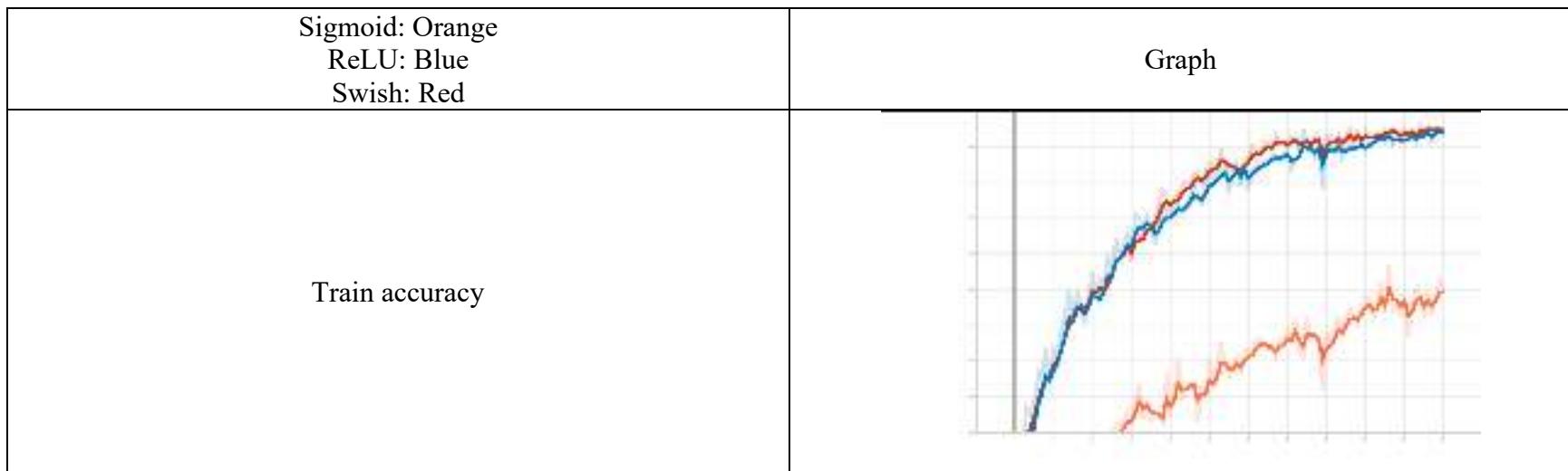


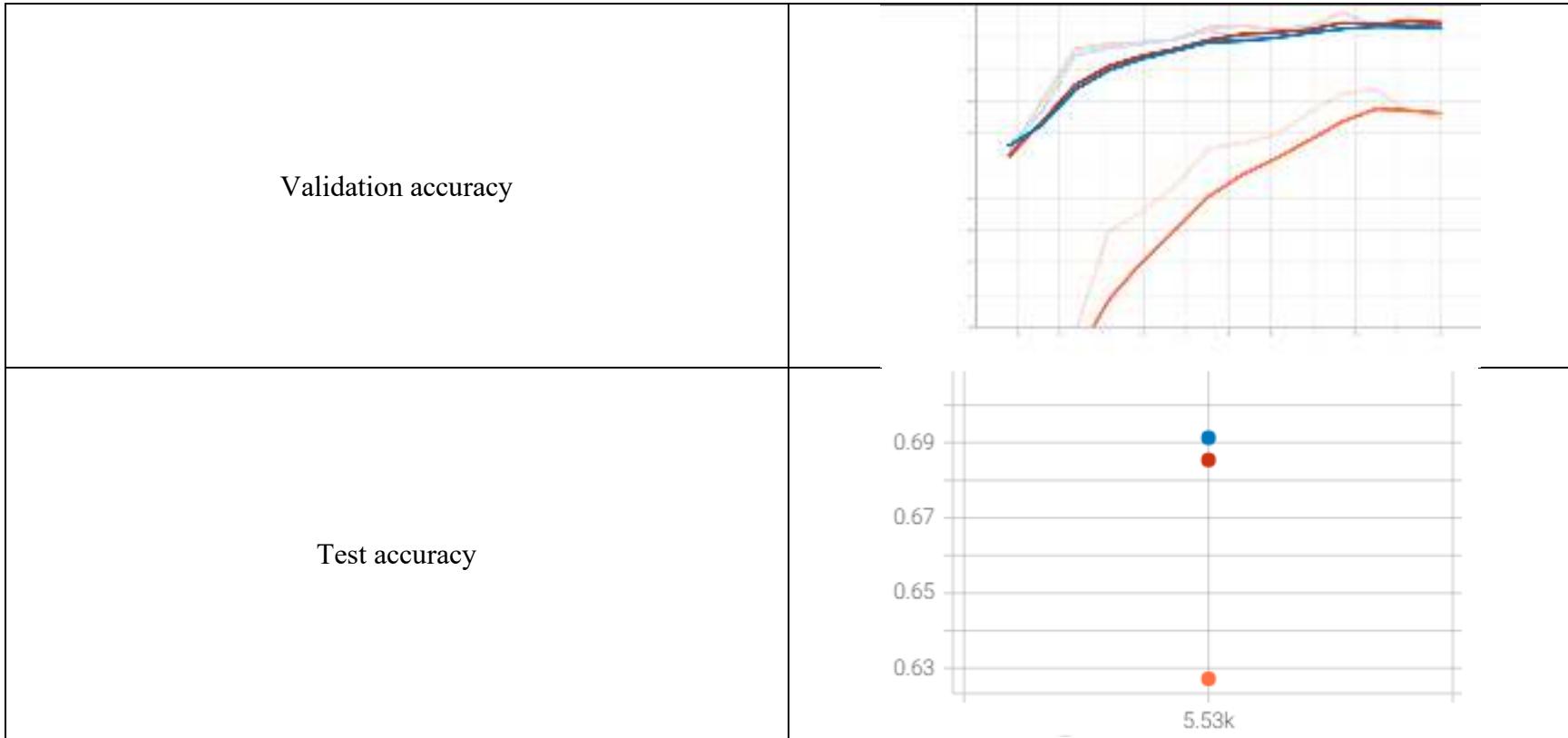
The followings will be my report about my model's performance on different hyperparameters:

Activation: ReLU - Swish - Sigmoid

Same (hyper)parameters with the global seed set to 42:

- Loss: CrossEntropyLoss
- DropOut: $p = 0.4$
- Optimizer: AdamW with learning rate 0.001 with weight decay 0.00001
- Epochs: 70





In all phases, Sigmoid underperforms while it is neck-to-neck between ReLU(96% train_acc and 69% test_acc) and Swish (96% train_acc and 68.8% test_acc)

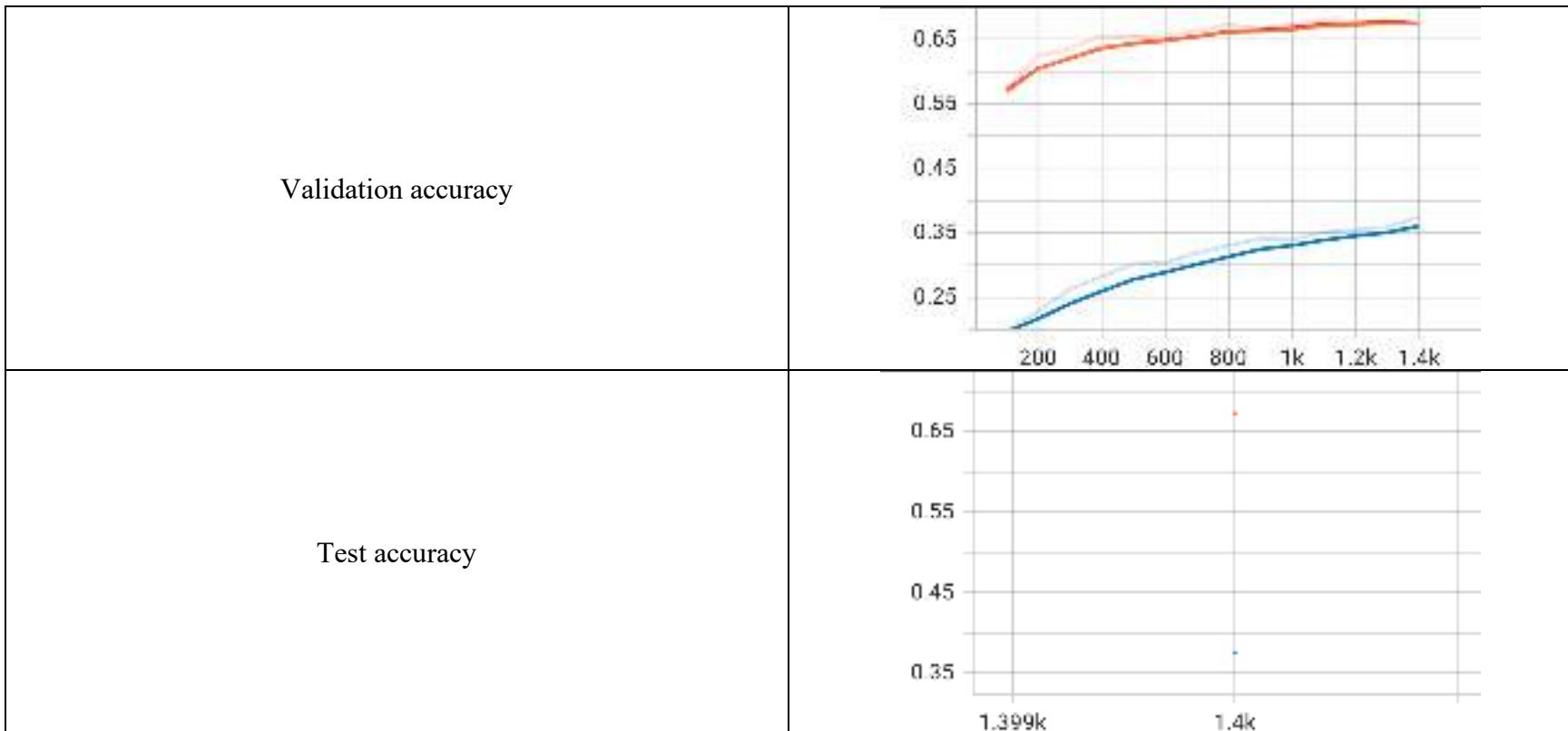
The graph shows that the model is probably under the effect of overfitting, but it is not the focus of this report.

Optimizer: Adam – SGD - AdamW

Same (hyper)parameters with the global seed set to 42:

- Loss: CrossEntropyLoss
- DropOut: $p = 0.4$
- Activation: Swish
- Epochs: 70





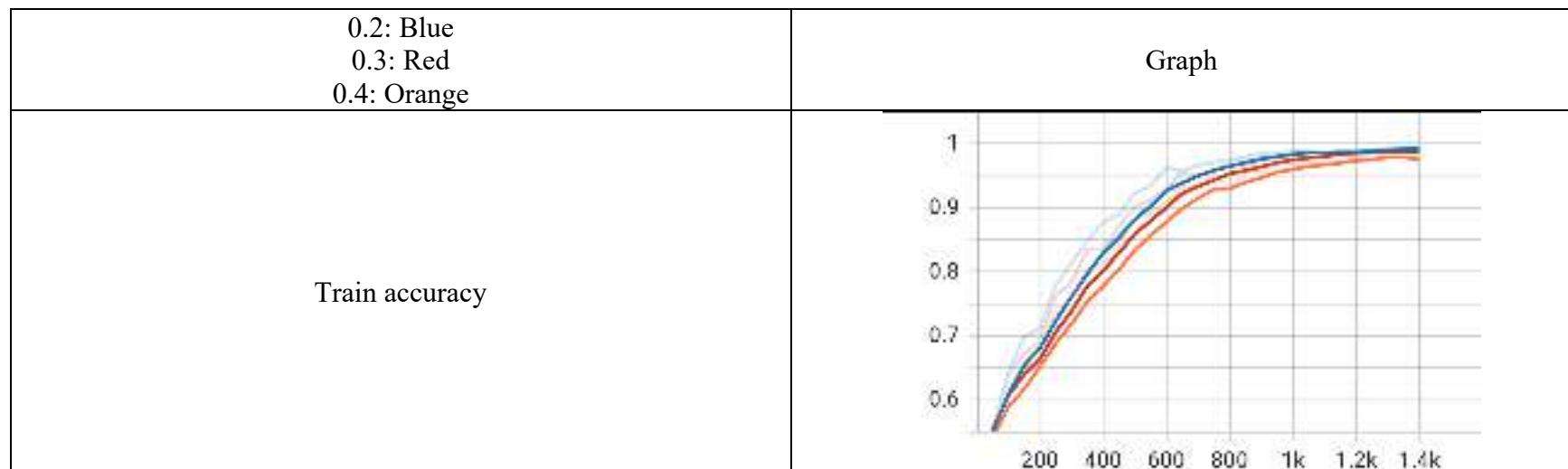
SGD underperforms in all phases while Adam and AdamW result in identical outcomes (they overlap) and achieve (98%, 67%, and 76% on training/validation/test accuracy).

The graph shows that the model is probably under the effect of overfitting, but it is not the focus of this report.

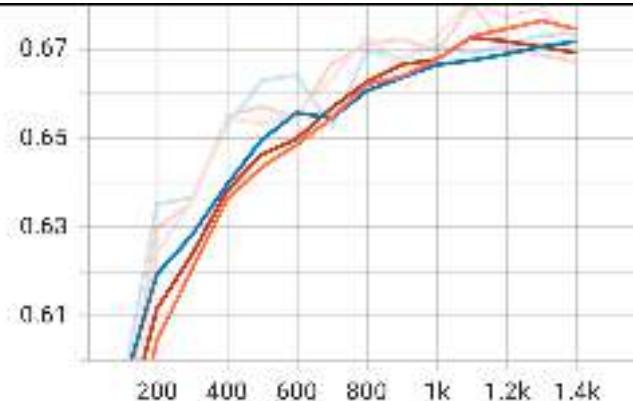
DropOut: $p = 0.2 - 0.3 - 0.4$

Same (hyper)parameters with the global seed set to 42:

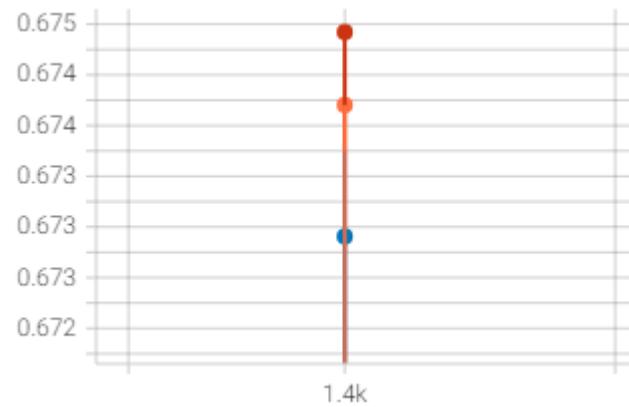
- Loss: CrossEntropyLoss
- Activation function: Swish
- Optimizer: AdamW with learning rate 0.001 and weight decay 0.00001
- Epochs: 70



Validation accuracy



Test accuracy



There is no significant difference among the three dropout ratios when the test accuracy in descending order is 0.3, 0.4, and 0.2

The graph shows that the model is probably under the effect of overfitting, but it is not the focus of this report.