Defense Implemented: Input Transformation

Input Transformation Approach: Differential Privacy

Metric	Clean	No Defense Attack	w/ Defense Attack
Loss	0.00845	6.13	0.00301
Accuracy	100%	3%	100%

Example Misclassifications:

No misclassified images for stage: Clean
Attack: fgsm_pgd_attack
Dataset: MNIST
Training Epochs: 10
Trained Clean Images: 54210
Test Images: 140
Accuracy: 1.00
Precision: 1.00
Recall: 1.00
F1-score: 1.00
ROC AUC score is not defined for a single class.

Number of misclassified images for No Defense Attack: 62
Attack: fgsm_pgd_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 27105
Test Images: 140
Accuracy: 0.03
Precision: 0.06
Recall: 0.05
F1-score: 0.06
ROC AUC Score: 1.00

Misclassifications:
0 -> 2: 40
0 -> 6: 10
0 -> 5: 4
0 -> 8: 3
0 -> 7: 5

No misclassified images for stage: w/ Defense Attack
Attack: fgsm_pgd_attack

Dataset: MNIST
Training Epochs: 10
Retrained Clean and Adversarial Images: 81315
Test Images: 140
Accuracy: 1.00
Precision: 1.00
Recall: 1.00
F1-score: 1.00
ROC AUC score is not defined for a single class.
