Optimization on MNIST

Optimizers Used :

- Adam
- RMSprop
- Adagrad
- SGD
- AdamW
- Adamax

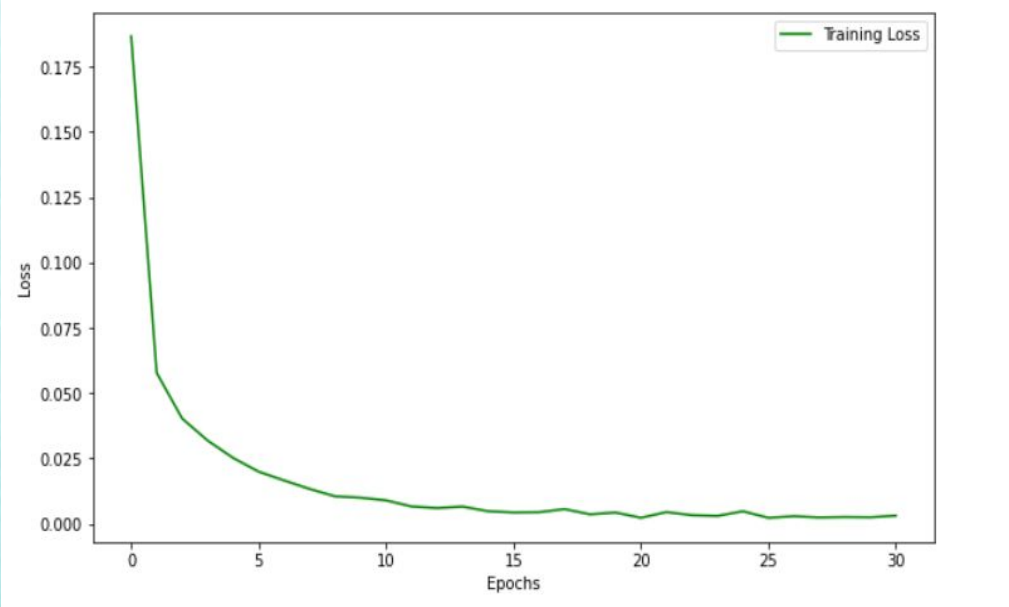
OVERVIEW

A two layered CNN based model is created and the initial weights and biases of the model is stored in a path file. We aim to find the working of various optimizers on the MNIST dataset and analyze the effectiveness of the optimizers on the CNN model implemented.
We then train the model for six different optimizers using the saved model and initialised weights and compare the accuracies, run-time, epochs and the mean and standard deviation taken to find the optimal model to be used and the optimum number of epochs needed to train the model for best results.

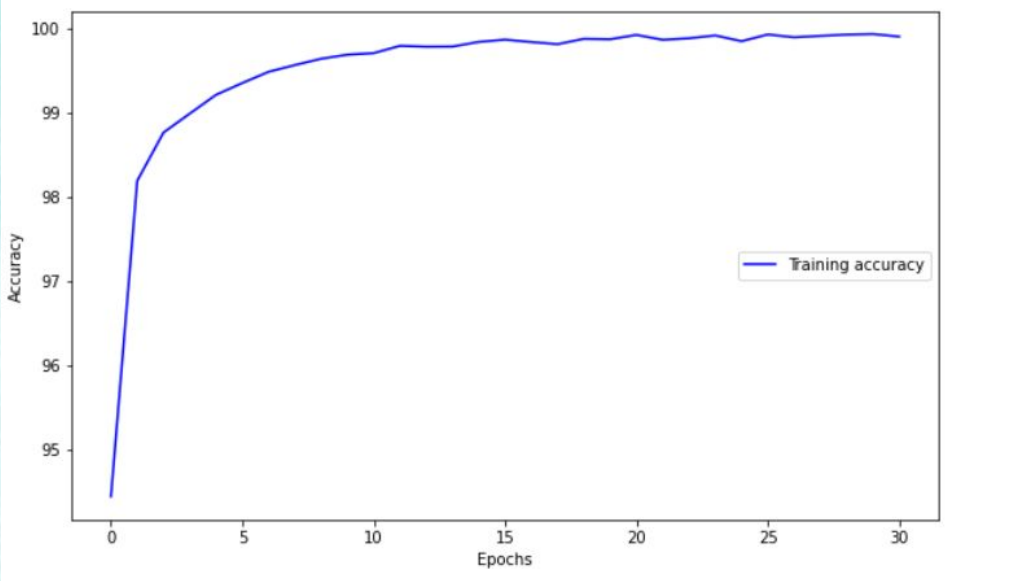
**To compare the optimizers the learning rate is taken constant as 0.001*

Adam

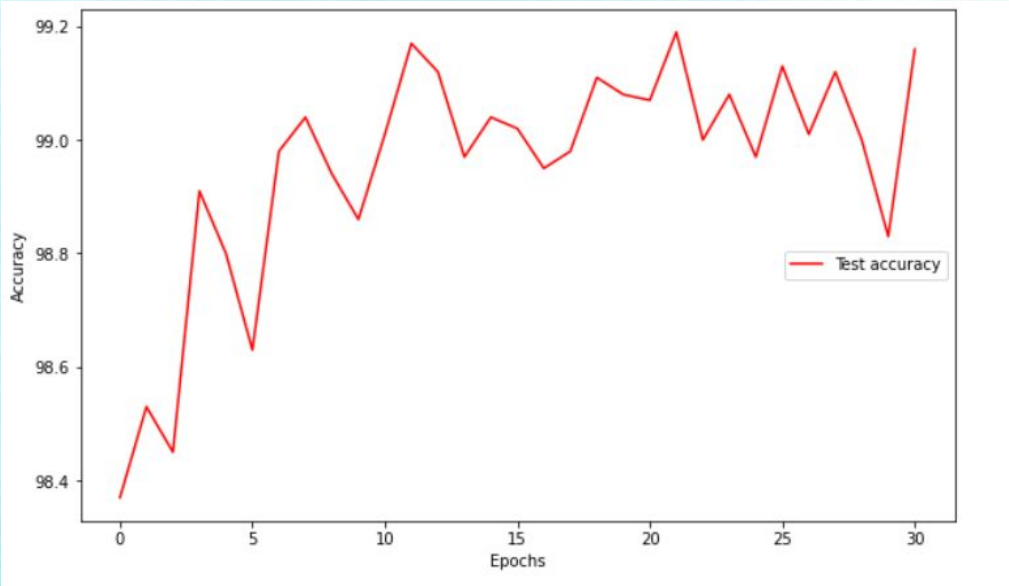
Loss:



Train Accuracy:



Test Accuracy:



The model is trained and validated for a maximum of 30 epochs or till its attains saturation (Test accuracy changes by an amount less than δ=0.02%), whichever happens earlier.

Results Obtained:

Max Test Accuracy = 99.19,

Which is obtained at epoch = 21

Time taken per epoch = 46.69s

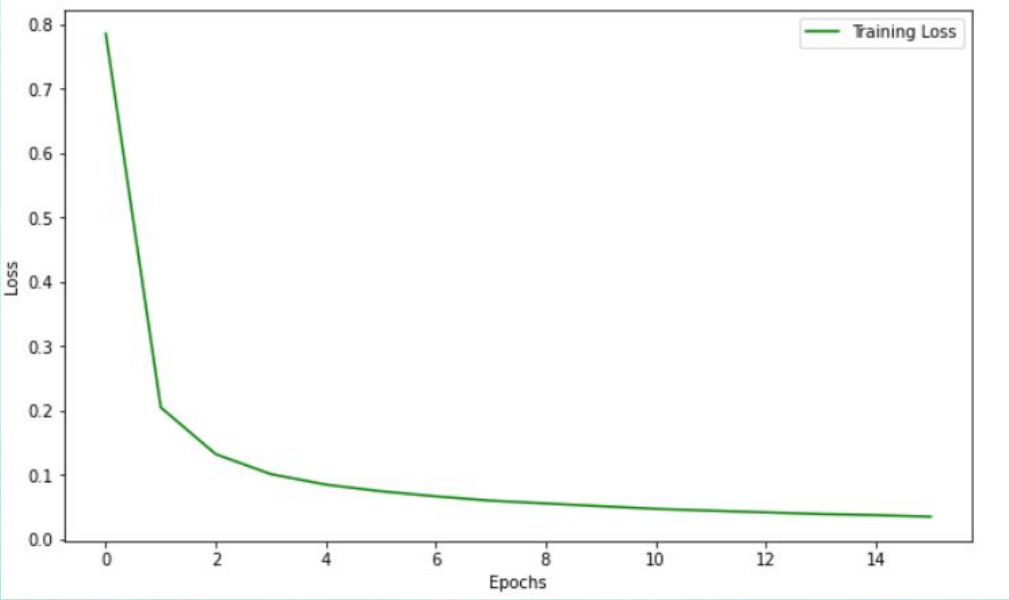
Total number of epochs = 30

Mean and Std. Dev of Train Acc = 99.49, 1.00

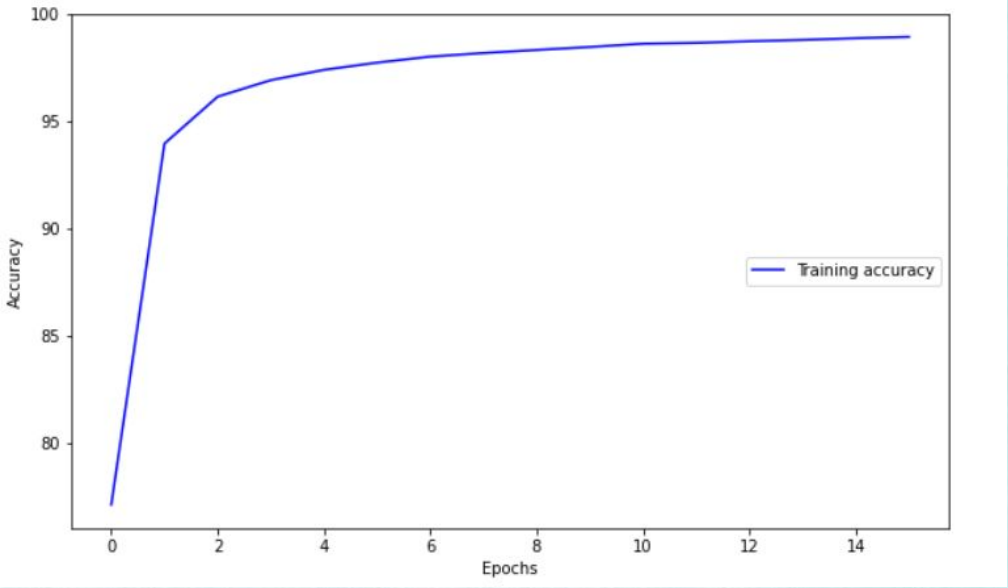
Mean and Std. Dev of Test Acc = 98.95, 0.58

SGD

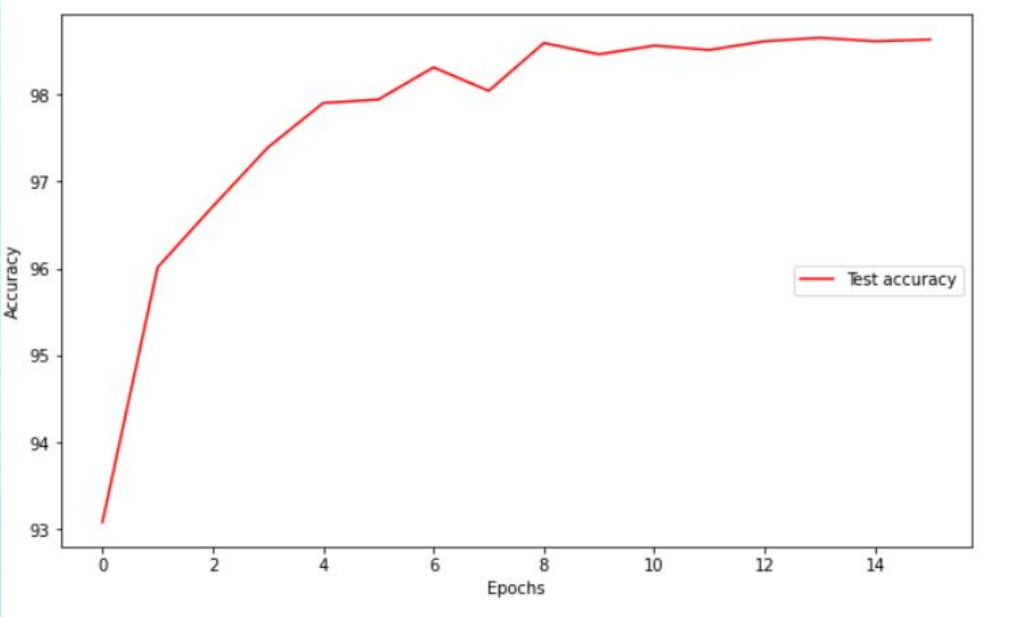
Loss:



Train Accuracy:



Test Accuracy:



The model is trained and validated for a maximum of 30 epochs or till its attains saturation (Test accuracy changes by an amount less than δ=0.02%), whichever happens earlier.

Results Obtained:

Max Test Accuracy = 98.65,

Which is obtained at epoch = 13

Time taken per epoch = 41.67s

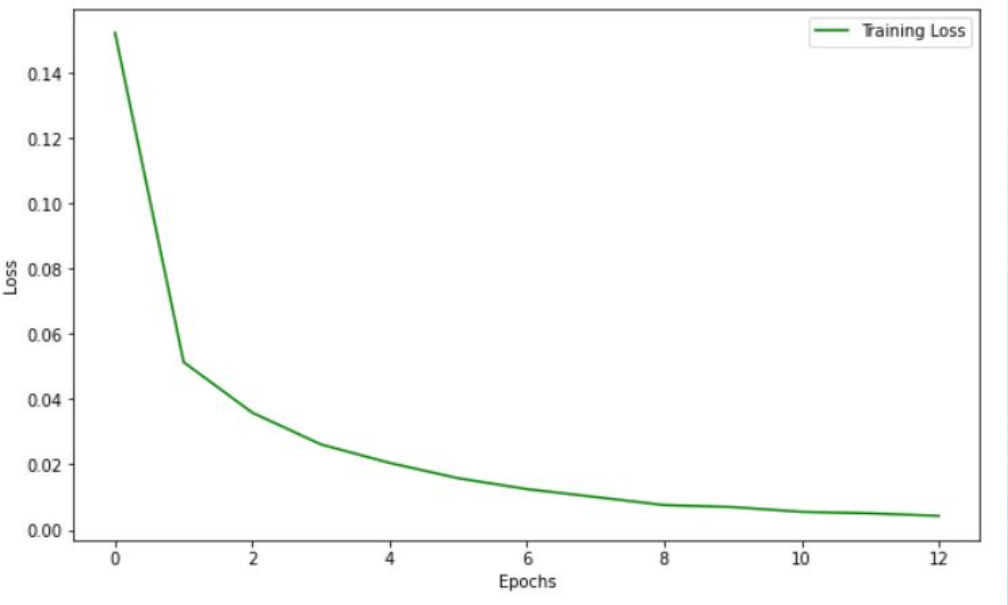
Total number of epochs = 15

Mean and Std. Dev of Train Acc = 96.55, 5.16

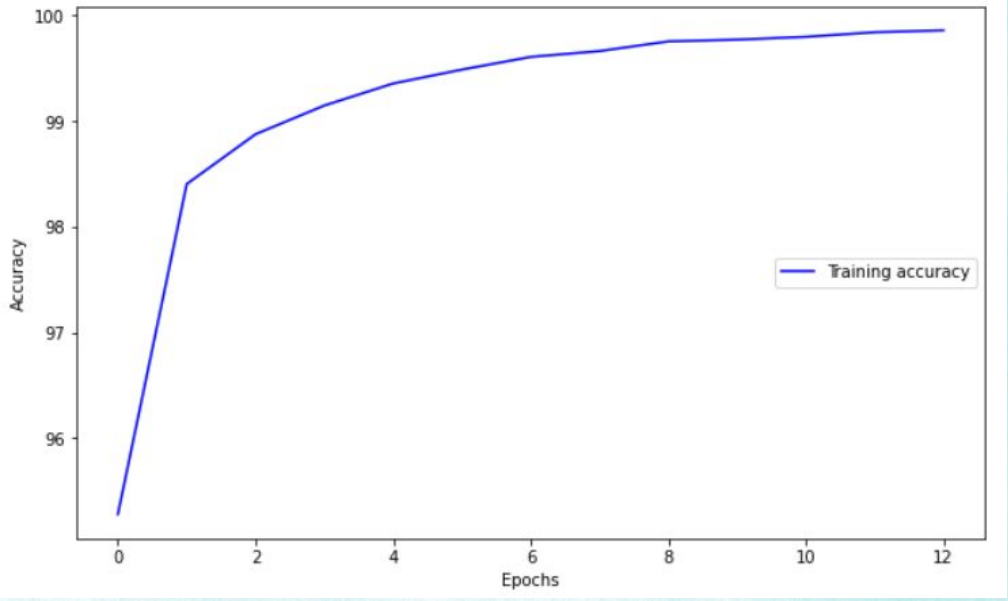
Mean and Std. Dev of Test Acc = 97.75, 1.85

RMSProp

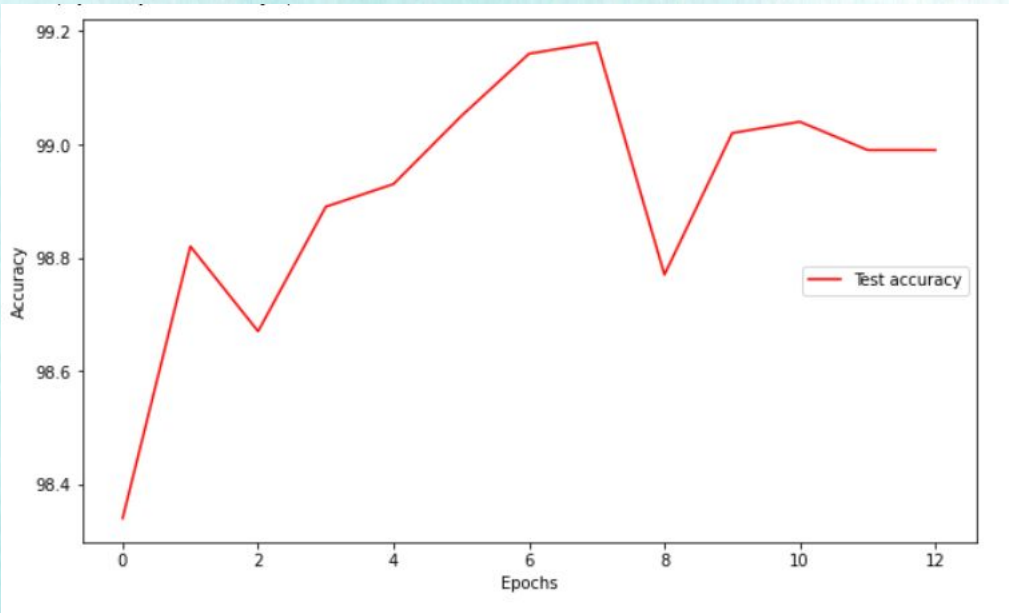
Loss:



Train Accuracy:



Test Accuracy:



The model is trained and validated for a maximum of 30 epochs or till its attains saturation (Test accuracy changes by an amount less than δ=0.02%), whichever happens earlier.

Results Obtained:

Max Test Accuracy = 99.18

Which is obtained at epoch = 7

Time taken per epoch = 45.80s

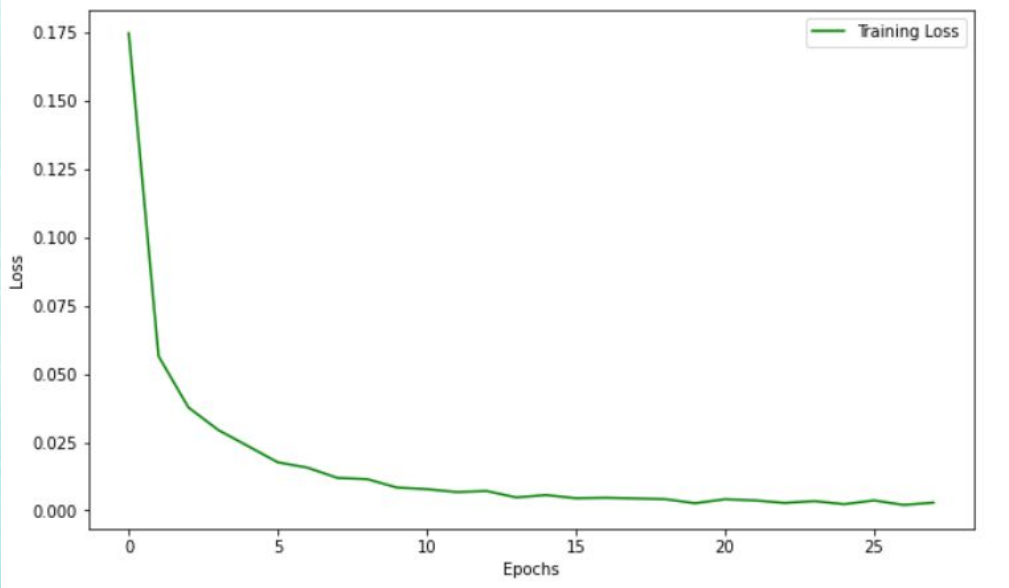
Total number of epochs = 12

Mean and Std. Dev of Train Acc = 99.14, 1.18

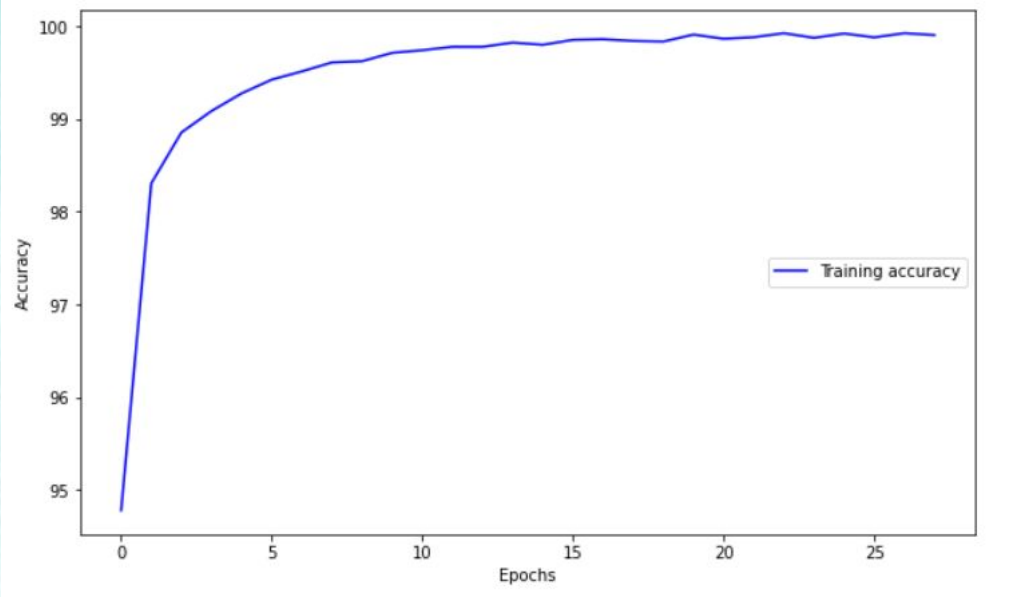
Mean and Std. Dev of Test Acc = 98.91, 0.31

AdamW

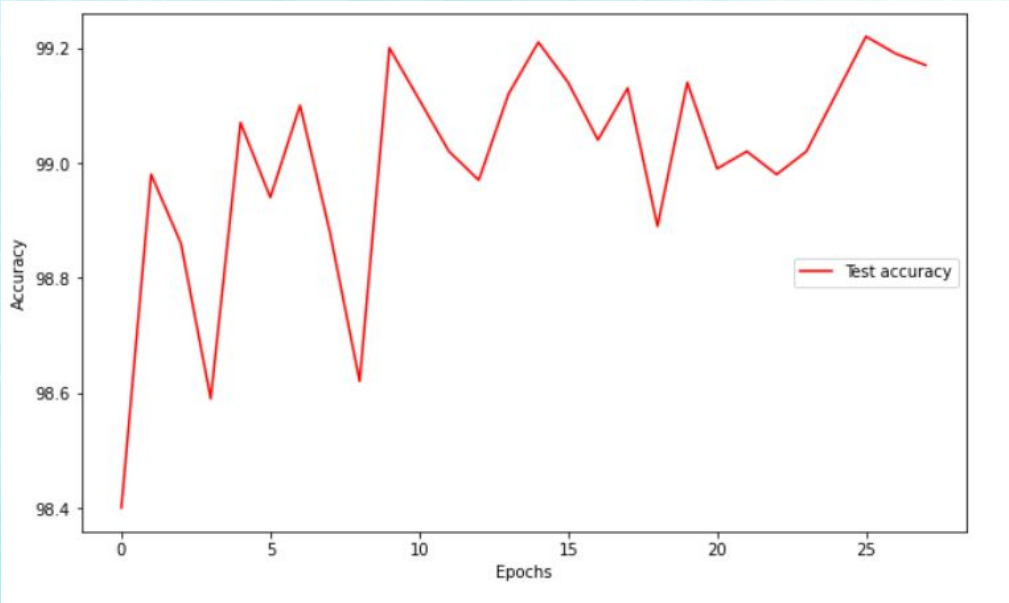
Loss:



Train Accuracy:



Test Accuracy:



The model is trained and validated for a maximum of 30 epochs or till its attains saturation (Test accuracy changes by an amount less than δ=0.02%), whichever happens earlier.

Results Obtained:

Max Test Accuracy = 99.22

Which is obtained at epoch = 25

Time taken per epoch = 43.32s

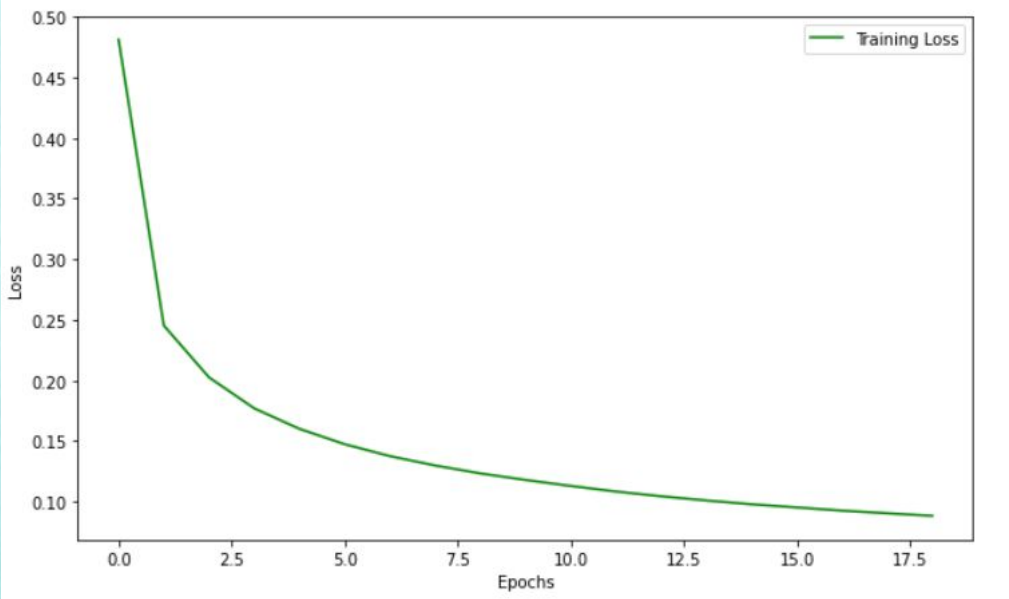
Total number of epochs = 27

Mean and Std. Dev of Train Acc = 99.48,0.97

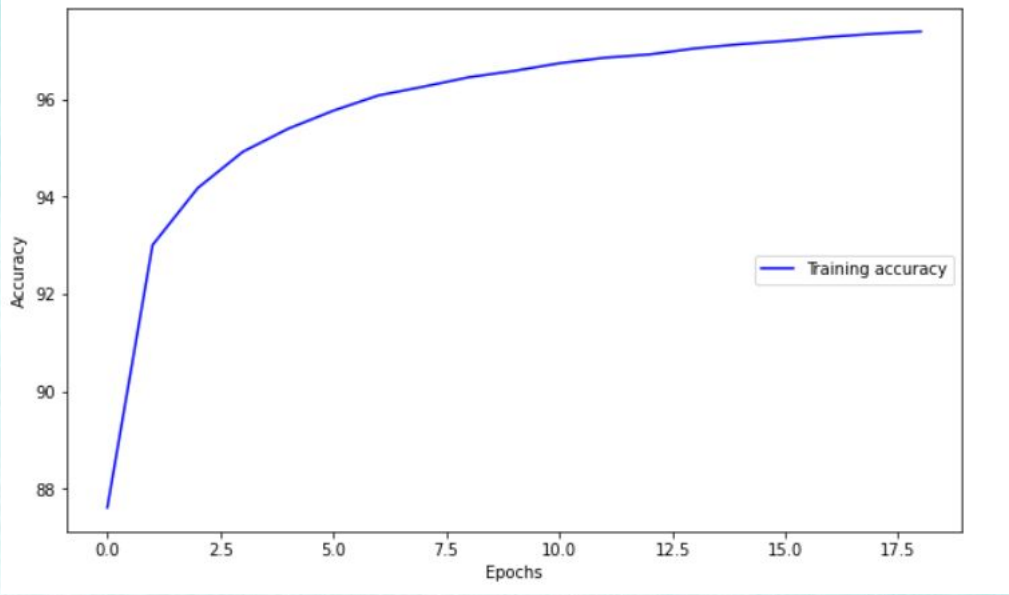
Mean and Std. Dev of Test Acc = 99.00, 0.51

Adagrad

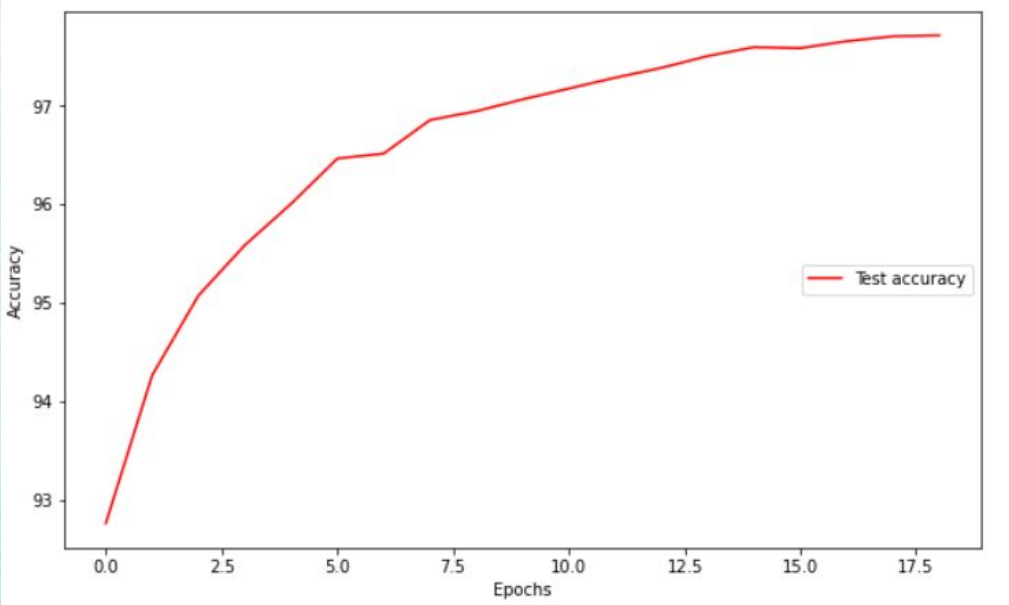
Loss:



Train Accuracy:



Test Accuracy:



The model is trained and validated for a maximum of 30 epochs or till its attains saturation (Test accuracy changes by an amount less than δ=0.02%), whichever happens earlier.

Results Obtained:

Max Test Accuracy = 97.71

Which is obtained at epoch = 18

Time taken per epoch = 41.73s

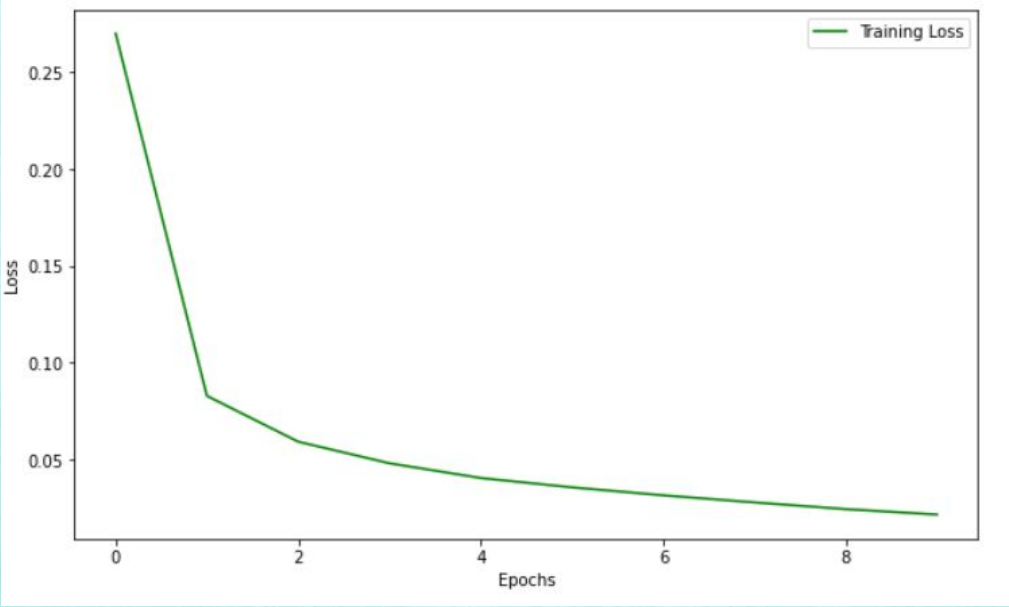
Total number of epochs = 18

Mean and Std. Dev of Train Acc = 95.79,2.24

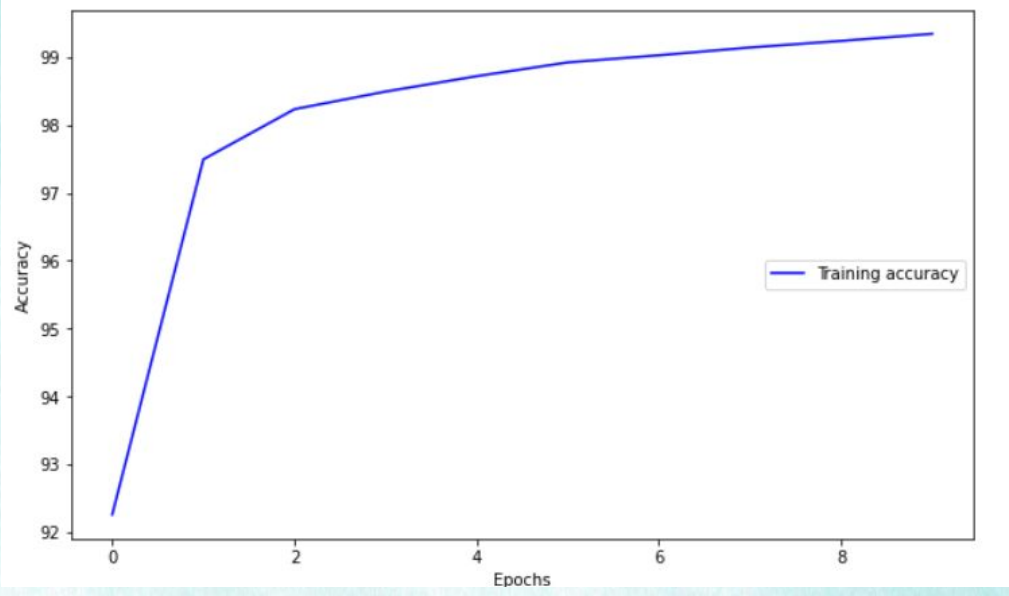
Mean and Std. Dev of Test Acc = 96.58, 1.51

Adamax

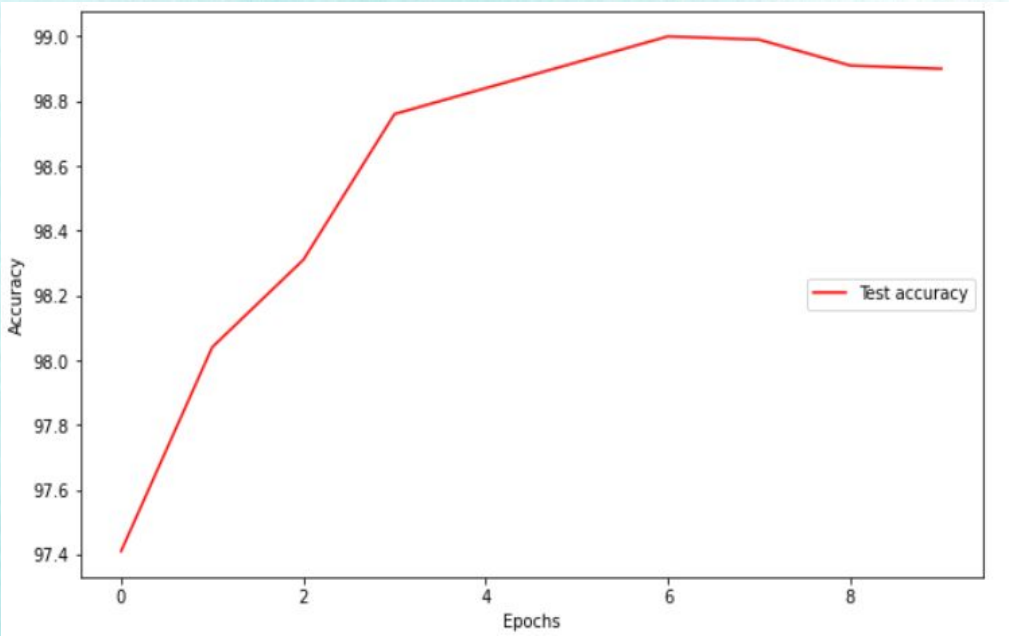
Loss:



Train Accuracy:



Test Accuracy:



The model is trained and validated for a maximum of 30 epochs or till its attains saturation (Test accuracy changes by an amount less than δ=0.02%), whichever happens earlier.

Results Obtained:

Max Test Accuracy = 99.00

Which is obtained at epoch = 6

Time taken per epoch = 47.30s

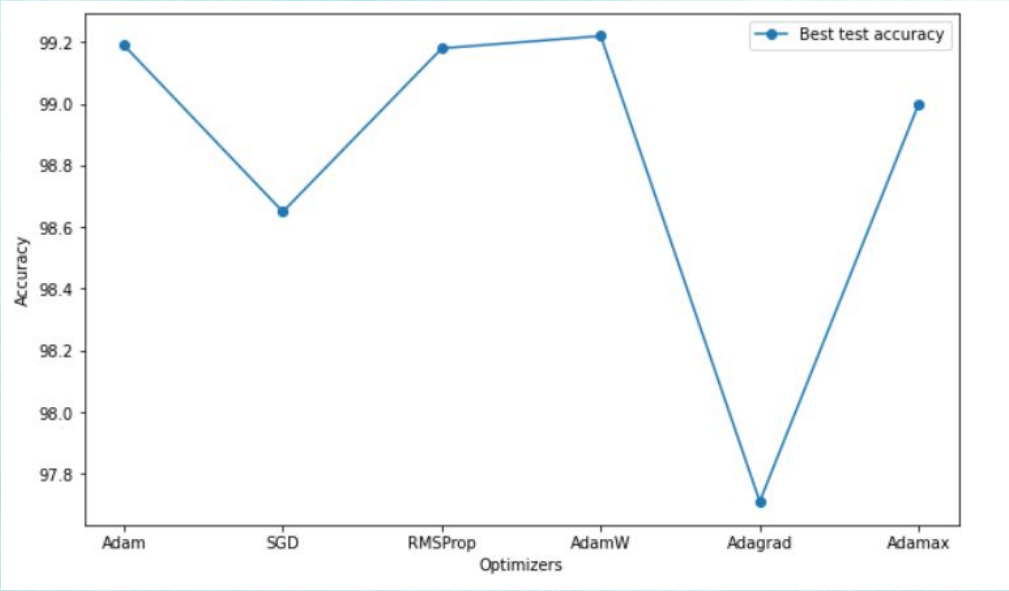
Total number of epochs = 9

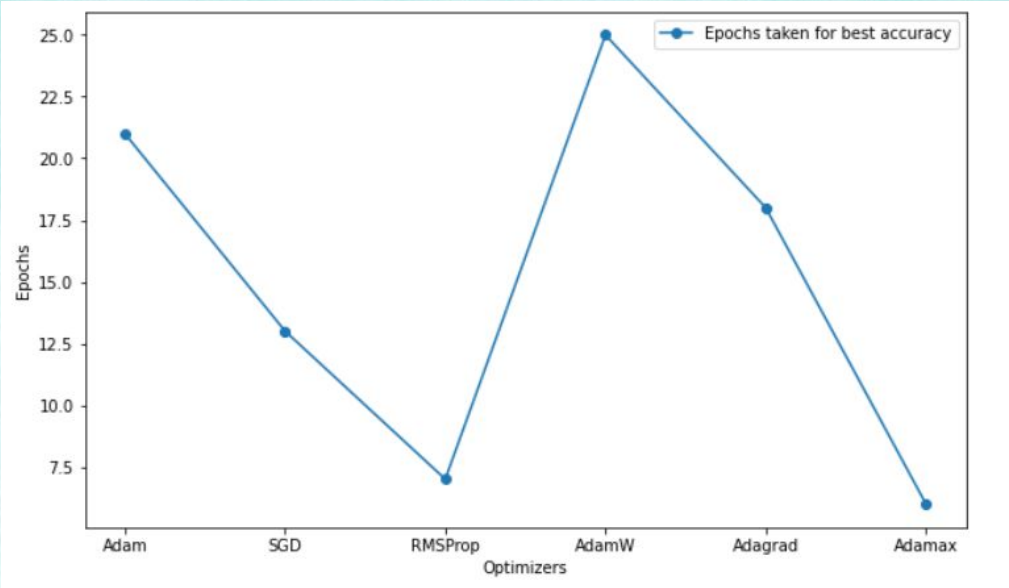
Mean and Std. Dev of Train Acc = 98.08,2.01

Mean and Std. Dev of Test Acc = 98.60, 0.72

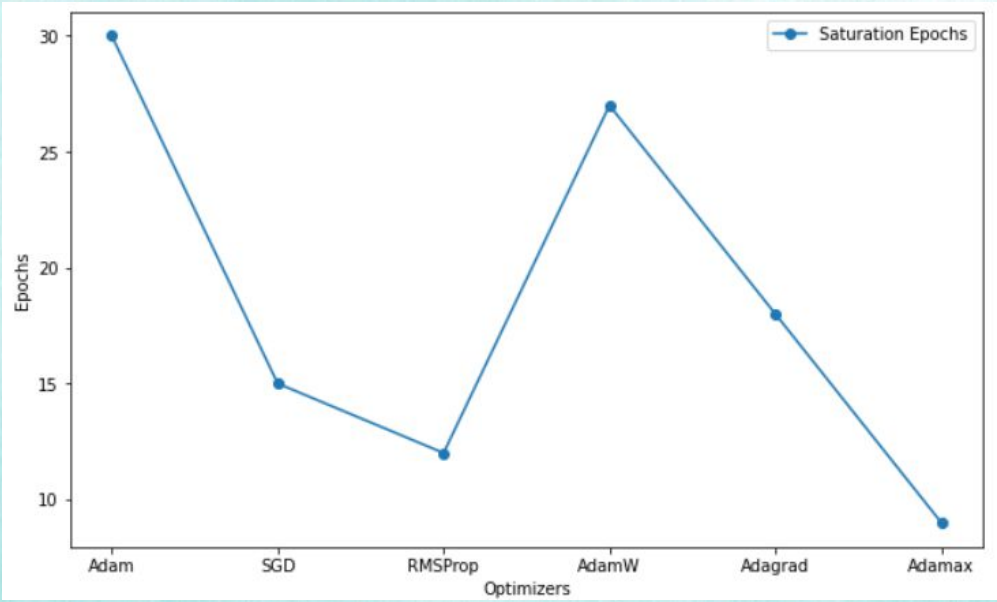
OBSERVATIONS

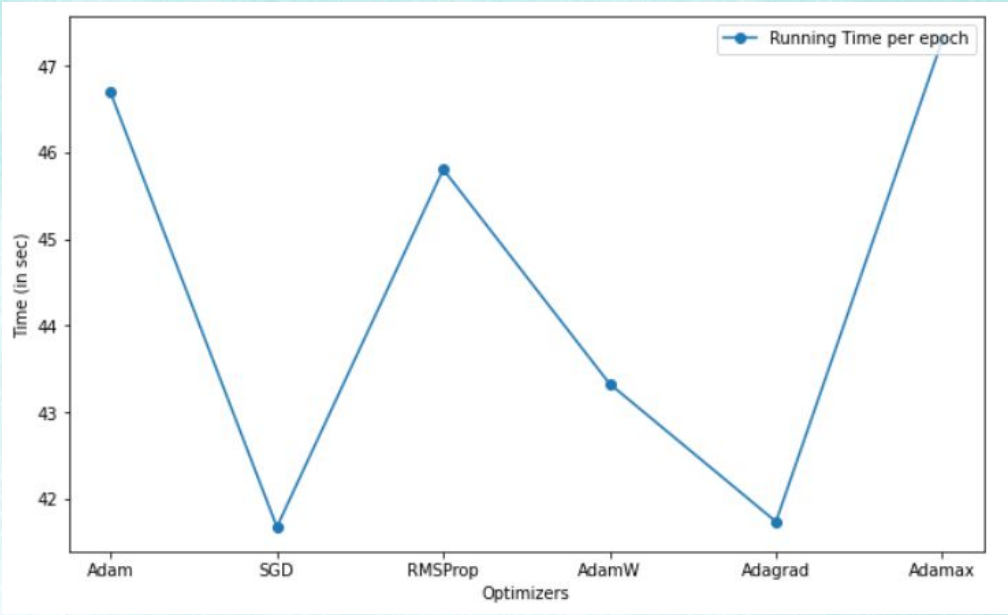
Test Accuracy and epochs taken:



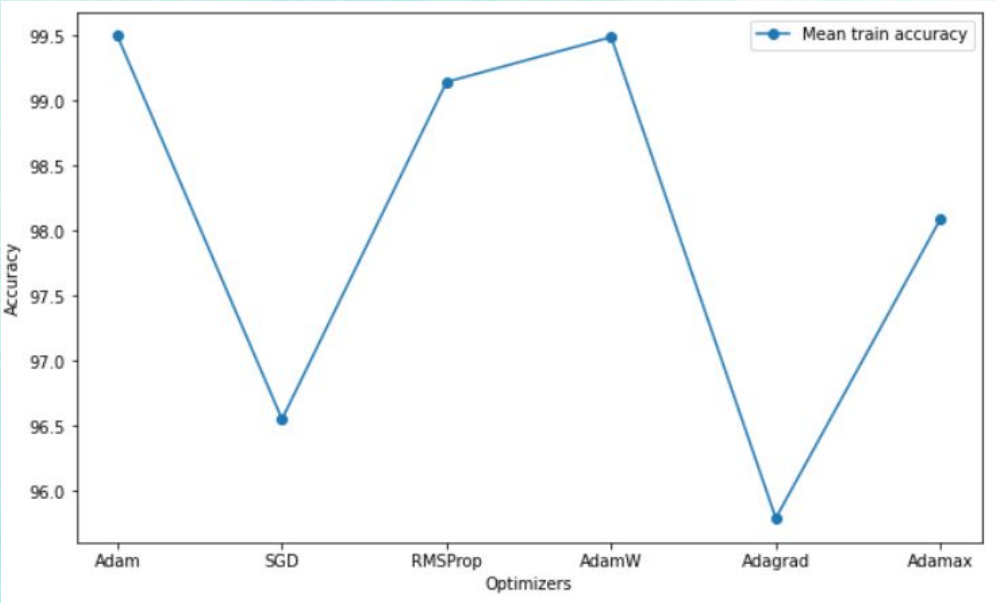


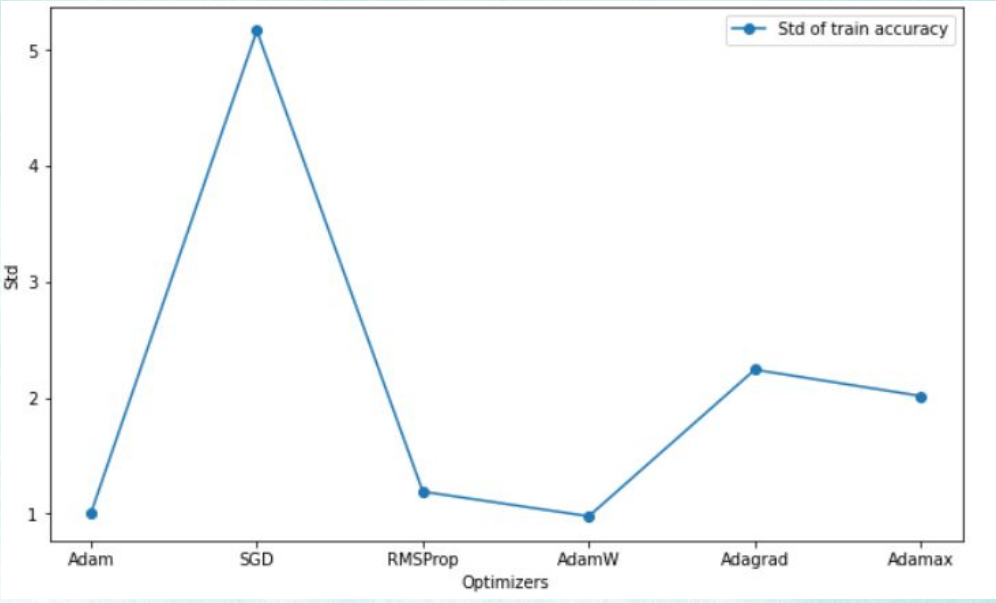
Total epochs and run-time per epoch:



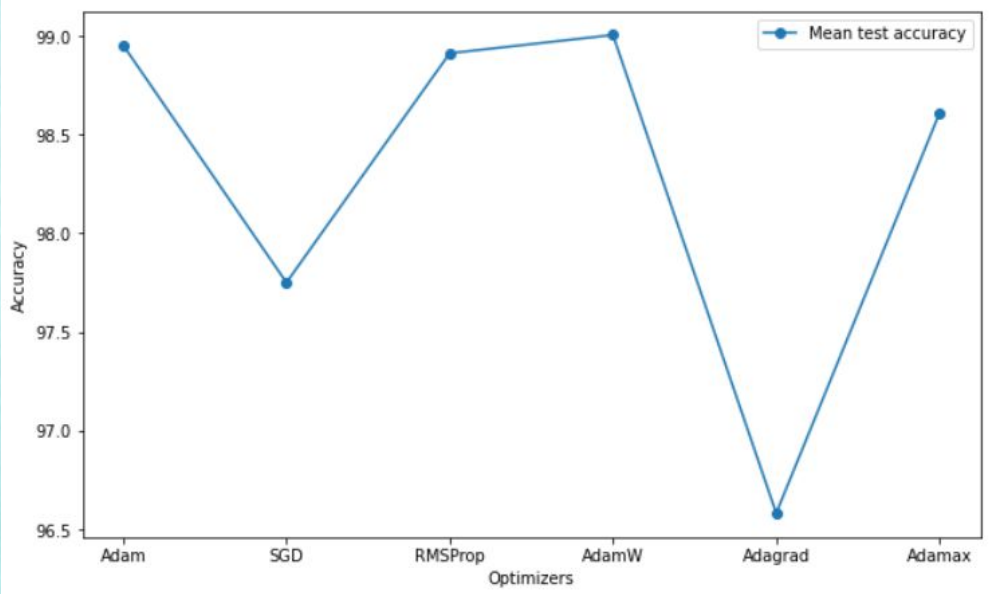


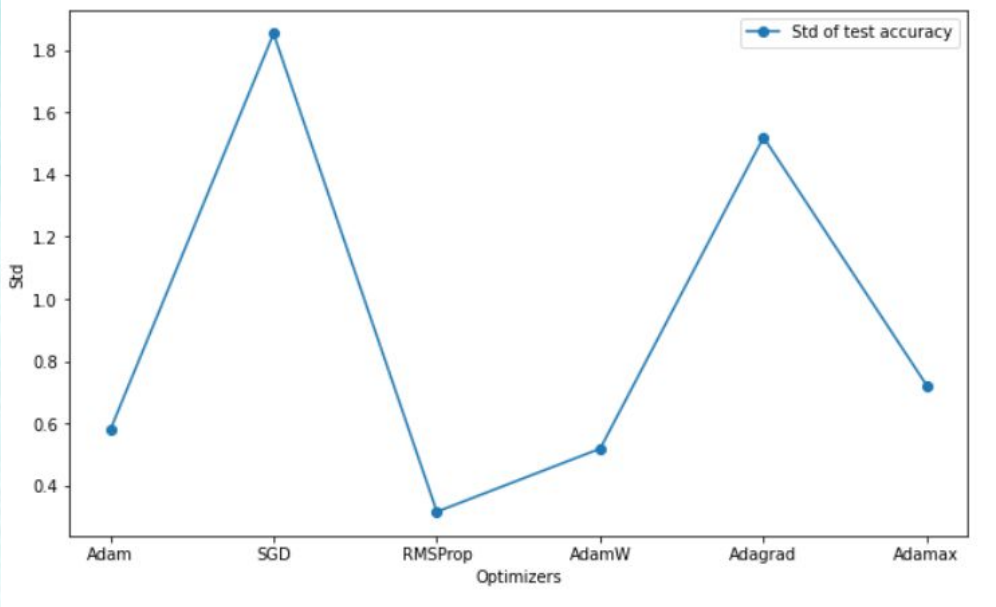
Mean and Standard Deviation of train accuracy:





Mean and Standard Deviation of test accuracy:





INFERENCE

The highest test accuracy for the model is calculated as 99.22% which is obtained by using AdamW optimizer after running 25 epochs. Also it is seen that there is a sharp convergence in loss and a high test accuracy using RMSProp and Adamax which gave a maximum test accuracy of 99.18% and 99% after 7 and 6 epochs respectively.
Both AdamW and RMSProp optimizers gave a high mean test accuracy and a low standard deviation in test accuracies.