BASELINE MODEL

| Experiment #	Epochs	Learning Rate	Batch Size	Optimizer	Test Accuracy	Average Loss
1	20	0.1	64	SGD	76.6%	0.51
2	20	0.01	64	SGD	83.6%	0.33
3	20	0.001	64	SGD	83.6%	0.42
4	20	0.0001	64	SGD	80.5%	0.82
5	20	0.001	16	SGD	83.9%	0.47
6	20	0.001	32	SGD	83.9%	0.53
7	20	0.001	64	SGD	83.6%	0.46
8	20	0.001	128	SGD	83.2%	0.50
9	20	0.001	512	SGD	81.0%	0.92
10	20	0.001	1024	SGD	78.9%	1.34
11	1	0.001	64	SGD	78.1%	0.78
12	10	0.001	64	SGD	83.1%	0.53
13	20	0.001	64	SGD	83.6%	0.46
14	50	0.001	64	SGD	83.9%	0.47
15	100	0.001	64	SGD	84.0%	0.43
16	20	0.001	64	Adam	82.9%	0.34
17	20	0.001	64	SGD	83.6%	0.46
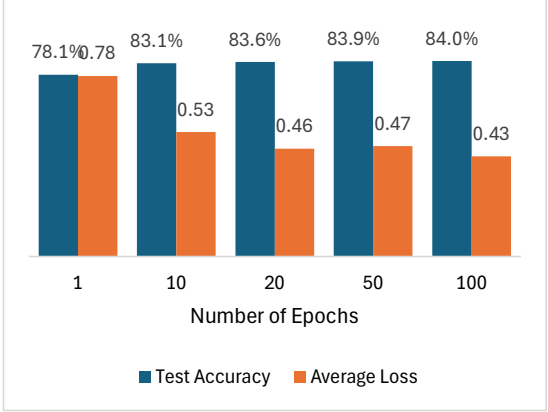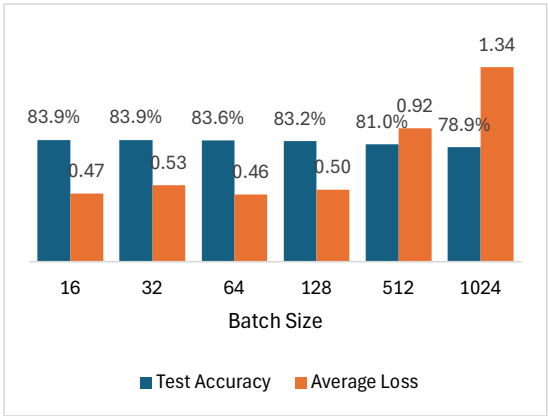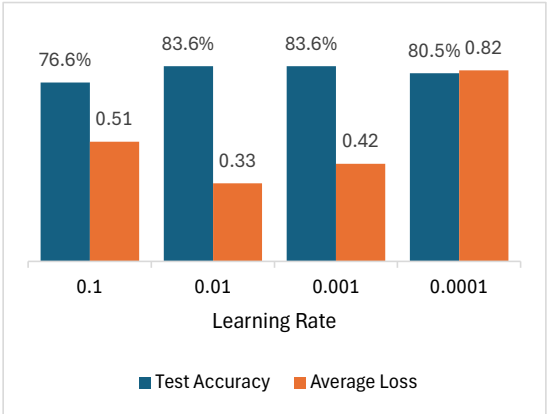
FINDINGS

* The learning rate significantly impacts model performance. Too high or too low rates hurt performance.

* Smaller batches are better. Batch sizes of 16-64 perform well (83.6-83.9% accuracy), while larger batches (512-1024) drop accuracy and increase loss significantly.

* More epochs help, but only up to a point. Accuracy improves from 78.1% at 1 epoch to 83.6% at 20 epochs. After that, the gains are minimized.

* Using Adam optimizer instead of SGD results in slightly lower accuracy (82.9%) but a lower average loss (0.34). This means that Adam may converge faster or more smoothly but does not necessarily mean it's better than SGD in this case.

IMPROVED MODEL

Experiment #	Epochs	Learning Rate	Batch Size	Optimizer	Test Accuracy	Average Loss
1	20	0.1	64	Adam	0.4%	6.04
2	20	0.01	64	Adam	72.5%	0.94
3	20	0.001	64	Adam	91.1%	0.11
4	20	0.0001	64	Adam	91.3%	0.31
5	20	0.001	16	Adam	89.9%	0.18
6	20	0.001	32	Adam	90.6%	0.14
7	20	0.001	64	Adam	91.3%	0.09
8	20	0.001	128	Adam	91.3%	0.07
9	20	0.001	512	Adam	90.8%	0.23
10	20	0.001	1024	Adam	91.2%	0.29
11	1	0.001	64	Adam	84.1%	0.35
12	10	0.001	64	Adam	90.5%	0.14
13	20	0.001	64	Adam	92.0%	0.10
14	50	0.001	64	Adam	91.8%	0.08
15	100	0.001	64	Adam	92.0%	0.06
16	20	0.001	64	SGD	91.6%	0.29
17	20	0.001	64	Adam	92.0%	0.10
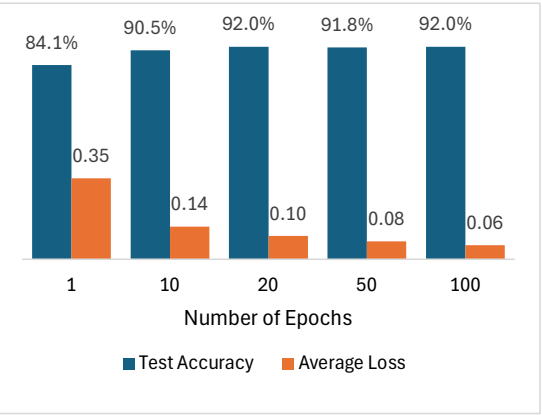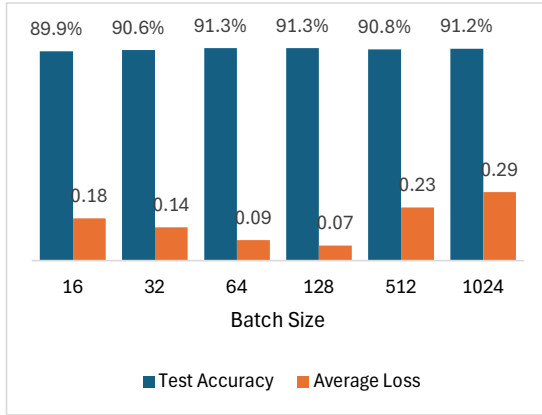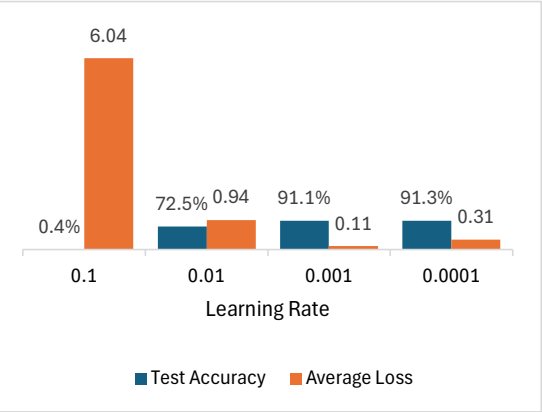
FINDINGS

* A learning rate of 0.001 works best, achieving 91.1-92.0% accuracy. Higher rates (e.g., 0.1) fail completely (0.4% accuracy), while lower rates (0.0001) still perform well but with slightly higher loss.

* Batch size has minimal effect on accuracy, with all sizes (16-1024) achieving 89.9-91.3% accuracy.

* Increasing epochs from 1 to 20 significantly improves accuracy (84.1% to 92.0%) but beyond 20 epochs, gains are minimal. Accuracy plateaus around 92.0%.

* Adam achieves higher accuracy (92.0%) and lower loss (0.10) compared to SGD (91.6% accuracy, 0.29 loss).

Learning Rate chart — Test Accuracy and Average Loss:
- 0.1: 0.4%, 6.04
- 0.01: 72.5%, 0.94
- 0.001: 91.1%, 0.11
- 0.0001: 91.3%, 0.31

Batch Size chart — Test Accuracy and Average Loss:
- 16: 89.9%, 0.18
- 32: 90.6%, 0.14
- 64: 91.3%, 0.09
- 128: 91.3%, 0.07
- 512: 90.8%, 0.23
- 1024: 91.2%, 0.29

Number of Epochs chart — Test Accuracy and Average Loss:
- 1: 84.1%, 0.35
- 10: 90.5%, 0.14
- 20: 92.0%, 0.10
- 50: 91.8%, 0.08
- 100: 92.0%, 0.06