ELL-793 – Lab Assignment-2

Report By:

Mukku Vamsi Krishna Reddy (2019CS10374)

Guruvu Surya Sai Prakash (2019EE10481)

Q1)

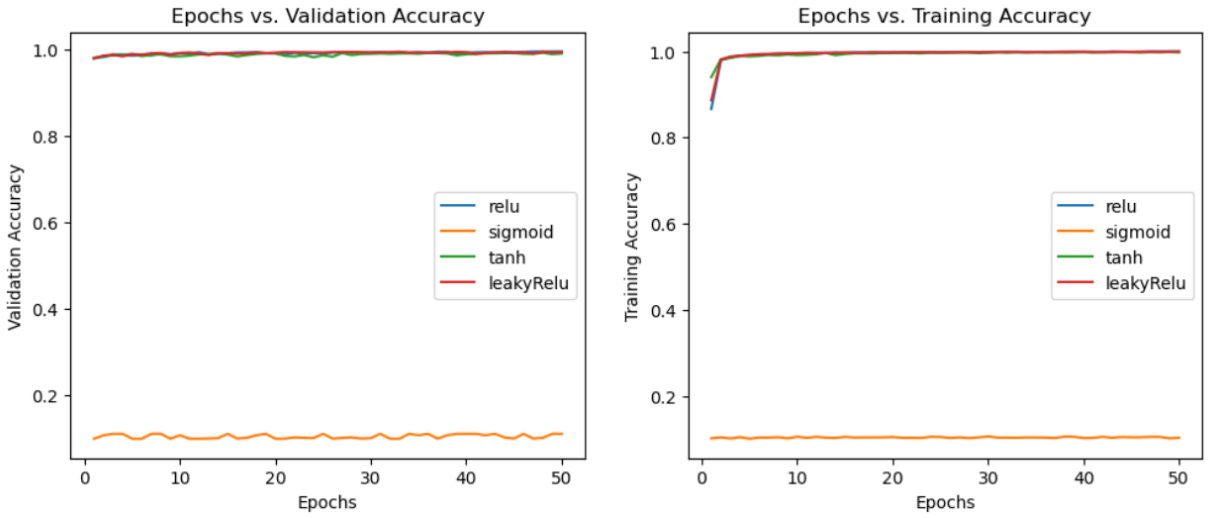
a) Which Activation is better?

For this experiment we chose the following setting

No of Epochs: 50
Optimizer: Adam
Learning rate: 0.0001
Batch size: 128

Now we train the network by varying the activation functions.

The results are shown in the below plot.



It can be observed that except the sigmoid function, all other functions are converging with almost the same speeds and any of these three functions would serve our purpose.

b) What learning rate is better?

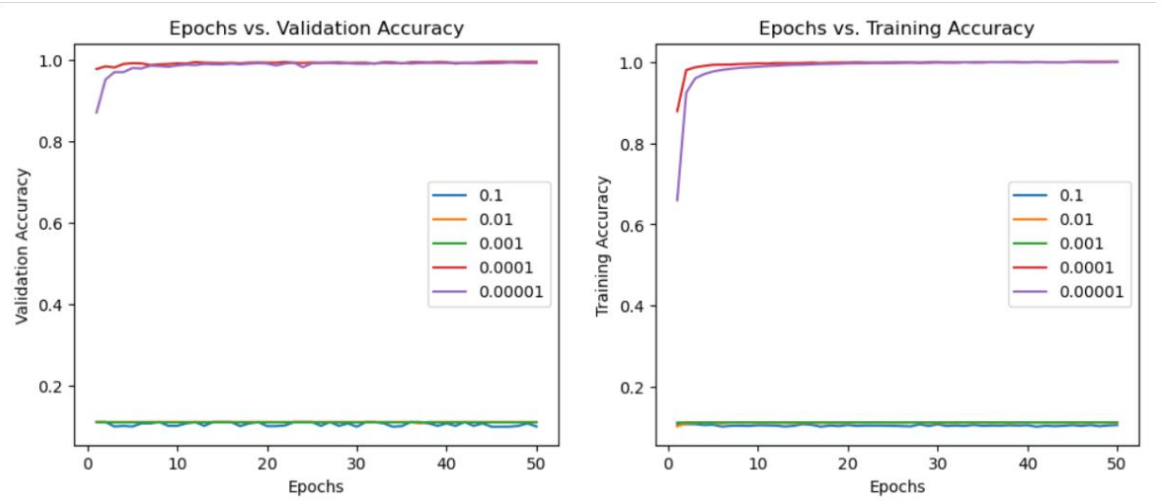
Since, different optimizers will have different best learning rates, we conducted the experiment over all the optimizers, and found the best optimizer for each case.

The parameter setting is as follows:

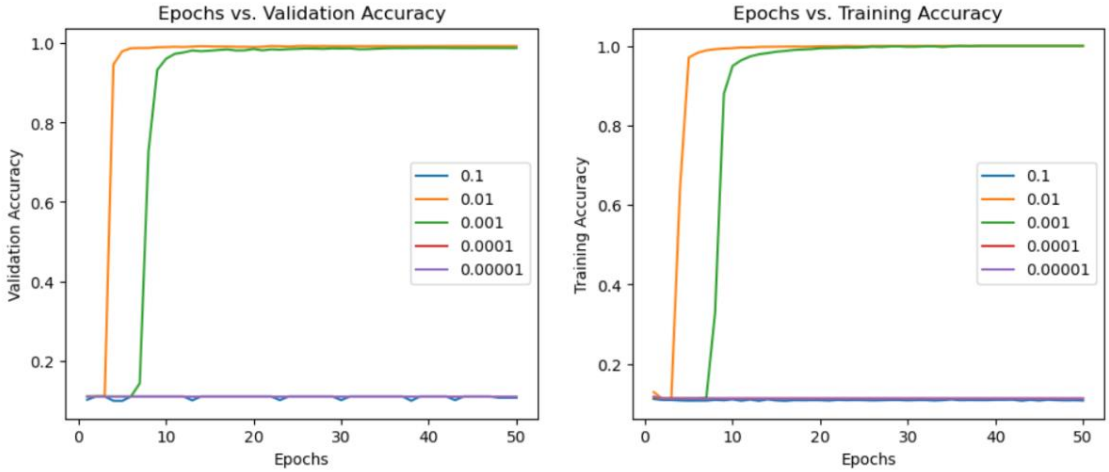
No of Epochs: 50
Batch size: 128
Activation function: ReLu

The results are shown in the below plots:

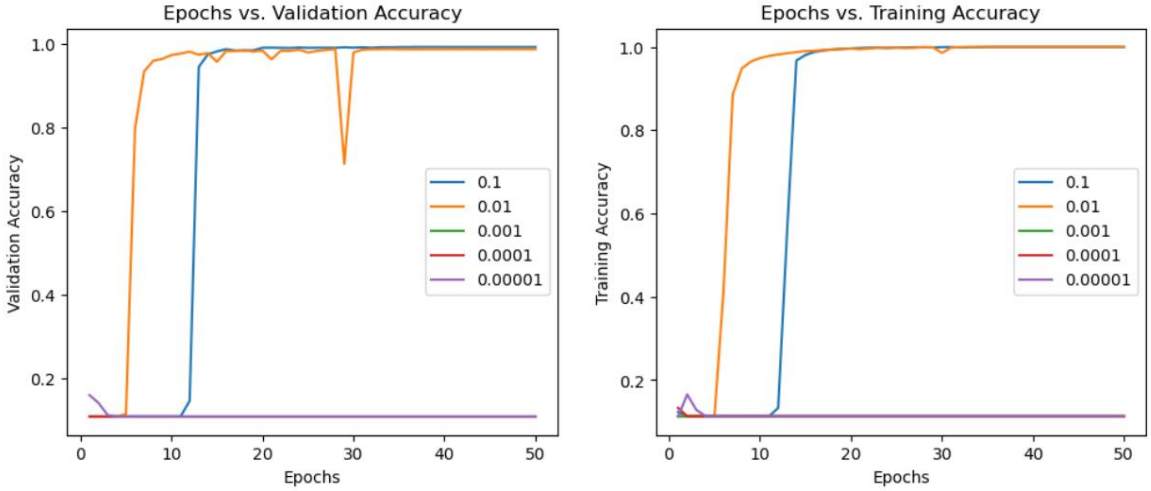
Adam Optimiser:



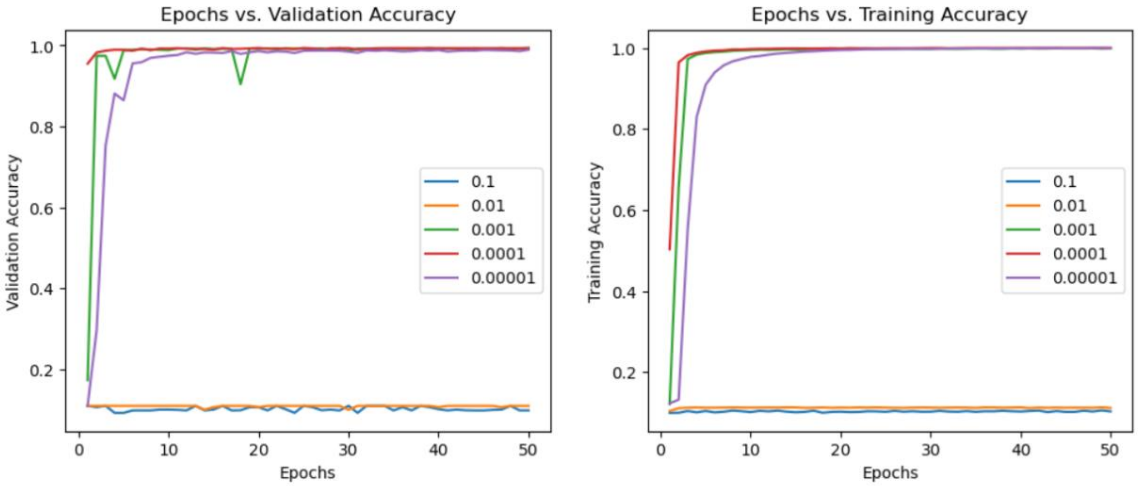
Momentum based gradient descent:



Vanilla Gradient descent:



RMS prop:



So, the final better learning rate is given is the following table:

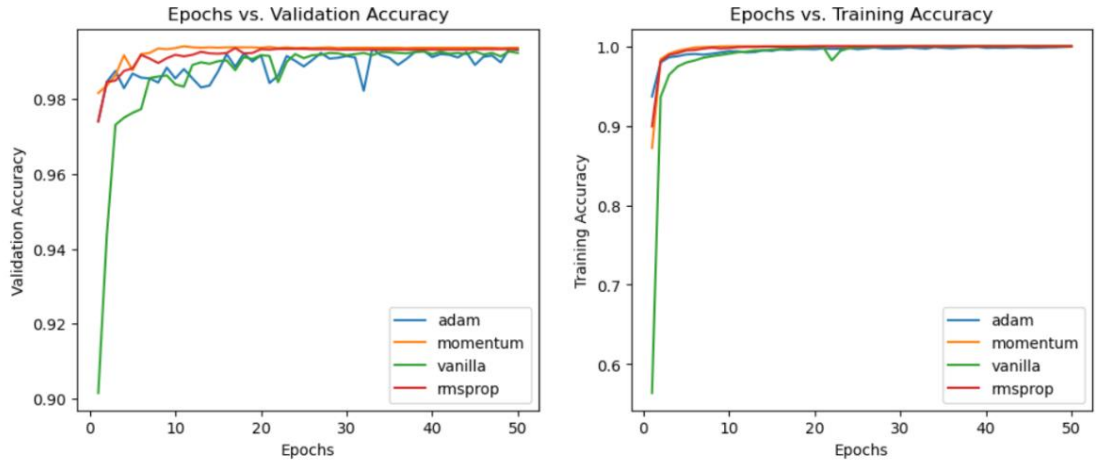
Optimizer	**Learning rate**
Adam	0.0001
Momentum based GD	0.01
Vanilla GD	0.01
RMS prop	0.0001

c) Role of Optimizer

For this experiment, the parameter setting is as follows:

No of Epochs: 50
Batch size: 128
Activation function: ReLu

The learning rate is chosen depending on the optimizer(the values are taken from the results of experiment (**b)**).



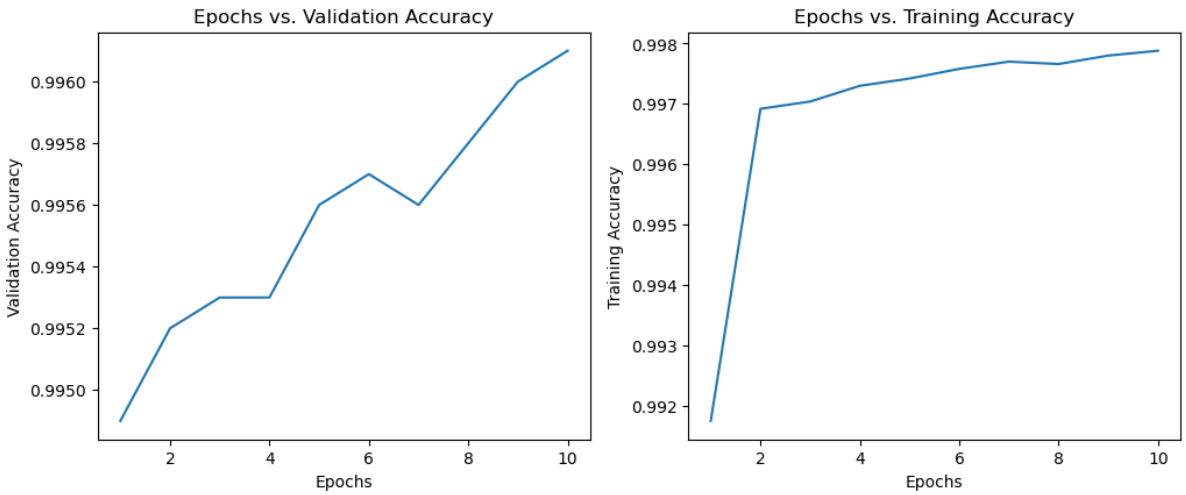
It can be observed that except vanilla_gd all other optimizers converge with similar speeds and any of them would serve our purpose.

d) Transfer Learning

Now, performing transfer learning on the new task by choosing the following setting.

No of epochs = 10
Optimizer = Adam
Learning rate = 0.0001
Activation function = tanh

The result is as follows:



It is observed that the convergence is achieved in less than 5 epochs.

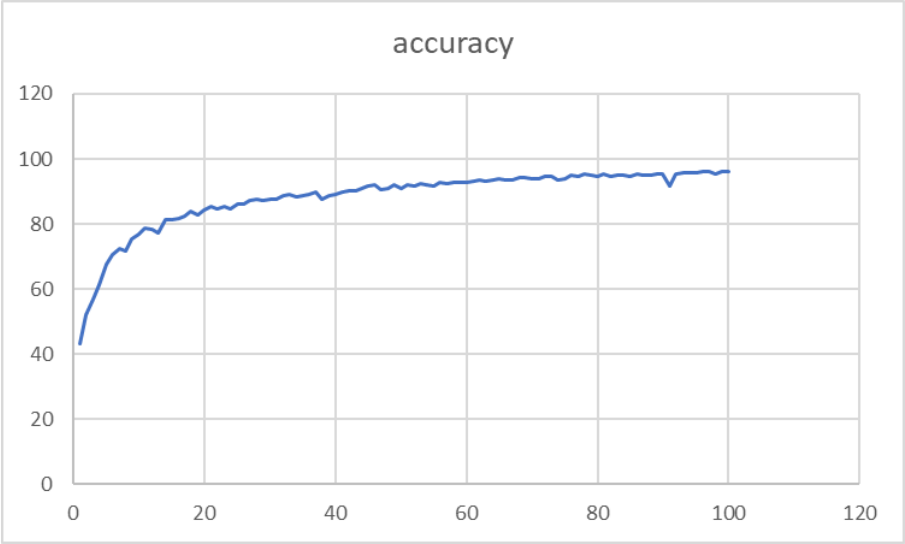
The final accuracies are as follows:

Training Accuracy: 99.8
Validation Accuracy: 99.6
Testing Accuracy: 99.5

Q2)

a)

(i) training all the layers of resnet-18 :

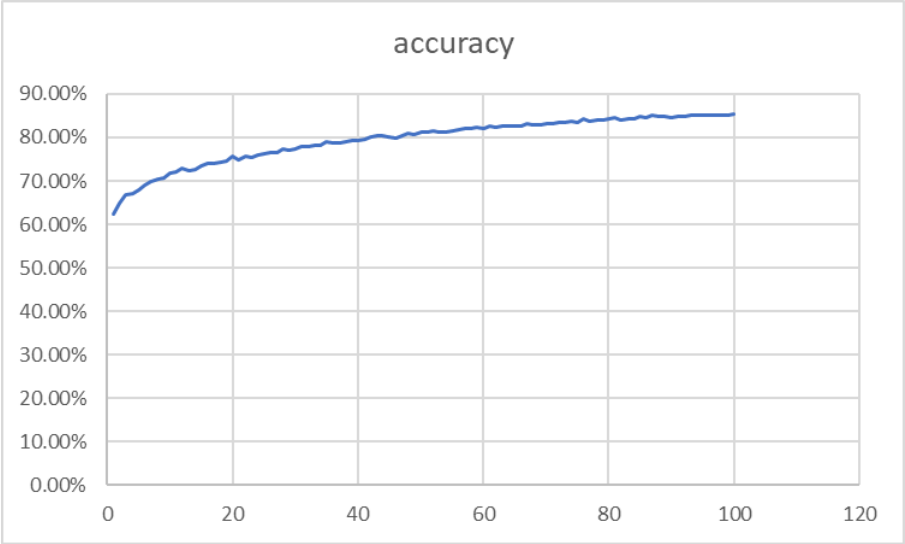


plot of training accuracy vs no.of epochs

Final Training accuracy: 96%

Test accuracy : 85%

(ii) **training only the last two layers of resnet-18:**



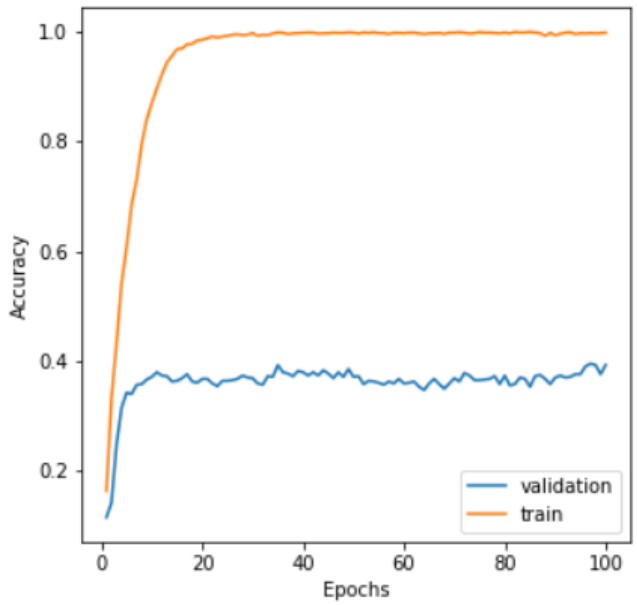
plot of training accuracy vs no.of epochs

Final Train accuracy: 85.50%

Test accuracy: 71%

b)

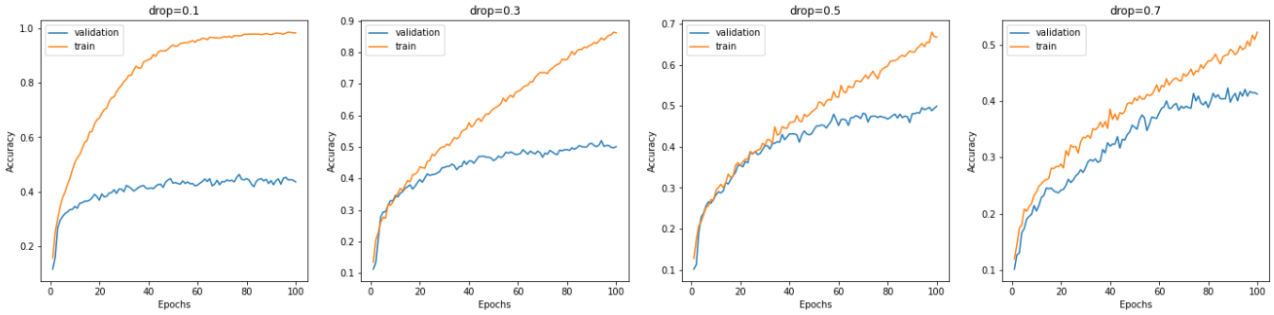
Results without any dropouts:




```
Train accuracy: 99.78
Validation accuracy: 39.33
Test accuracy: 38.92
```

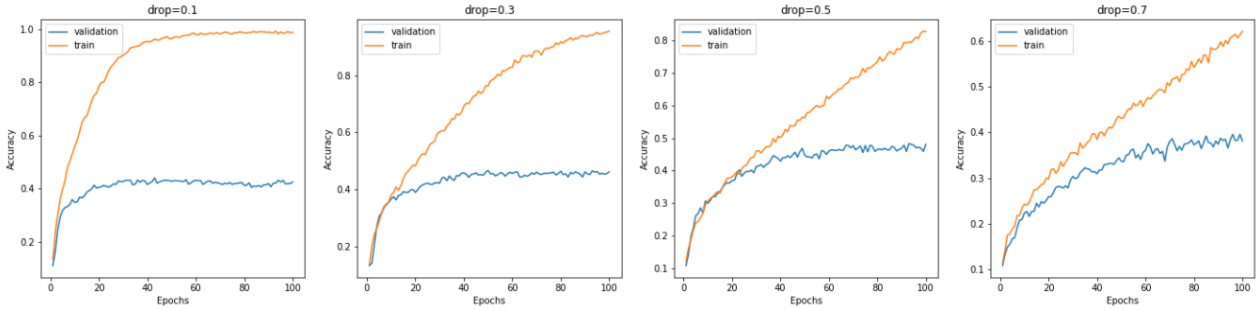
It is overfitting.

Drop out after layer 5:

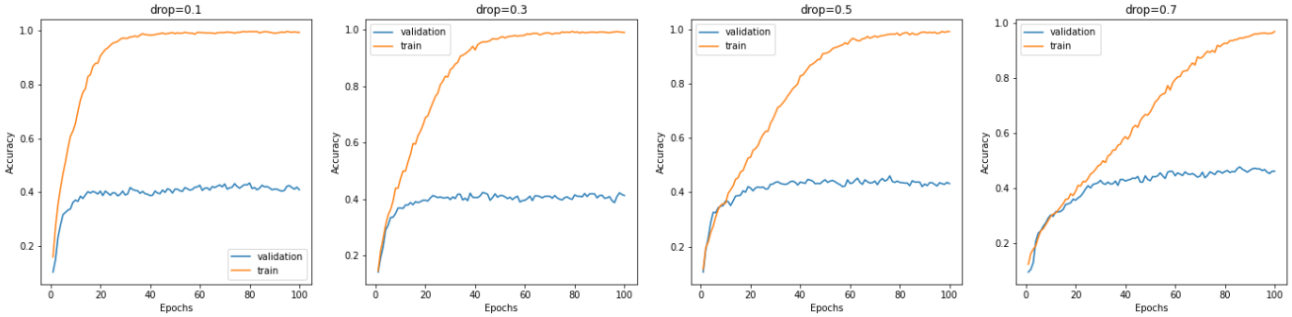


It is observed that, for drop_out_rate = 0.1, it is overfitting. For drop_out_rate=0.3, it is better, and it has 44% validation accuracy. For other dropout rates it is underfitting

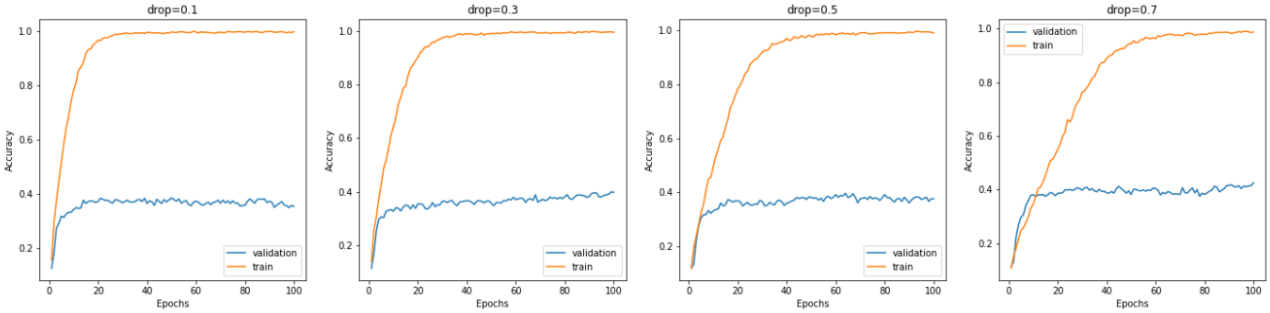
Drop out after layer 9:



Drop out after layer-13:



Drop out after layer-17:



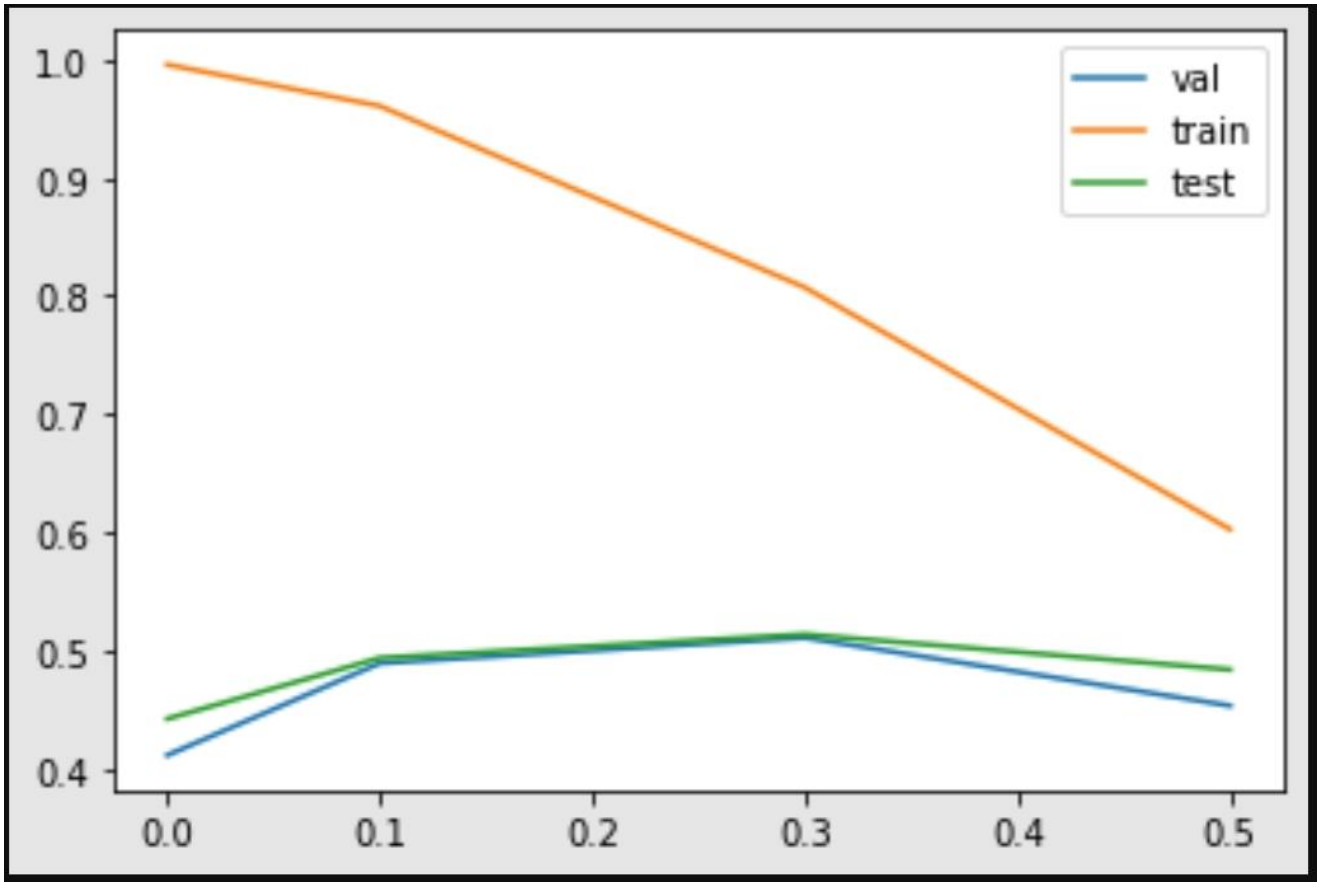
It is observable that the dropouts after layer-9,13,17 are not creating much impact as compared to layer-5.

Data Augmentations:

1) Randomly flipping the image horizontally.
2) Randomly rotating the image by 90 degrees.
3) Randomly adjusting the brightness
4) Randomly adjusting the contrast of the image.
5) Randomly zooming the image.

After these Data Augmentations and performing dropout after layer 5 (since, we observed better accuracy here in previous experiment)

The train, test, validation plot is shown below,

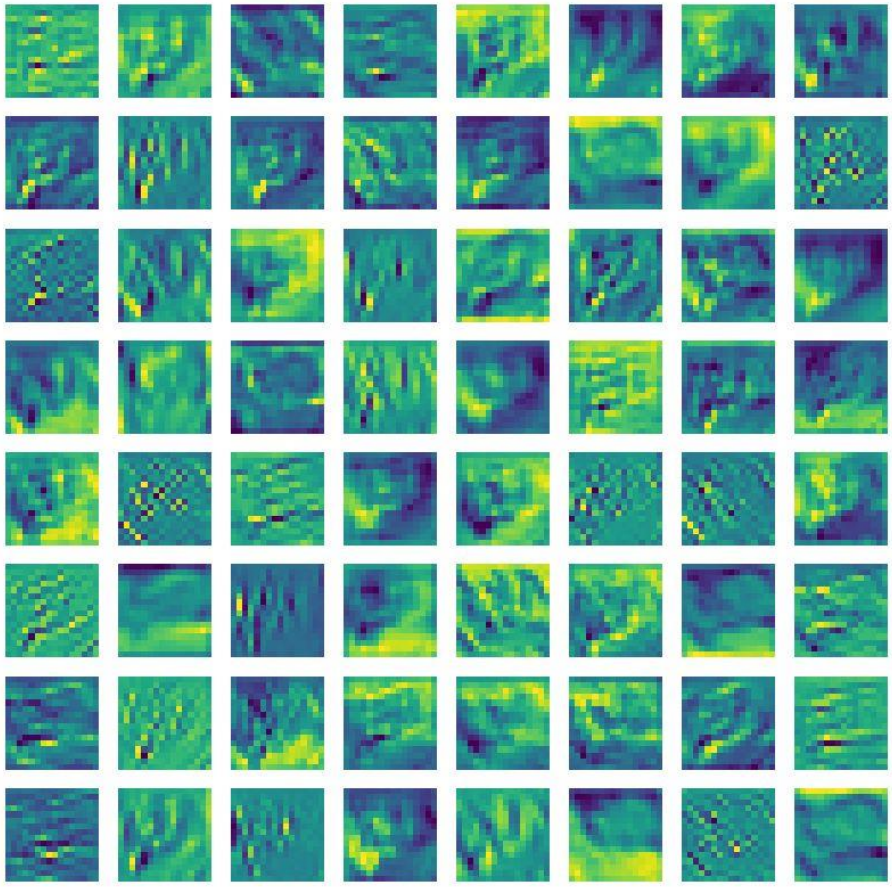


It is observed that for dropout of 0.3, the overfitting has improved, with validation and test accuracy nearly equal to 52%.

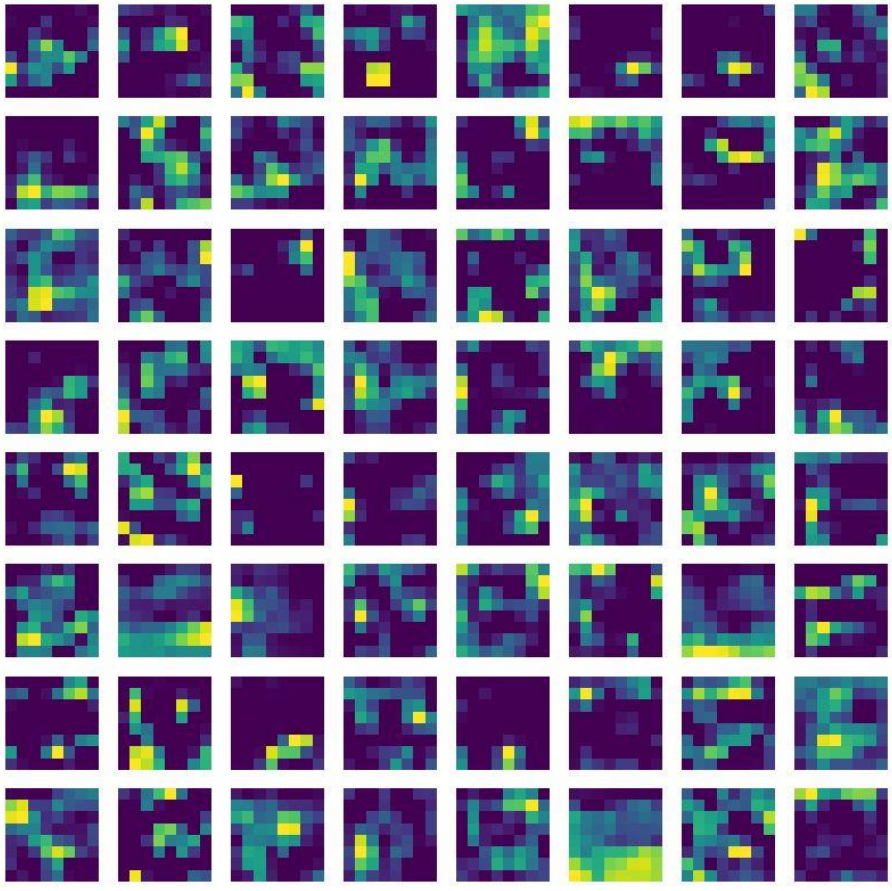
c)

For the resnet-18 after training,

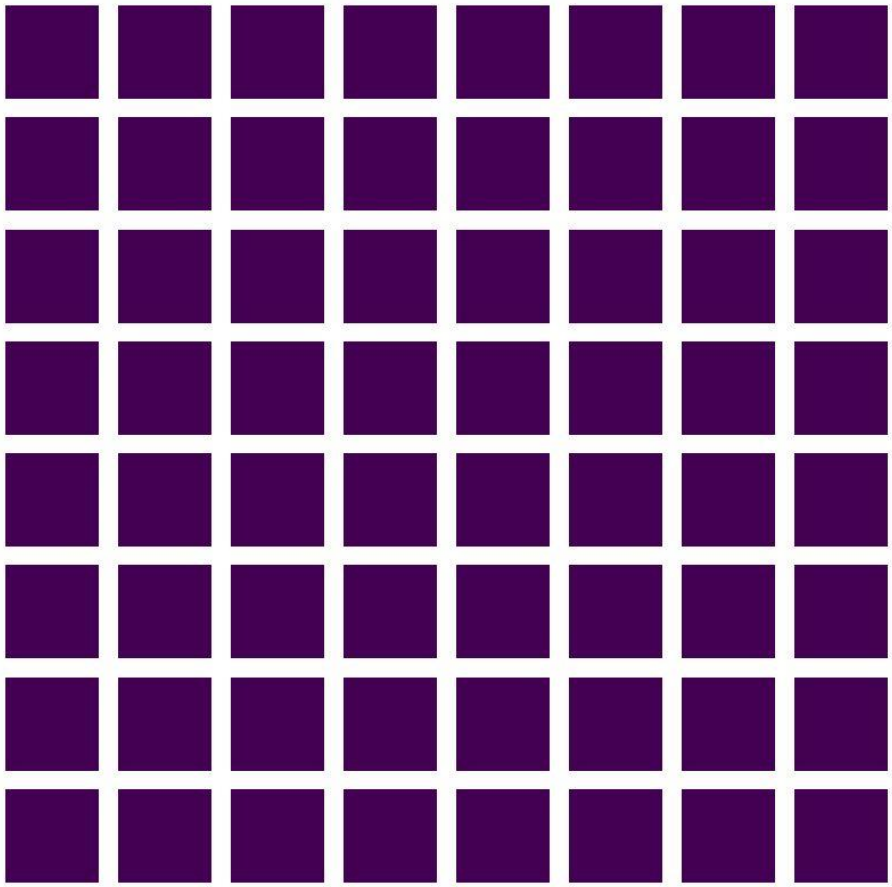
For the first image, first layer:



For the first image,fifth layer:



For the eight layer:



Heat map is getting darker as we move deeper into the network is because the activations of the deeper layers are more sparse and less diverse compared to the earlier layers. This might be due to the effect of the non-linear activation functions and pooling layers that are applied to the feature maps as the information is processed through the network.