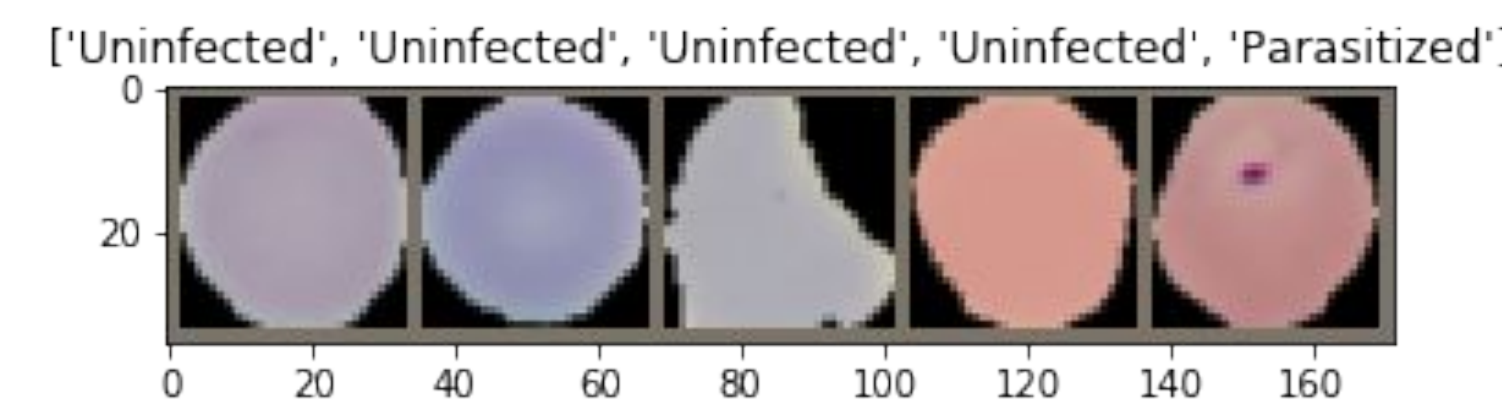


ABSTRACT

Malaria is a disease caused by a parasite, female mosquito bite. In this project we aim to classify parasitized v/s uninfected red blood cell images using convolutional neural network. Our dataset source is kaggle.

DATA

- Our dataset consists of two classes parasitized and uninfected red blood cell images. Total images = 27,560
- Training
 - Parasitized = 10,780
 - Uninfected = 10,780
- Validation
 - Parasitized = 3000
 - Uninfected = 3000
- The split of this was 78% for training and 22% for validation.
- The parasitized cells can be identified as a red dot in the cell. If there is no red dot then the cell is uninfected.
- We set up all our experiments by processing data in batches of size 5.



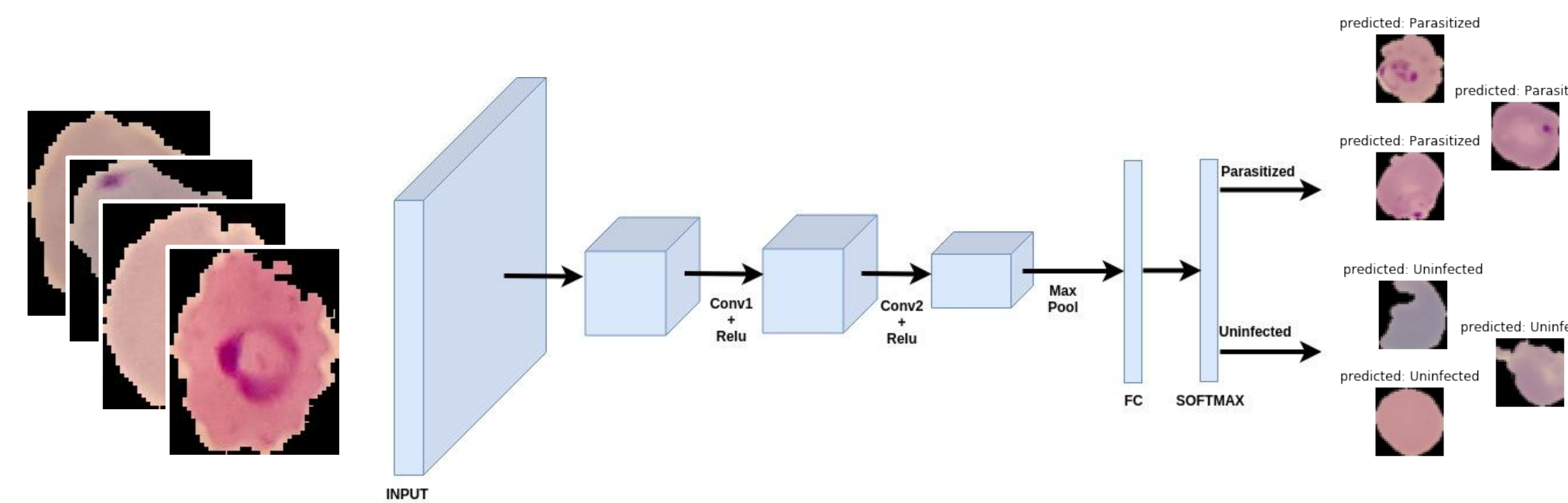
MATERIALS

- Our methods include a custom build convolutional neural network.
- For this project we used pytorch framework by facebook.
- We trained our model on NVIDIA corporation GP104M [Geforce GTX 1070 Mobile].

References

- Data Source : <https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria>

NETWORK ARCHITECTURE



- Our model consists of two convolution layers, one max pool layer.
- The filter size for the first convolutional layer we use is 3*3 and number of filters we use is 12.
- This layer is followed by a RELU activation layer.
- For the second convolutional layer we use 12 3*3 filters followed by a RELU activation layer.
- We then pass our output to the fully connected layer.
- There are two output neurons in the final layer representing the two classes. We normalize probabilities using softmax.

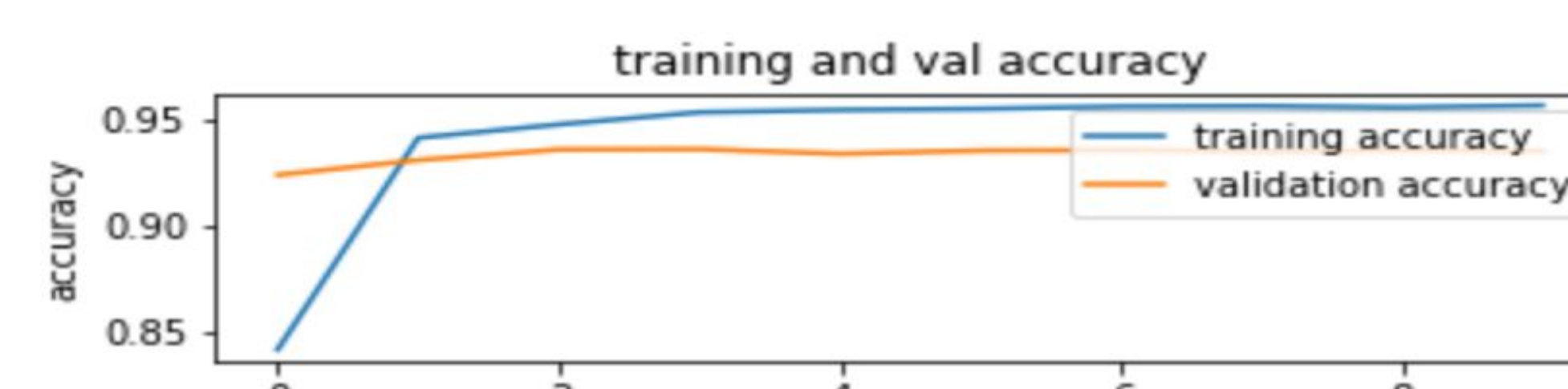
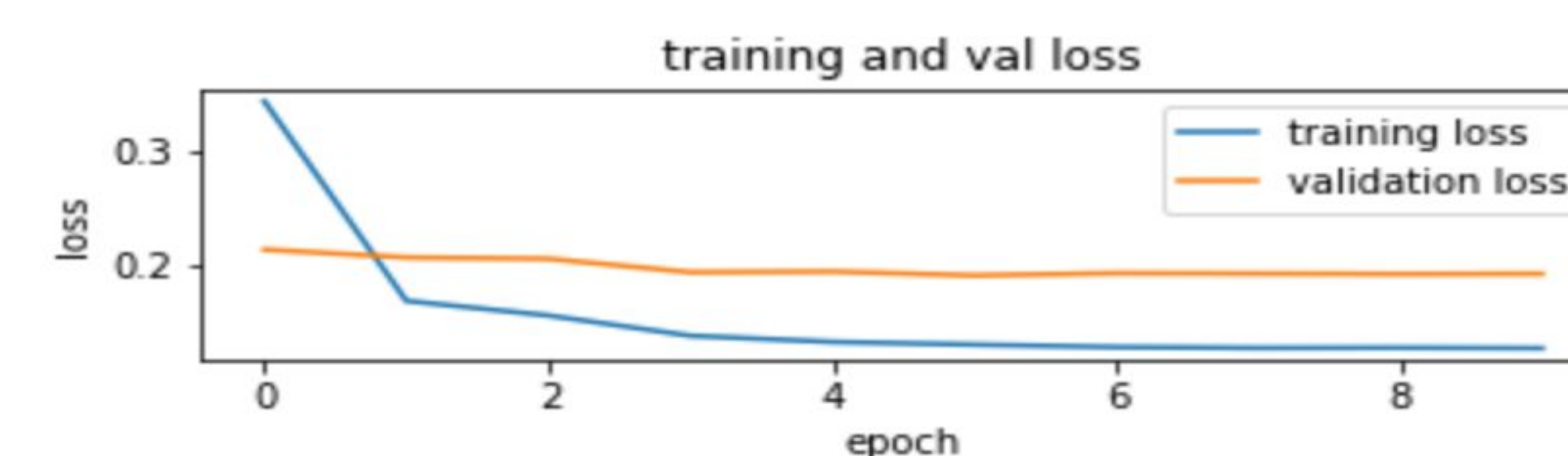
EXPERIMENTS WITH OPTIMIZATION TECHNIQUE

Loss function: All errors were calculated using cross entropy loss formula for binary classification is as follows:

$$-(y \log(p) + (1 - y) \log(1 - p))$$

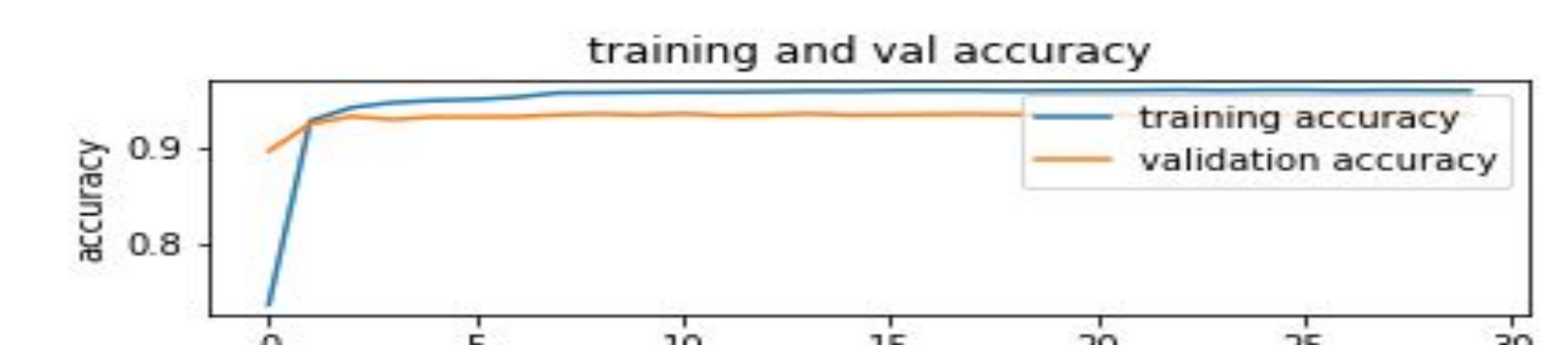
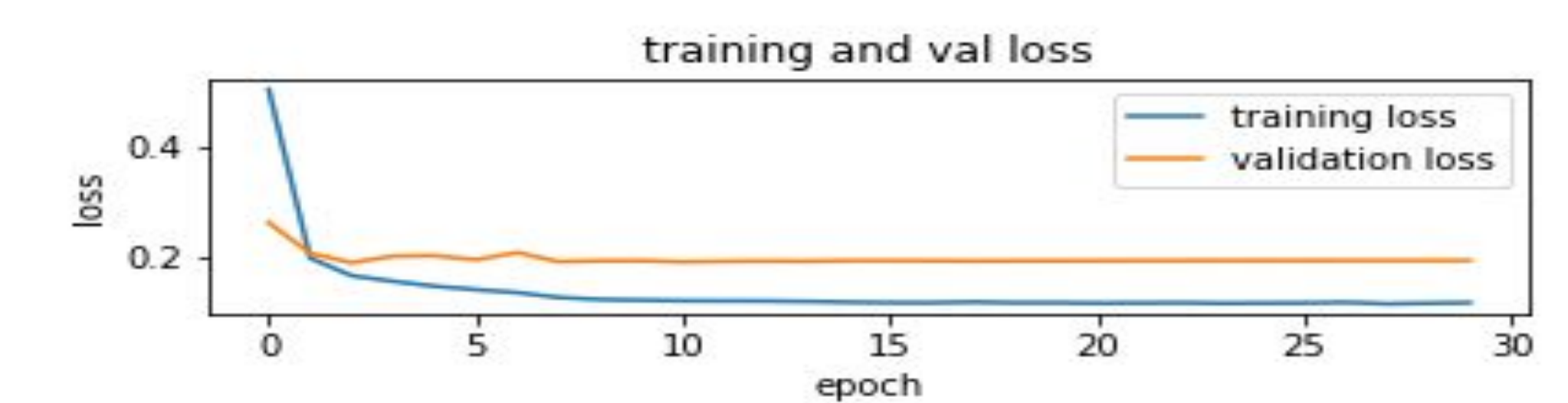
Experiment 1 : Adam optimizer.

We achieved Best validation acc of 0.939333 with learning rate 0.001 , step_size=3, gamma=0.1 for 10 epoch.

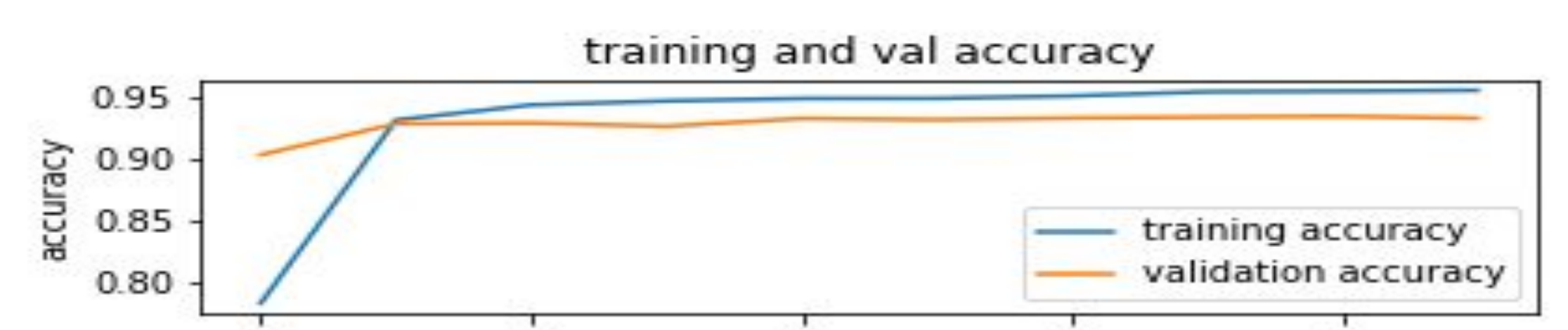
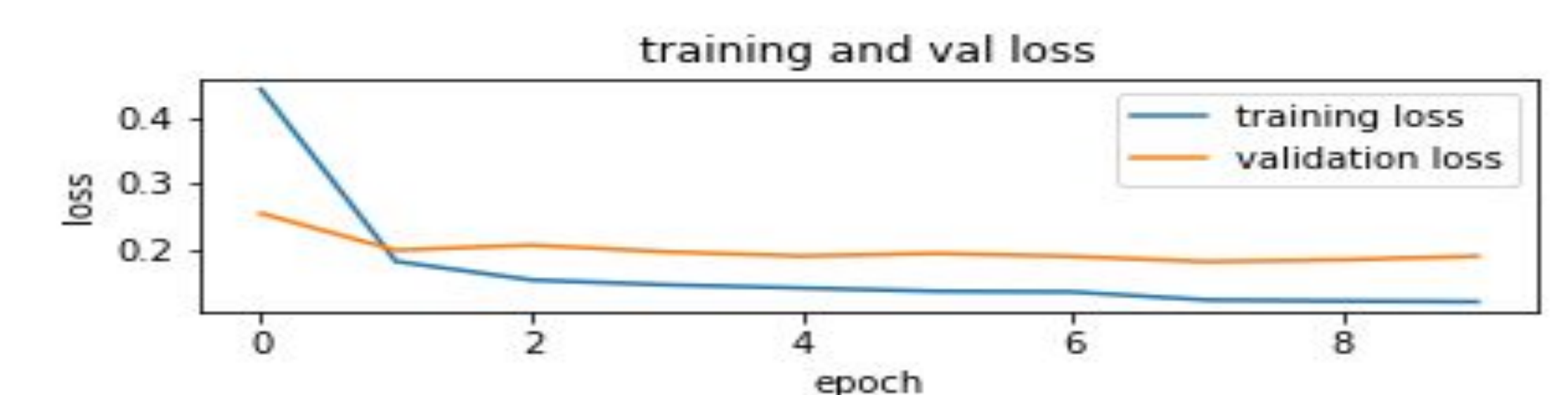


RESULTS AND EXPERIMENTS

Experiment 2(SGD with momentum): we trained our network for 30 epoch, it gave us best validation accuracy of 0.934333. Observing the curves below we noted that our model was not learning anything after 15 epoch as there was no change in the accuracy after that point accuracy was stagnant. Best learning rate selected was 0.001 with step_size=7, gamma=0.1 momentum = 0.9.



Experiment 3:(Early stopping) SGD with momentum Due to above observations from experiment 1 we decided to stop early at 10 epoch. This experiment gave us best validation accuracy of 0.935667. Learning rate selected was 0.001 with step_size=7, gamma=0.1 and momentum = 0.9.



CONCLUSIONS

- From the experiments conducted above we conclude that Adam optimizer did 0.4% better than the stostacic gradient descent with nesterov momentum with 10 epoch.
- Second we observe that this model is not a deep but yet it performs well on this dataset. We concluded that the problem of classification is very easy hence the model does well.