Defense Implemented: Input Transformation

Input Transformation Approach: Image Quilting

Metric	Clean	No Defense Attack	w/ Defense Attack
Loss	0.0573	4.69	0.00526
Accuracy	98%	14%	100%

Example Misclassifications:

Number of misclassified images for Clean: 1
Attack: fgsm_bim_attack
Dataset: MNIST
Training Epochs: 10
Trained Clean Images: 54210
Test Images: 140
Accuracy: 0.98
Precision: 0.99
Recall: 0.99
F1-score: 0.99
ROC AUC Score: 1.00

Misclassifications:
0 -> 2: 1

Number of misclassified images for No Defense Attack: 55
Attack: fgsm_bim_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 27105
Test Images: 140
Accuracy: 0.14
Precision: 0.26
Recall: 0.22
F1-score: 0.24
ROC AUC Score: 1.00

Misclassifications:
0 -> 9: 11
0 -> 2: 19
0 -> 6: 10
0 -> 7: 5
0 -> 3: 1
0 -> 8: 3
0 -> 5: 6

No misclassified images for stage: w/ Defense Attack
Attack: fgsm_bim_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.
