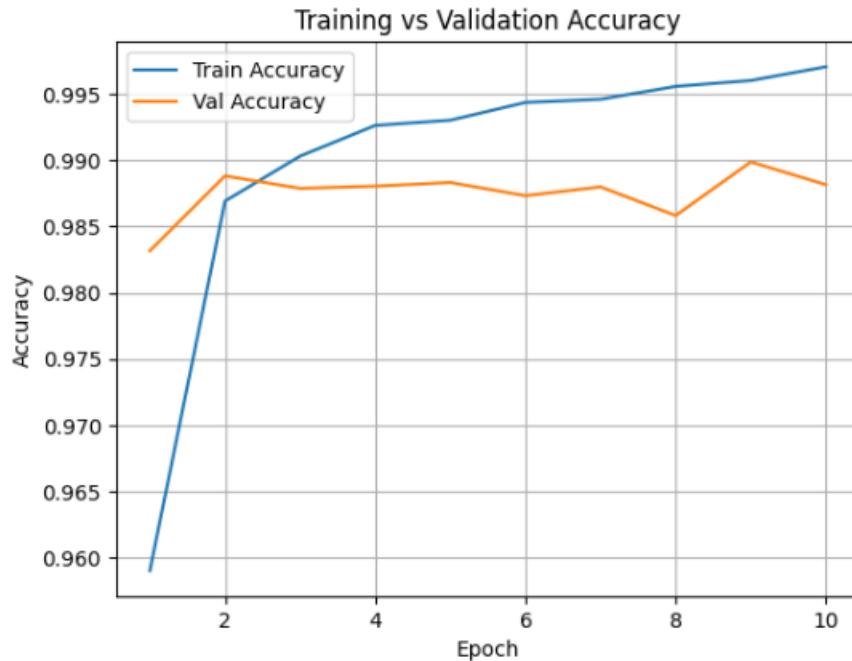


Qamar Uddin

Exercise 3 Report.

I have selected **Adam** ($lr = 0.001$, $\text{betas} = (0.9, 0.999)$, weight decay = $1e-4$) because it converged faster and more stable than SGD, RMSProp and AdaGrad on the MNIST dataset with CNN. With these parameters, the model achieved a best validation accuracy of approximately 0.9899% at epoch 9, and a test accuracy of approximately 0.9883%. The learning rate $1e-3$ is a standard and empirically effective choice for Adam, while the default betas provide good smoothing of gradient noise; a small weight decay $1e-4$ helps regularize the model and prevent overfitting.

Below, the graph shows the accuracy level and loss level of training and testing phase.



```
... 100%|██████████| 1.65M/1.65M [00:00<00:00, 14.8MB/s]
100%|██████████| 4.54k/4.54k [00:00<00:00, 5.57MB/s]
Epoch [1/10] Train Loss: 0.1310, Val Loss: 0.0522, Train Acc: 0.9590, Val Acc: 0.9832
Epoch [2/10] Train Loss: 0.0435, Val Loss: 0.0392, Train Acc: 0.9869, Val Acc: 0.9888
Epoch [3/10] Train Loss: 0.0296, Val Loss: 0.0404, Train Acc: 0.9903, Val Acc: 0.9879
Epoch [4/10] Train Loss: 0.0232, Val Loss: 0.0384, Train Acc: 0.9926, Val Acc: 0.9880
Epoch [5/10] Train Loss: 0.0203, Val Loss: 0.0386, Train Acc: 0.9930, Val Acc: 0.9883
Epoch [6/10] Train Loss: 0.0172, Val Loss: 0.0413, Train Acc: 0.9943, Val Acc: 0.9873
Epoch [7/10] Train Loss: 0.0151, Val Loss: 0.0419, Train Acc: 0.9946, Val Acc: 0.9880
Epoch [8/10] Train Loss: 0.0129, Val Loss: 0.0480, Train Acc: 0.9956, Val Acc: 0.9858
Epoch [9/10] Train Loss: 0.0125, Val Loss: 0.0352, Train Acc: 0.9960, Val Acc: 0.9899
Epoch [10/10] Train Loss: 0.0094, Val Loss: 0.0380, Train Acc: 0.9970, Val Acc: 0.9881
```

Best validation accuracy: 0.9899 at epoch 9

Test Accuracy: 0.9883

Training vs Validation Loss

