Optimization Algorithm Performance Report

Tigist Wondimneh & Nahom Senay

1. Introduction

In this experiment, we systematically evaluated the effectiveness of several optimization algorithms on a **“Fashion MNISTclassification”** task using a neural network. The main goal was to compare their impact on training convergence, generalization (validation accuracy), and robustness using three primary metrics: **Loss**, **Accuracy**, and **F1 Score**.

The following optimizers were evaluated:

- Batch Gradient Descent
- Stochastic Gradient Descent
- Mini Batch Gradient Descent
- Gradient Descent with Momentum
- Gradient Descent with Nesterov Momentum
- AdaGrad
- RMSProp
- Adam

Dataset Description: Fashion MNIST

Fashion MNIST is a more challenging and modern alternative to the original MNIST dataset. It includes grayscale images of 10 categories of **Zalando fashion products**, providing a benchmark for image classification models.

- **Classes**:
  0 - T-shirt/top
  1 - Trouser
  2 - Pullover
  3 - Dress
  4 - Coat
  5 - Sandal
  6 - Shirt
  7 - Sneaker
  8 - Bag
  9 - Ankle boot
- **Image dimensions**: 28×28 pixels
- **Channels**: 1 (grayscale)
- **Train/Test Split**: 60,000 training images and 10,000 test images
- **Preprocessing**: Pixel values were normalized to the range [0, 1]

2. Evaluation Metrics

Each optimizer was assessed based on:

- **Loss**: The cross-entropy loss on the validation set.
- **Accuracy**: The percentage of correctly predicted labels.
- **F1 Score**: The harmonic mean of precision and recall, representing a balanced view of performance especially in imbalanced classification.

Optimizer	Loss	Accuracy (%)	F1 Score
Batch Gradient Descent	2.2848	28.27	0.1753
Mini batch GD	0.3219	87.87	0.8797
Gradient Descent with Momentum	0.2500	91.56	0.9160
Gradient Descent with Nesterov	0.2585	91.05	0.9101
AdaGrad	0.2463	91.15	0.9112
RMSProp	0.2803	91.72	0.9156
Adam	0.2895	91.81	0.9175

Due to high computational requirements the SGD was not able to run on our devices.

3. Visual Analysis

3.1 Part 1 Optimizer

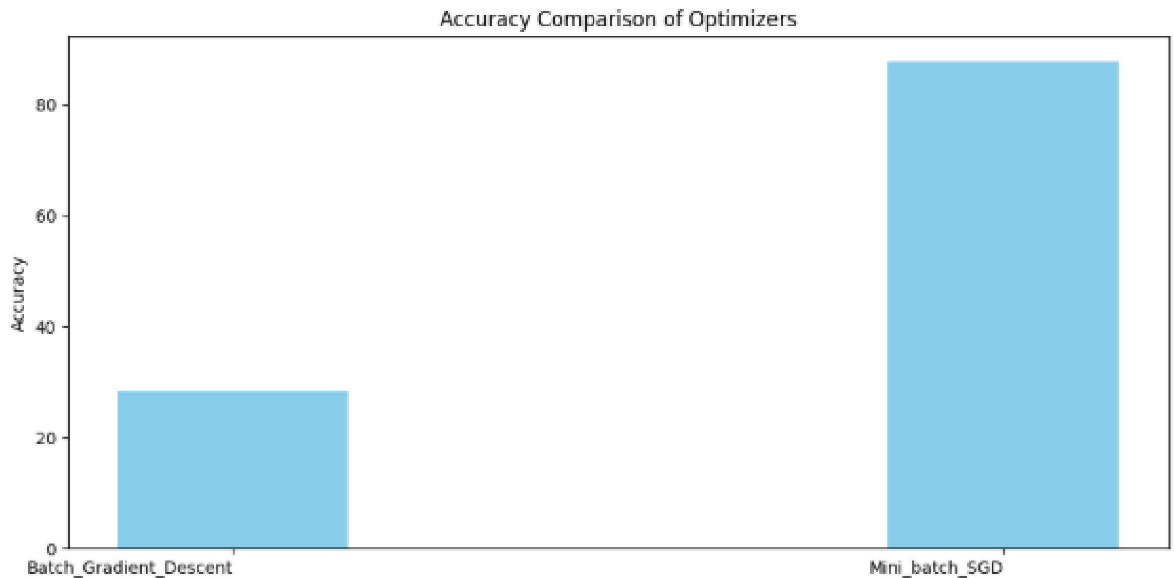

Fig: Accuracy comparison of optimizers

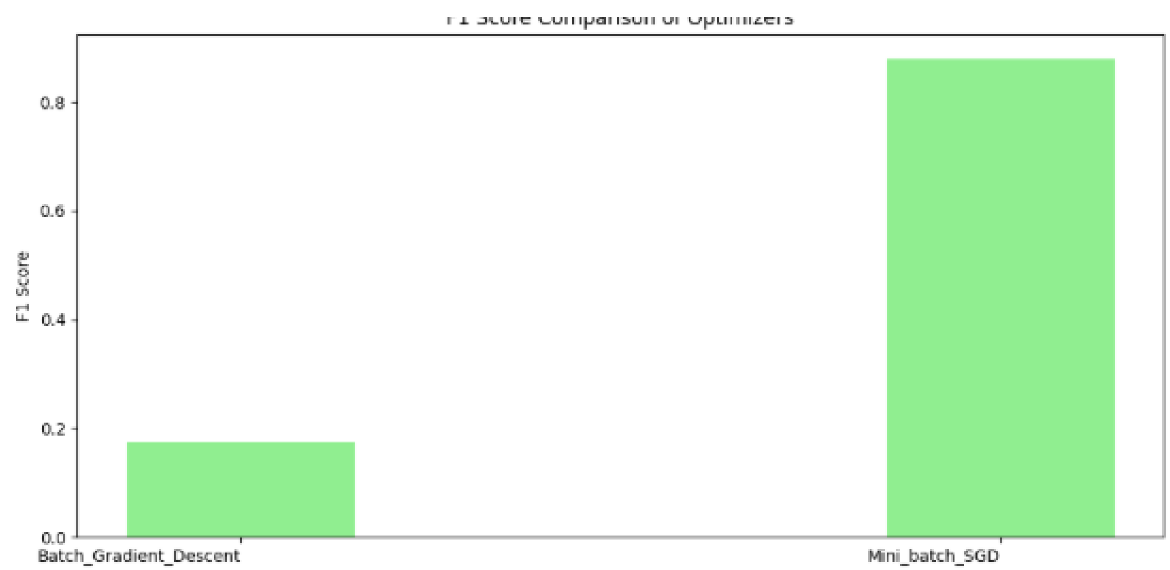

Fig: F1 score comparison of optimizers

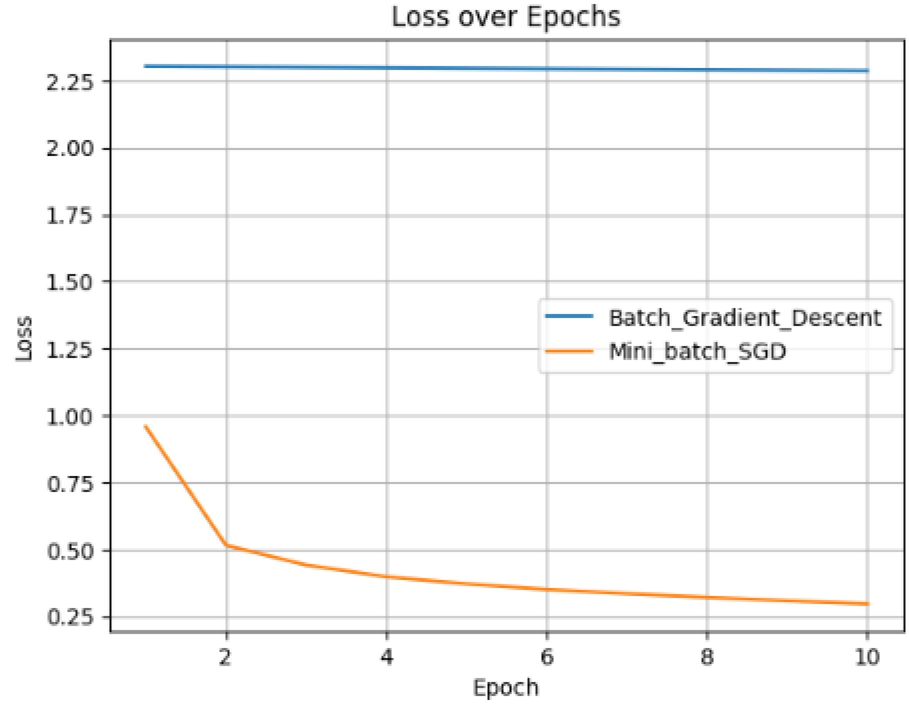

Fig: Loss over epochs

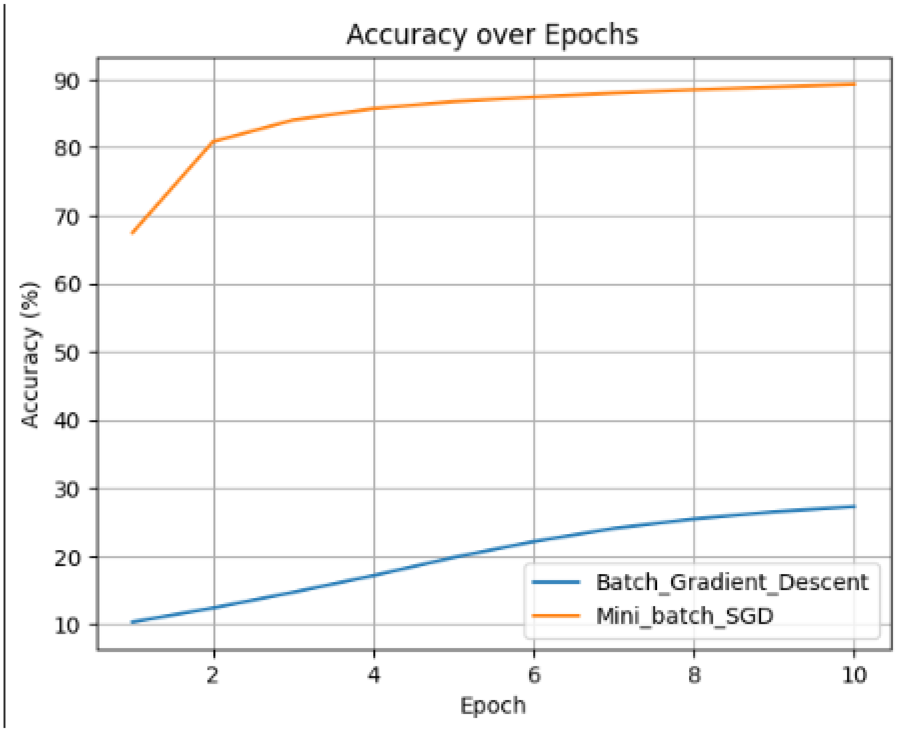

Fig: Accuracy over epoches

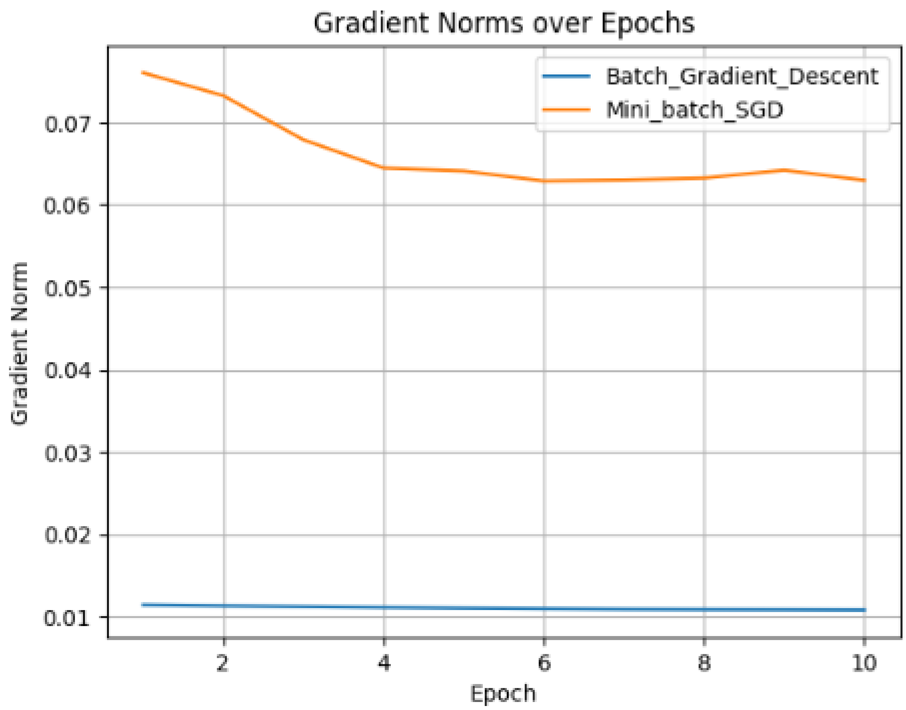

Fig: Gradient norms over epoches

3.2 Part 2 Optimizer

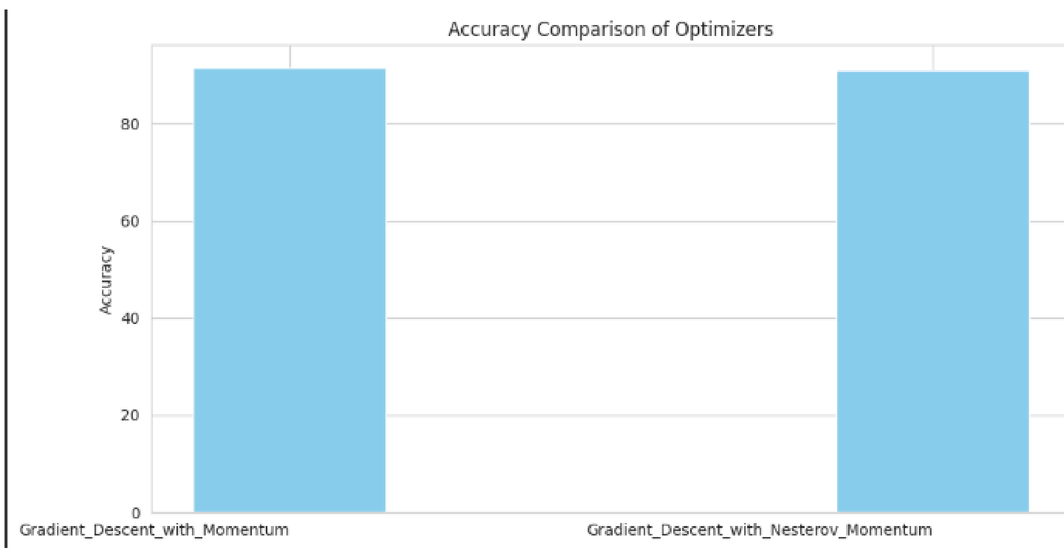

Fig: Accuracy comparison of GD with moment and D with Nesterov Moment

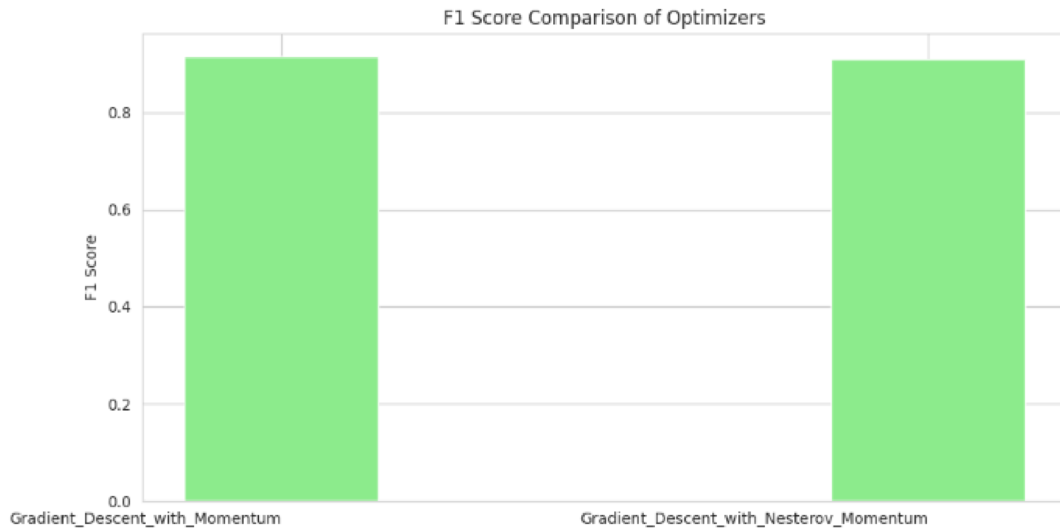

Fig: F1 score comparison of GD with moment and D with Nesterov Moment

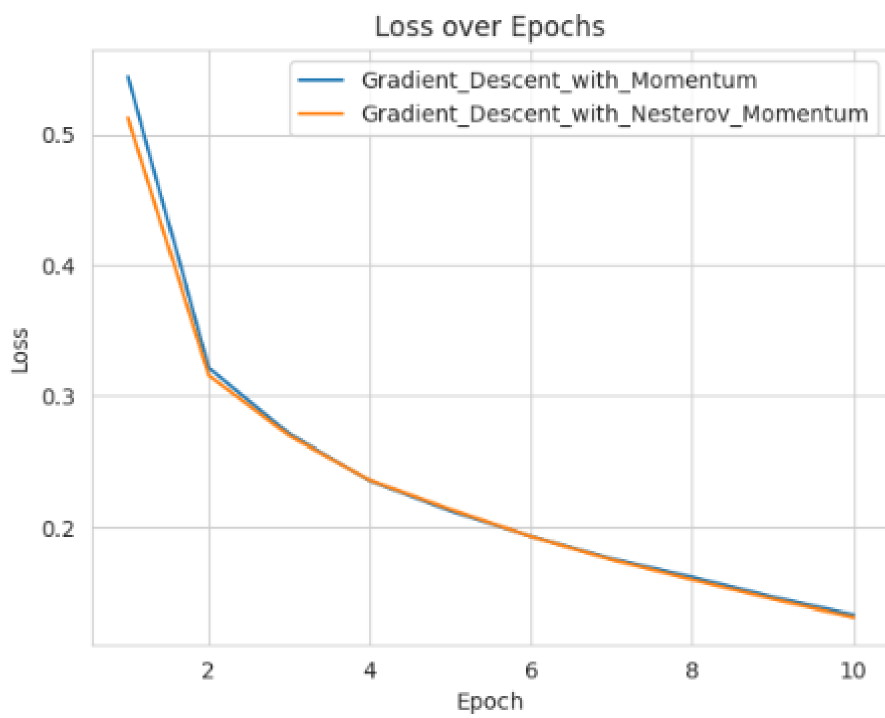

Fig: Loss over epochs

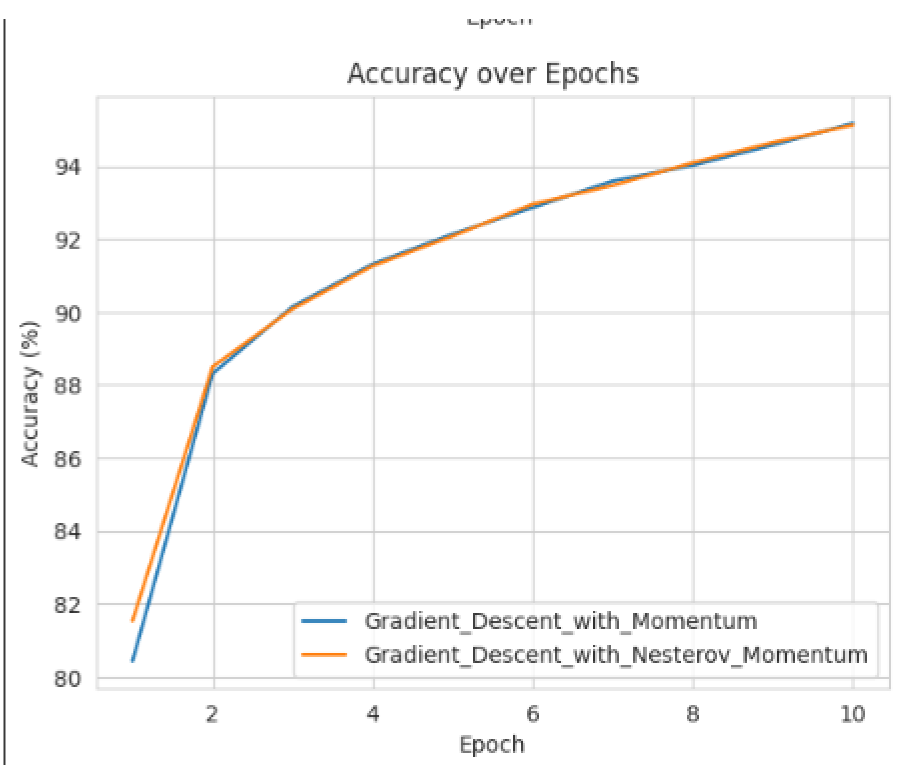

Fig: Accuracy for epoches

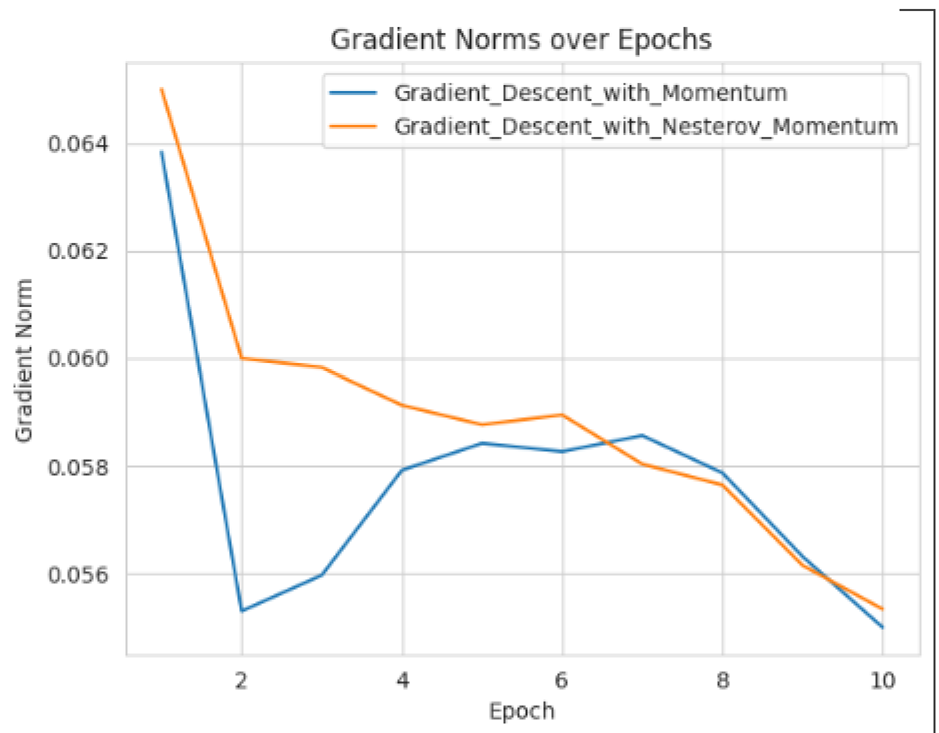

Fig: Gradient norms over epoches

3.3 Part 3 Optimizer

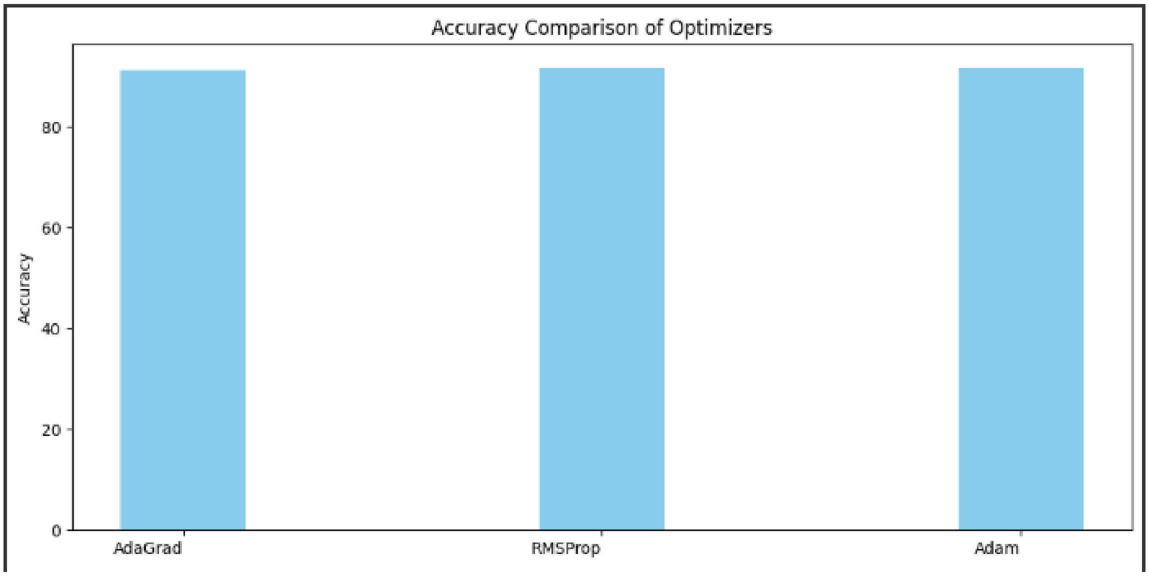

Fig: Accuracy comparison of Adagrad, RMSProp and Adam

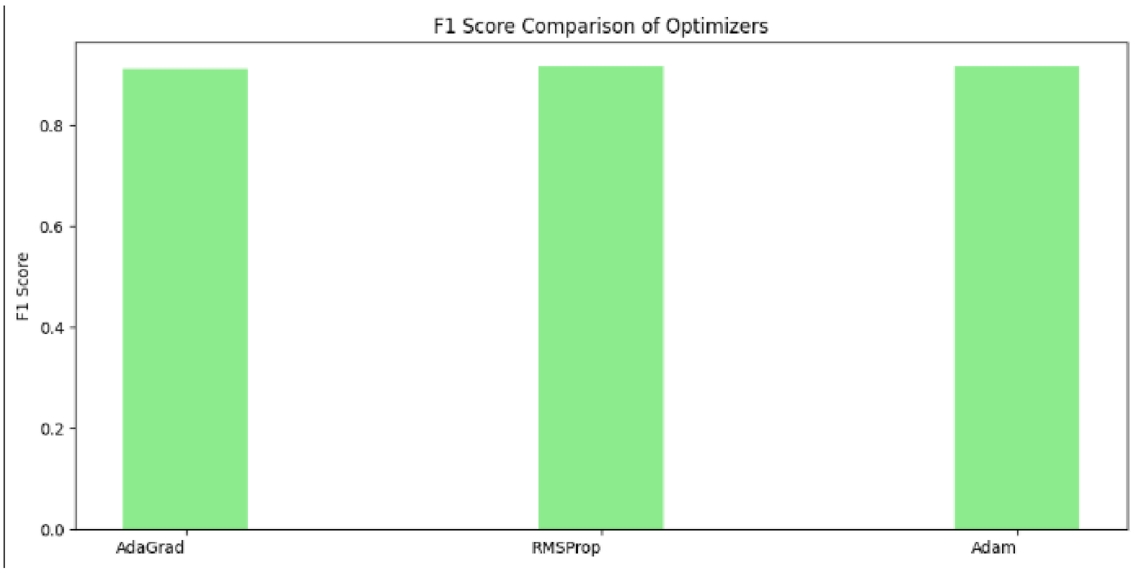

Fig: F1 score comparison of Adagrad, RMSProp and Adam

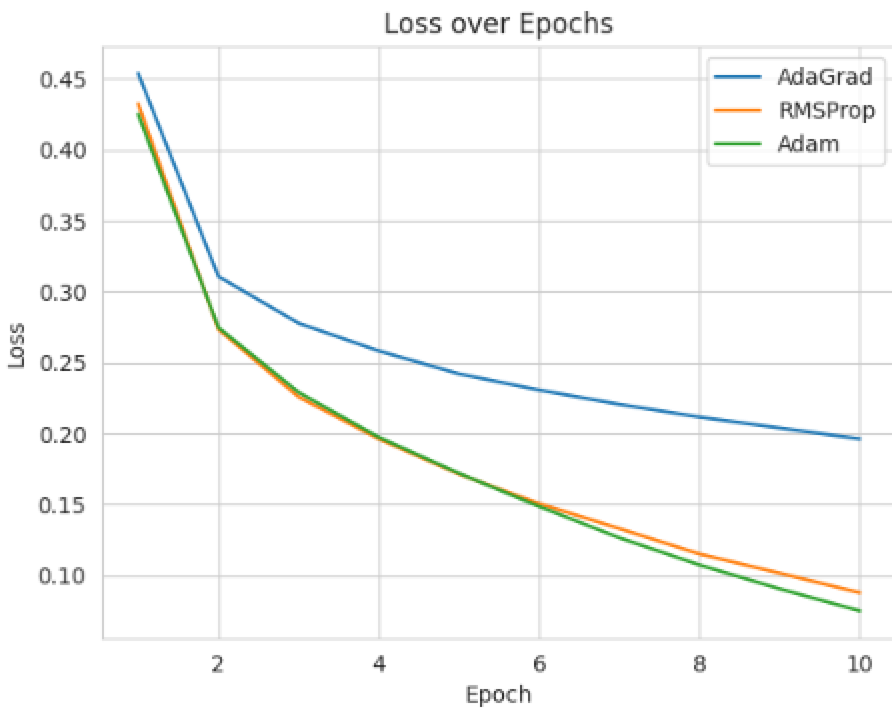

Fig: Loss over epoches

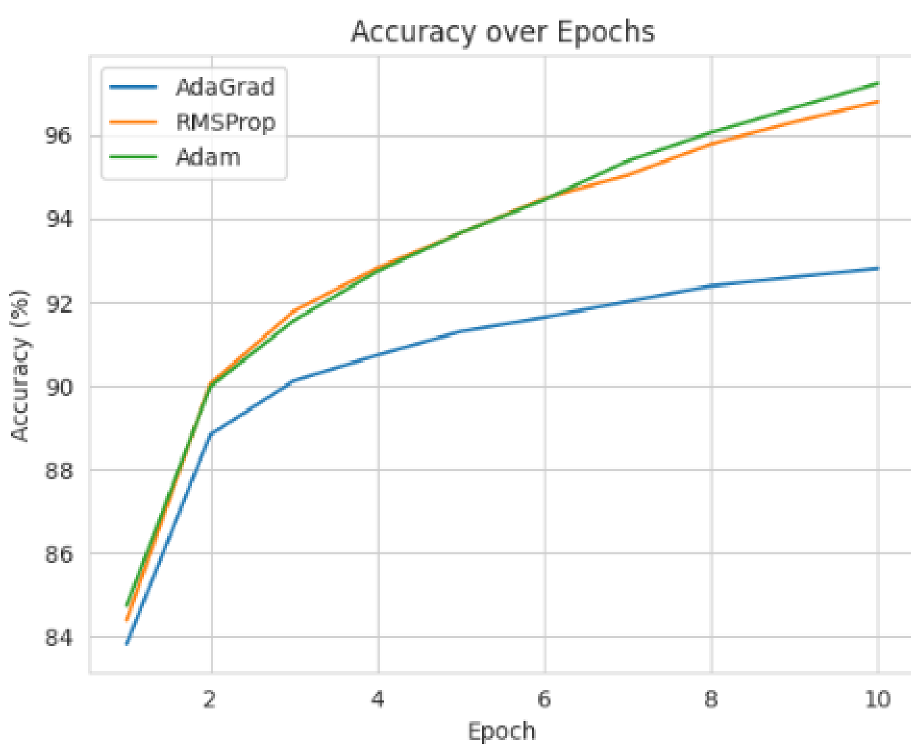

Fig: Accuracy over epochs

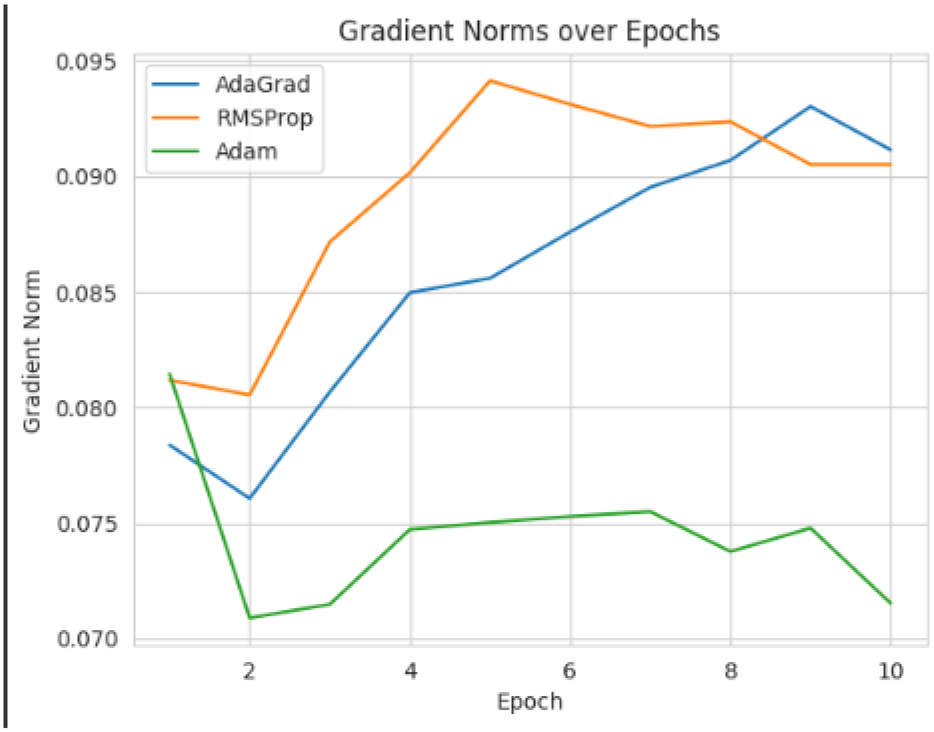

Fig: Gradient norms over epoches

4. Ranking and Insights

```
Basic Optimizer Ranking (based on F1 Score):
1. Mini_batch_SGD - Accuracy: 87.8700, F1 Score: 0.8798
2. Batch_Gradient_Descent - Accuracy: 28.2700, F1 Score: 0.1754
```

```
Basic Optimizer Ranking (based on F1 Score):
1. Gradient_Descent_with_Momentum - Accuracy: 91.5600, F1 Score: 0.9160
2. Gradient_Descent_with_Nesterov_Momentum - Accuracy: 91.0500, F1 Score: 0.9101
```

```
Basic Optimizer Ranking (based on F1 Score):
1. Adam - Accuracy: 91.8100, F1 Score: 0.9175
2. RMSProp - Accuracy: 91.7200, F1 Score: 0.9156
3. AdaGrad - Accuracy: 91.1500, F1 Score: 0.9112
```

```
Basic Optimizer Ranking (based on F1 Score):
1. Adam - Accuracy: 91.8100, F1 Score: 0.9175
2. Gradient_Descent_with_Momentum - Accuracy: 91.5600, F1 Score: 0.9160
3. RMSProp - Accuracy: 91.7200, F1 Score: 0.9156
4. AdaGrad - Accuracy: 91.1500, F1 Score: 0.9112
5. Gradient_Descent_with_Nesterov_Momentum - Accuracy: 91.0500, F1 Score: 0.9101
6. Mini_batch_SGD - Accuracy: 87.8700, F1 Score: 0.8798
7. Batch_Gradient_Descent - Accuracy: 28.2700, F1 Score: 0.1754
```

5. Conclusion

In this controlled experiment:

- **Adam** emerged as the most balanced and performant optimizer across all metrics with 91.8 % accuracy and F1 Score 0.9175.

- All optimizers performed within close range, reinforcing the importance of task-specific tuning over blind selection.