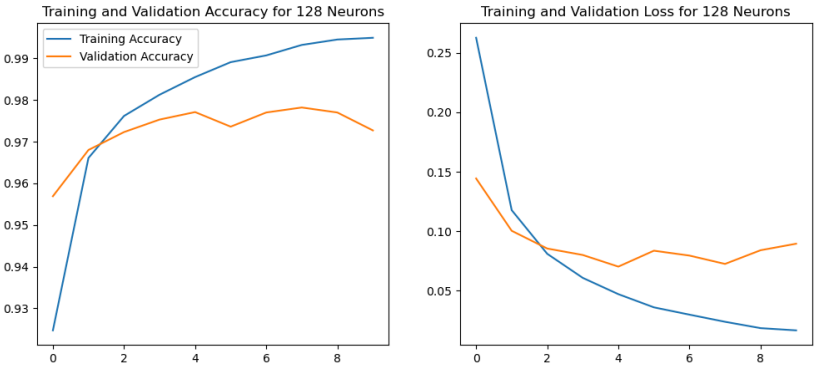
Layer (type)	Output Shape	Param #
dense_10 (Dense)	(None, 128)	100,480
dense_11 (Dense)	(None, 10)	1,290

Total params: 101,770 (397.54 KB)

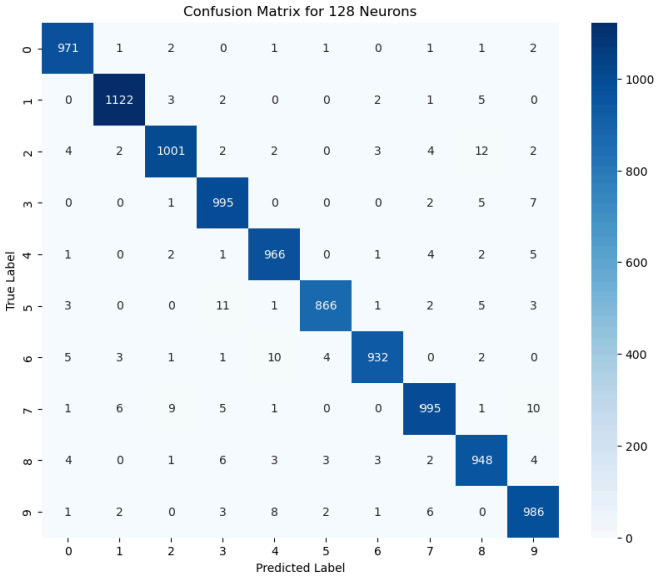
Trainable params: 101,770 (397.54 KB)

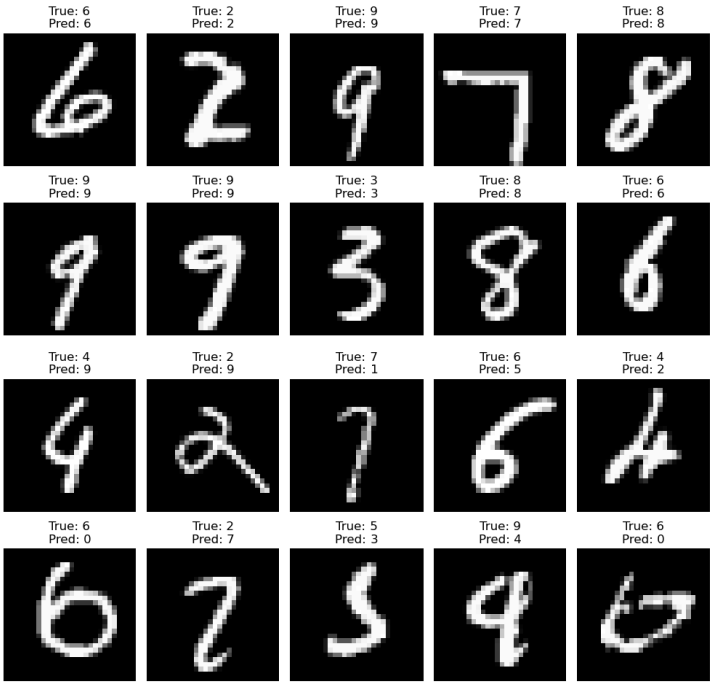
Non-trainable params: 0 (0.00 B)

```
Epoch 1/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 3s 1ms/step - accuracy: 0.8786 - loss: 0.4312 - val_accuracy: 0.9569 - val_loss: 0.1443
Epoch 2/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 1ms/step - accuracy: 0.9646 - loss: 0.1263 - val_accuracy: 0.9680 - val_loss: 0.1003
Epoch 3/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 1ms/step - accuracy: 0.9772 - loss: 0.0804 - val_accuracy: 0.9723 - val_loss: 0.0853
Epoch 4/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 1ms/step - accuracy: 0.9809 - loss: 0.0611 - val_accuracy: 0.9753 - val_loss: 0.0799
Epoch 5/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 3s 1ms/step - accuracy: 0.9864 - loss: 0.0462 - val_accuracy: 0.9771 - val_loss: 0.0701
Epoch 6/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 3s 1ms/step - accuracy: 0.9898 - loss: 0.0341 - val_accuracy: 0.9736 - val_loss: 0.0835
Epoch 7/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 3s 1ms/step - accuracy: 0.9915 - loss: 0.0275 - val_accuracy: 0.9770 - val_loss: 0.0794
Epoch 8/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 1ms/step - accuracy: 0.9942 - loss: 0.0220 - val_accuracy: 0.9782 - val_loss: 0.0724
Epoch 9/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 1ms/step - accuracy: 0.9951 - loss: 0.0173 - val_accuracy: 0.9770 - val_loss: 0.0839
Epoch 10/10
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 2s 1ms/step - accuracy: 0.9955 - loss: 0.0159 - val_accuracy: 0.9727 - val_loss: 0.0894
313/313 ━━━━━━━━━━━━━━━━━━━━ 1s 953us/step - accuracy: 0.9740 - loss: 0.0882
Test Accuracy for 128-Neuron Model: 0.9782
313/313 ━━━━━━━━━━━━━━━━━━━━ 0s 773us/step
              precision    recall  f1-score   support

           0       0.98      0.99      0.99       980
           1       0.99      0.99      0.99      1135
           2       0.98      0.97      0.98      1032
           3       0.97      0.99      0.98      1010
           4       0.97      0.98      0.98       982
           5       0.99      0.97      0.98       892
           6       0.99      0.97      0.98       958
           7       0.98      0.97      0.97      1028
           8       0.97      0.97      0.97       974
           9       0.97      0.98      0.97      1009

    accuracy                           0.98     10000
   macro avg       0.98      0.98      0.98     10000
weighted avg       0.98      0.98      0.98     10000
```


Training and Validation Accuracy for 128 Neurons
Training Accuracy
Validation Accuracy
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0
2
4
6
8

Training and Validation Loss for 128 Neurons
0.25
0.20
0.15
0.10
0.05
0
2
4
6
8


Confusion Matrix for 128 Neurons
True Label
Predicted Label
971 1 2 0 1 1 0 1 1 2
0 1122 3 2 0 0 2 1 5 0
4 2 1001 2 2 0 3 4 12 2
0 0 1 995 0 0 0 2 5 7
1 0 2 1 966 0 1 4 2 5
3 0 0 11 1 866 1 2 5 3
5 3 1 1 10 4 932 0 2 0
1 6 9 5 1 0 0 995 1 10
4 0 1 6 3 3 3 2 948 4
1 2 0 3 8 2 1 6 0 986
0 1 2 3 4 5 6 7 8 9
1000
800
600
400
200
0


True: 6
Pred: 6
True: 2
Pred: 2
True: 9
Pred: 9
True: 7
Pred: 7
True: 8
Pred: 8
True: 9
Pred: 9
True: 9
Pred: 9
True: 3
Pred: 3
True: 8
Pred: 8
True: 6
Pred: 6
True: 4
Pred: 9
True: 2
Pred: 9
True: 7
Pred: 1
True: 6
Pred: 5
True: 4
Pred: 2
True: 6
Pred: 0
True: 2
Pred: 7
True: 5
Pred: 3
True: 9
Pred: 4
True: 6
Pred: 0

Part 4: Custom Neural Network for 99% average F1 Score

Results, graphs, reports:

Model: "sequential_8"

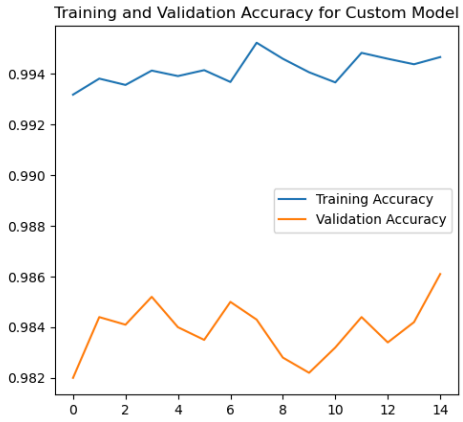
Layer (type)	Output Shape	Param #
dense_20 (Dense)	(None, 1024)	803,840
dropout_5 (Dropout)	(None, 1024)	0
dense_21 (Dense)	(None, 512)	524,800
dropout_6 (Dropout)	(None, 512)	0
dense_22 (Dense)	(None, 256)	131,328
dropout_7 (Dropout)	(None, 256)	0
dense_23 (Dense)	(None, 128)	32,896
dense_24 (Dense)	(None, 10)	1,290

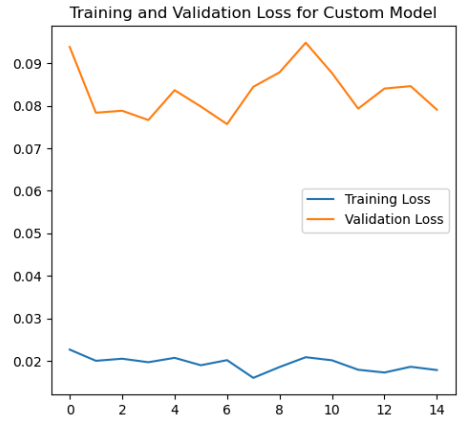
Total params: 1,494,154 (5.70 MB)

Trainable params: 1,494,154 (5.70 MB)

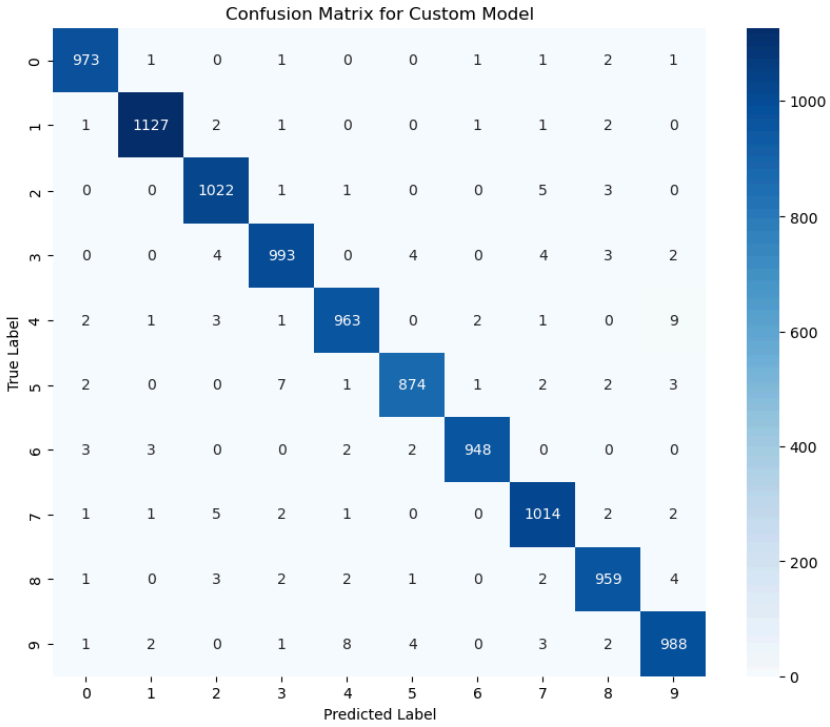
Non-trainable params: 0 (0.00 B)

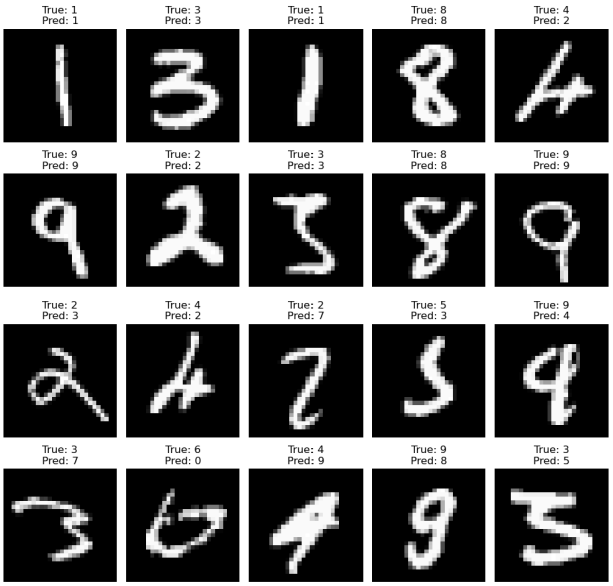
```
Epoch 1/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 5s 4ms/step - accuracy: 0.9935 - loss: 0.0209 - val_accuracy: 0.9820 - val_loss: 0.0939
Epoch 2/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.9935 - loss: 0.0216 - val_accuracy: 0.9844 - val_loss: 0.0784
Epoch 3/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.9933 - loss: 0.0215 - val_accuracy: 0.9841 - val_loss: 0.0788
Epoch 4/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 4s 4ms/step - accuracy: 0.9940 - loss: 0.0187 - val_accuracy: 0.9852 - val_loss: 0.0767
Epoch 5/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.9947 - loss: 0.0177 - val_accuracy: 0.9840 - val_loss: 0.0837
Epoch 6/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.9945 - loss: 0.0182 - val_accuracy: 0.9835 - val_loss: 0.0798
Epoch 7/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.9942 - loss: 0.0179 - val_accuracy: 0.9850 - val_loss: 0.0757
Epoch 8/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.9957 - loss: 0.0159 - val_accuracy: 0.9843 - val_loss: 0.0845
Epoch 9/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.9947 - loss: 0.0184 - val_accuracy: 0.9828 - val_loss: 0.0879
Epoch 10/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 4s 4ms/step - accuracy: 0.9950 - loss: 0.0170 - val_accuracy: 0.9822 - val_loss: 0.0948
Epoch 11/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 4s 4ms/step - accuracy: 0.9932 - loss: 0.0200 - val_accuracy: 0.9832 - val_loss: 0.0877
Epoch 12/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.9947 - loss: 0.0179 - val_accuracy: 0.9844 - val_loss: 0.0793
Epoch 13/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 4s 4ms/step - accuracy: 0.9944 - loss: 0.0175 - val_accuracy: 0.9834 - val_loss: 0.0841
Epoch 14/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 4s 4ms/step - accuracy: 0.9946 - loss: 0.0185 - val_accuracy: 0.9842 - val_loss: 0.0846
Epoch 15/15
938/938 ━━━━━━━━━━━━━━━━━━━━ 4s 4ms/step - accuracy: 0.9950 - loss: 0.0168 - val_accuracy: 0.9861 - val_loss: 0.0791
```





```
313/313 ━━━━━━━━━━━━━━━━━━━━ 1s 1ms/step - accuracy: 0.9829 - loss: 0.1059
Test Accuracy for Custom Model: 0.9861
313/313 ━━━━━━━━━━━━━━━━━━━━ 0s 1ms/step
              precision    recall  f1-score   support

           0       0.99      0.99      0.99       980
           1       0.99      0.99      0.99      1135
           2       0.98      0.99      0.99      1032
           3       0.98      0.98      0.98      1010
           4       0.98      0.98      0.98       982
           5       0.99      0.98      0.98       892
           6       0.99      0.99      0.99       958
           7       0.98      0.99      0.98      1028
           8       0.98      0.98      0.98       974
           9       0.98      0.98      0.98      1009

    accuracy                           0.99     10000
   macro avg       0.99      0.99      0.99     10000
weighted avg       0.99      0.99      0.99     10000
```





Part 5: Discussion Questions

1. Epochs: When neural networks are being trained, an epoch is a single run of the whole training dataset. The model uses the training data to minimize the loss function in order to update its parameters during each epoch.

Epochs are used to iteratively train the model, allowing it to gradually learn and improve its predictions across multiple cycles through the data.

Though too many epochs might cause overfitting, in which the model memorizes training data rather than generalizing, increasing the number of epochs can enhance model performance by giving it more time to learn. On the other hand, underfitting, in which the model does not learn enough from the data, might occur when there are insufficient epochs.

2. Batch Size: Batch size refers to the number of training samples processed before the model's internal parameters are updated. For example, in a batch size of 32, the model computes the gradient and updates the weights after processing 32 samples.

A smaller batch size allows the model to update its weights more frequently, potentially leading to faster convergence but noisier updates. Larger batch sizes provide smoother updates and are more computationally efficient but may require more memory and converge more slowly.

3. Dropout: Dropout is a regularization technique where a random subset of neurons is "dropped" (set to zero) during training. This prevents the model from relying too heavily on specific neurons and forces it to learn more robust features.

Dropout helps prevent overfitting by reducing the model's reliance on specific neurons, effectively encouraging it to distribute learning across the network. It also improves generalization, enabling the model to perform better on unseen data.