

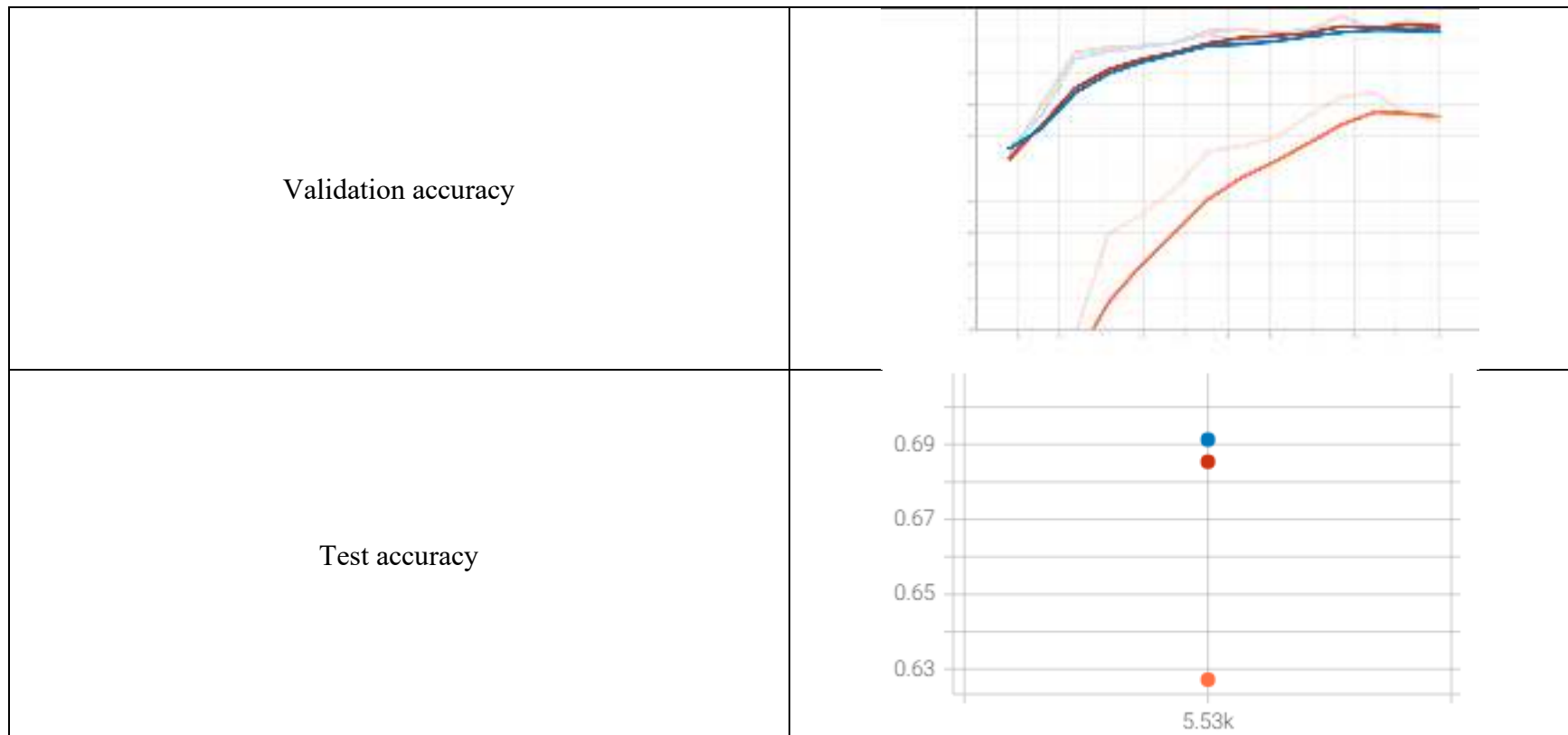
The followings will be my report about my model's performance on different hyperparameters:

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

Sigmoid: Orange ReLU: Blue Swish: Red	Graph
Train accuracy	



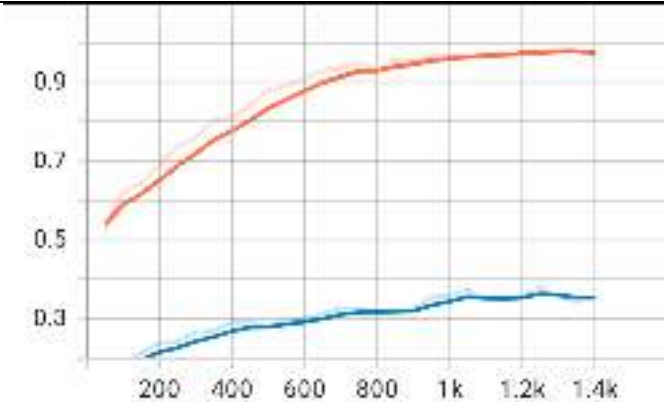
In all phases, Sigmoid underperforms while it is neck-to-neck between ReLU(96% train_acc and 69% test_acc) and Swish (96% train_acc and 68.8% test_acc)

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

Adam: Orange SGD: Blue AdamW: Red	Graph																																
Train accuracy	 <p>The graph displays the training accuracy of three optimizers over 1400 epochs. The y-axis represents accuracy from 0.3 to 0.9, and the x-axis represents epochs from 200 to 1400. Adam (orange) and AdamW (red) show rapid growth, reaching a plateau near 0.95. SGD (blue) shows much slower growth, plateauing around 0.35.</p> <table><tr><th>Epochs</th><th>Adam (Orange)</th><th>SGD (Blue)</th><th>AdamW (Red)</th></tr><tr><td>200</td><td>0.55</td><td>0.15</td><td>0.55</td></tr><tr><td>400</td><td>0.70</td><td>0.25</td><td>0.70</td></tr><tr><td>600</td><td>0.85</td><td>0.30</td><td>0.85</td></tr><tr><td>800</td><td>0.90</td><td>0.32</td><td>0.90</td></tr><tr><td>1000</td><td>0.92</td><td>0.35</td><td>0.92</td></tr><tr><td>1200</td><td>0.94</td><td>0.35</td><td>0.94</td></tr><tr><td>1400</td><td>0.95</td><td>0.35</td><td>0.95</td></tr></table>	Epochs	Adam (Orange)	SGD (Blue)	AdamW (Red)	200	0.55	0.15	0.55	400	0.70	0.25	0.70	600	0.85	0.30	0.85	800	0.90	0.32	0.90	1000	0.92	0.35	0.92	1200	0.94	0.35	0.94	1400	0.95	0.35	0.95
Epochs	Adam (Orange)	SGD (Blue)	AdamW (Red)																														
200	0.55	0.15	0.55																														
400	0.70	0.25	0.70																														
600	0.85	0.30	0.85																														
800	0.90	0.32	0.90																														
1000	0.92	0.35	0.92																														
1200	0.94	0.35	0.94																														
1400	0.95	0.35	0.95																														



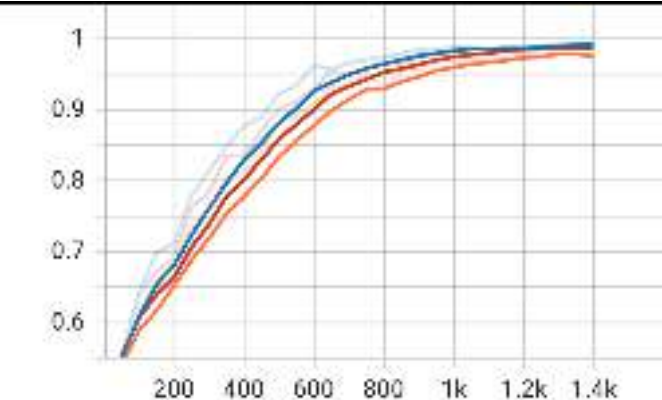
SGD underperforms in all phases while Adam and AdamW result in identical outcomes (they overlap) and achieve (98%, 67%, and 76% on training/validation/test accuracy).

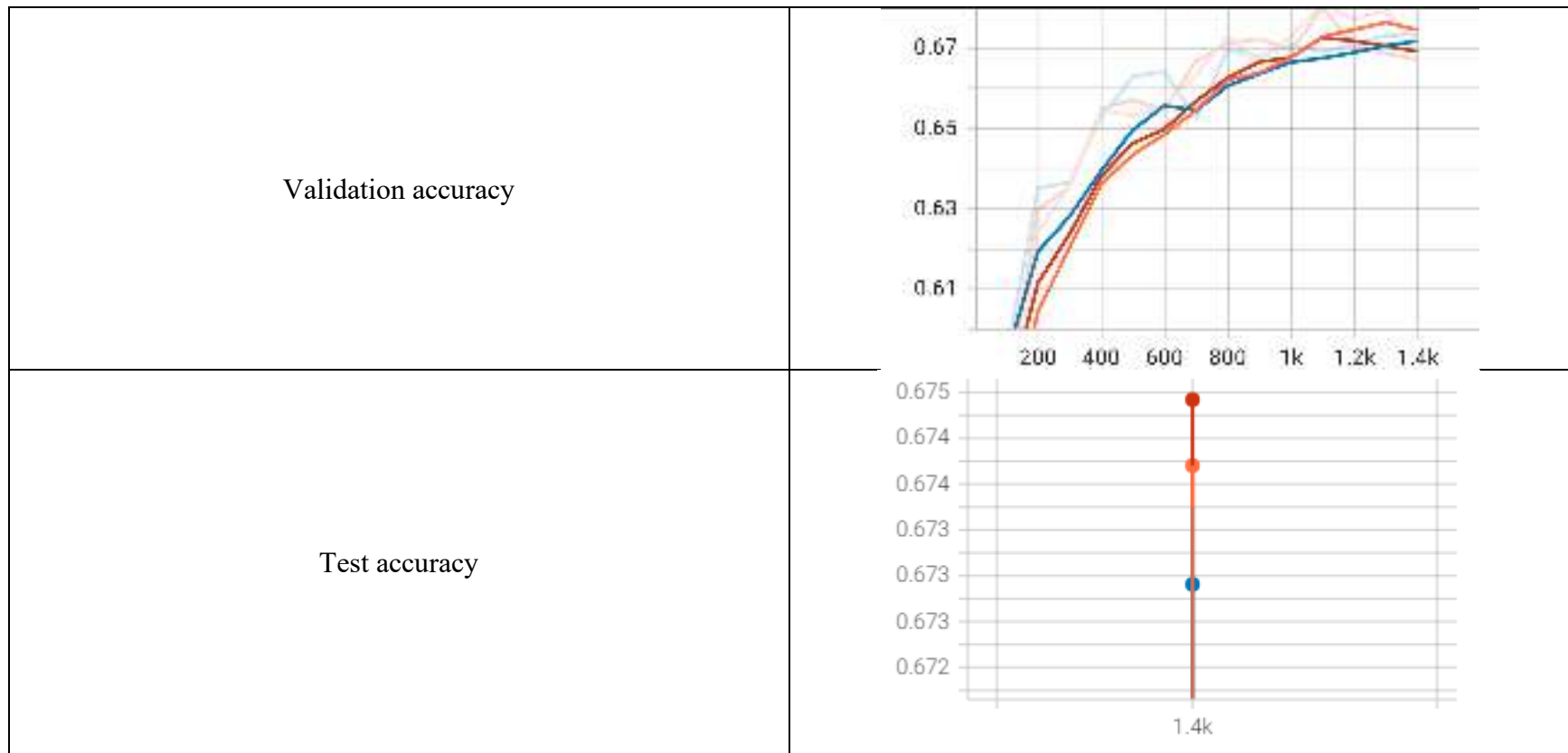
The graph shows that the model is probably under the effect of overfitting, but it is not the focus of this report.

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

0.2: Blue 0.3: Red 0.4: Orange	Graph
Train 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.