Defense Implemented: Input Transformation

Input Transformation Approach: Image Quilting

Metric	Clean	No Defense Attack	w/ Defense Attack
Loss	0.00247	3.61	0.00267
Accuracy	100%	22%	100%

Example Misclassifications:

No misclassified images for stage: Clean
Attack: fgsm_cw_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: 50
Attack: fgsm_cw_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 27105
Test Images: 140
Accuracy: 0.22
Precision: 0.38
Recall: 0.33
F1-score: 0.35
ROC AUC Score: 1.00

Misclassifications:
0 -> 2: 11
0 -> 7: 4
0 -> 5: 8
0 -> 1: 3
0 -> 6: 10
0 -> 3: 3
0 -> 4: 6
0 -> 9: 4
0 -> 8: 1

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