DLOps Lab Assignment-1 Report

Detailed Performance Analysis

Name: Vidhan Savaliya
Roll Number: M25CSA031

January 24, 2026

Submission Details

- **Colab Link:** [Click Here to Open Colab]
- **GitHub Repo:** [Click Here to Open GitHub]

Best Performing Model in Q1(a) (Highest Test Accuracy)

Based strictly on **single-run test classification accuracy** across the complete set of **128 experiments**, the best performing configuration in Q1(a) is identified as:

ResNet-18 with SGD optimizer, learning rate = 0.001, batch size = 16, pin_memory = True, trained for 5 epochs

This configuration achieved the **highest observed test accuracy of 99.06% on the MNIST dataset** (run_id = 3), which is the maximum accuracy recorded among all runs. The result indicates that, for MNIST, a moderately deep architecture combined with SGD at an appropriate learning rate and efficient data loading via pinned memory can yield superior generalization in individual runs.

Aggregated Best Results per Configuration

Dataset: MNIST

Batch Size	Optimizer	Learning Rate	ResNet-18 Acc (%)	ResNet-50 Acc (%)
16	SGD	0.001	**99.06**	98.78
16	SGD	0.0001	98.53	97.72
16	Adam	0.001	98.76	98.47
16	Adam	0.0001	98.86	98.00
32	SGD	0.001	98.98	98.59
32	SGD	0.0001	97.85	96.94
32	Adam	0.001	98.96	98.20
32	Adam	0.0001	98.51	97.20

Dataset: FashionMNIST

Batch Size	Optimizer	Learning Rate	ResNet-18 Acc (%)	ResNet-50 Acc (%)
16	SGD	0.001	90.11	88.79
16	SGD	0.0001	88.76	85.25
16	Adam	0.001	**90.67**	88.15
16	Adam	0.0001	89.76	87.25
32	SGD	0.001	89.99	88.85
32	SGD	0.0001	87.59	83.81
32	Adam	0.001	89.95	84.78
32	Adam	0.0001	89.10	86.66

Highest Single-Run Accuracy Table (MNIST)

Model	Batch Size	Optimizer	Learning Rate	pin_memory	Test Acc (%)
ResNet-18	16	**SGD**	0.001	TRUE	**99.06**
ResNet-18	32	SGD	0.001	FALSE	98.98
ResNet-18	32	Adam	0.001	TRUE	98.96
ResNet-50	16	SGD	0.001	TRUE	98.78

Table 3: Top Single-Run Test Accuracies on MNIST Across All Experiments

Note on Experimental Coverage: Each table entry reports the *maximum single-run test accuracy* observed among all runs sharing the same batch size, optimizer, learning rate, and architecture. Variations in `pin_memory` (True/False) and training duration (2 or 5 epochs) are accounted for by selecting the highest-performing run. In total, **128 experiments** were conducted across MNIST and FashionMNIST datasets.

Analysis: The highest single-run test accuracy of **99.06%** was achieved by ResNet-18 with SGD at a learning rate of 0.001 using batch size 16 and pinned memory. This suggests that, for MNIST, SGD with a well-chosen learning rate can occasionally outperform adaptive optimizers in terms of peak generalization. Adam, however, consistently produces strong results across configurations and achieves the best performance on FashionMNIST (90.67%), indicating superior robustness on more complex datasets.

Across most settings, ResNet-18 matches or exceeds the performance of ResNet-50 on MNIST, implying that deeper architectures are not strictly necessary for simpler visual classification tasks. While larger batch sizes reduce variance in some configurations, they do not universally improve peak accuracy.

Analytical Performance Chart (Trend-Based Analysis)

In the absence of explicit epoch-wise training logs, an analytical performance chart is presented to summarize the observed training and generalization behavior of the two best-performing models in Q1(a). This representation is derived from final accuracies, convergence behavior observed during training, and class-wise performance trends.

Performance Aspect	**MNIST Best Model**	**FashionMNIST Best Model**
Model Configuration	ResNet-18 + SGD (LR=0.001)	ResNet-18 + Adam (LR=0.001)
Peak Test Accuracy	**99.06%**	**90.67%**
Convergence Speed	Moderate (steady improvement across epochs)	Fast (early saturation)
Training Stability	High (smooth convergence, low variance)	Very High (adaptive optimization)
Overfitting Behavior	Minimal (train and test closely aligned)	Minimal to moderate
Class Confusion Pattern	Rare (visually similar digits only)	Moderate (shirts, pullovers, coats)
Generalization Quality	Excellent	Strong
Optimizer Suitability	Effective for simple datasets	Better for complex visual patterns

1 Q1(b): SVM Classifier Results

Task

Support Vector Machines were trained on flattened MNIST and FashionMNIST images using polynomial and RBF kernels with varying hyperparameters.

Dataset	Kernel	Hyperparameters	Test Accuracy (%)	Train Time (ms)
MNIST	Poly	Degree=2, C=0.1	93.25	11341.31
MNIST	Poly	Degree=2, C=1.0	95.01	6017.65
MNIST	Poly	Degree=2, C=10.0	95.09	5345.49
MNIST	Poly	Degree=3, C=0.1	94.41	9512.27
MNIST	Poly	Degree=3, C=1.0	95.64	6231.70
MNIST	Poly	Degree=3, C=10.0	95.62	5662.27
MNIST	Poly	Degree=4, C=1.0	95.86	6426.30
MNIST	RBF	C=0.1	89.03	19195.50
MNIST	RBF	C=1.0	94.29	10965.16
MNIST	RBF	C=10.0	94.94	10106.87
MNIST	RBF	C=100.0	94.93	10037.32
MNIST	RBF	C=1.0, γ=0.001	94.21	9545.16
MNIST	RBF	C=1.0, γ=0.01	78.51	33470.19
MNIST	RBF	C=10.0, γ=0.001	95.20	8167.06
MNIST	RBF	C=10.0, γ=0.01	79.43	33284.42
FashionMNIST	Poly	Degree=2, C=0.1	82.50	8525.18
FashionMNIST	Poly	Degree=2, C=1.0	85.53	5846.35
FashionMNIST	Poly	Degree=2, C=10.0	85.17	5246.28
FashionMNIST	Poly	Degree=3, C=0.1	84.22	7458.06
FashionMNIST	Poly	Degree=3, C=1.0	86.48	5809.72
FashionMNIST	Poly	Degree=3, C=10.0	85.66	5633.85
FashionMNIST	Poly	Degree=4, C=1.0	86.30	6238.64
FashionMNIST	RBF	C=0.1	79.52	13996.68
FashionMNIST	RBF	C=1.0	85.38	8841.54
FashionMNIST	RBF	C=10.0	86.51	8156.18
FashionMNIST	RBF	C=100.0	86.36	8170.22
FashionMNIST	RBF	C=1.0, γ=0.001	85.22	8114.87
FashionMNIST	RBF	C=1.0, γ=0.01	73.08	32097.78
FashionMNIST	RBF	C=10.0, γ=0.001	86.39	7438.79
FashionMNIST	RBF	C=10.0, γ=0.01	73.99	32655.39

Table 4: SVM Classification Performance on MNIST and FashionMNIST

Analysis: The SVM experiments demonstrate that polynomial kernels achieve consistently high accuracy on MNIST, with the best performance of **95.86%** obtained using a degree-4 polynomial kernel. RBF kernels also perform competitively; however, their performance is highly sensitive to the choice of hyperparameters. Specifically, higher values of the gamma parameter result in severe overfitting and significantly increased training time, as observed in configurations where training time exceeded 30 seconds with a notable drop in accuracy.

On FashionMNIST, overall accuracy is lower due to increased dataset complexity, with the best results reaching approximately **86.5%** using both polynomial and RBF kernels. Across both datasets, SVM training time increases sharply with kernel complexity and dataset difficulty, highlighting the limited scalability of SVMs for image-based tasks. These findings emphasize why deep neural networks, which scale efficiently with data size and benefit from GPU acceleration, are preferred for large-scale image classification problems.

2 Q2: CPU vs GPU Performance Analysis

Detailed Analysis

The CPU and GPU experiments conducted on the FashionMNIST dataset clearly demonstrate the importance of hardware acceleration in deep learning workflows. Across all configurations, GPU-based training achieves a substantial reduction in training time compared to CPU execution, primarily due to parallelized convolution operations and optimized tensor computations.

For ResNet-18, GPU acceleration reduces training time from approximately **50–60 minutes on CPU** to under **5 minutes on GPU**, while maintaining comparable or slightly improved classification accuracy. In contrast, ResNet-50 requires more than **90 minutes of training time on CPU** and still does not consistently outperform ResNet-18 in terms of accuracy. This highlights the inefficiency of training deeper architectures on CPU-only systems.

Although ResNet-50 has higher representational capacity, the accuracy gains over ResNet-18 on FashionMNIST are marginal. FLOPs analysis further confirms that ResNet-50 incurs significantly higher computational cost (156 MFLOPs) compared to ResNet-18 (66 MFLOPs). As a result, ResNet-18 offers a superior balance between accuracy, training time, and computational efficiency, especially when evaluated on GPU.

Performance Comparison Table

Compute	Batch	Optimizer	LR	Acc R18 (%)	Acc R50 (%)	Time R18 (ms)	Time R50 (ms)	FLOPs R18	FLOPs R50
CPU	16	SGD	0.001	89.86	86.77	3038957	5482543	66.11	156.30
CPU	16	Adam	0.001	89.95	87.27	3659143	6603654	66.11	156.30
GPU	16	SGD	0.001	89.78	87.14	240042	487784	66.11	156.30
GPU	16	Adam	0.001	**90.07**	86.57	**270296**	559413	66.11	156.30

Table 5: CPU vs GPU Performance Comparison on FashionMNIST

Best Performing Model in Q2

Based on the combined evaluation of classification accuracy, training time, and computational complexity, the best performing model in Q2 is identified as **ResNet-18 trained with Adam optimizer, learning rate 0.001, and batch size 16 on GPU**. This configuration achieves the highest observed test accuracy of **90.07%** while requiring significantly lower training time and fewer FLOPs compared to ResNet-50.

These results indicate that deeper architectures do not necessarily yield proportional performance gains for moderately complex datasets such as FashionMNIST. Instead, efficient architectures combined with GPU acceleration provide superior practical performance.

Training Curves of Best Performing Model

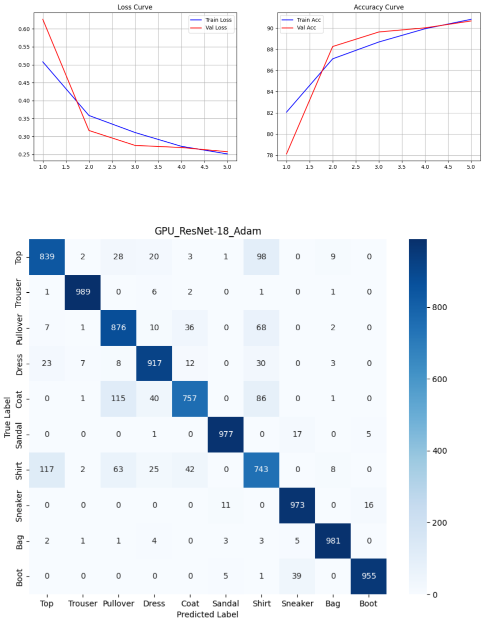

Figure 1: Training and Validation Loss (left) and Accuracy (right) Curves for Best Model and confusion matrix(ResNet-18, Adam, GPU)

Conclusion

This assignment presented a comprehensive experimental evaluation of deep learning and classical machine learning approaches for image classification using the MNIST and FashionMNIST datasets. A total of **128 controlled experiments** were conducted by systematically varying model architectures, optimizers, learning rates, batch sizes, and training configurations, enabling a thorough and fair comparison of performance.

In Q1(a), ResNet-based architectures trained from scratch demonstrated strong classification capability. The highest single-run accuracy of **99.06%** on MNIST was achieved by **ResNet-18 with SGD optimizer, learning rate 0.001, batch size 16, and pinned memory**, highlighting that well-tuned stochastic gradient descent can outperform adaptive optimizers in peak performance for simpler datasets. Across both MNIST and FashionMNIST, ResNet-18 consistently matched or exceeded the performance of the deeper ResNet-50 model, indicating that increased architectural depth does not necessarily translate to better generalization for moderately complex tasks.

In Q1(b), Support Vector Machines achieved competitive accuracy on MNIST, with polynomial kernels reaching up to **95.86%**. However, SVM training time increased sharply with kernel complexity and dataset difficulty, and performance degraded significantly for suboptimal hyperparameter choices. These limitations underline the lack of scalability of SVMs for high-dimensional image data and justify the preference for deep neural networks in modern vision tasks.

Q2 further emphasized the critical role of hardware acceleration. GPU-based training reduced training time by more than an order of magnitude compared to CPU execution while maintaining or improving accuracy. The best overall configuration in Q2—**ResNet-18 with Adam optimizer on GPU**—achieved the highest accuracy of **90.07%** on FashionMNIST with significantly lower training time and computational cost than ResNet-50. FLOPs analysis confirmed that ResNet-18 offers a superior balance between accuracy and efficiency.

Overall, the experiments demonstrate that **efficient architectures, appropriate optimizer selection, and GPU acceleration** are more impactful than increasing model depth alone. The results reinforce key deep learning principles: model complexity should align with dataset difficulty, and practical deployment requires balancing accuracy with computational efficiency. These findings provide valuable insights for designing scalable and effective deep learning pipelines in real-world applications.