Defense Implemented: Adversarial Training

Metric	Clean	No Defense Attack	w/ Defense Attack
Loss	0.0267	0.118	0.00246
Accuracy	100%	92%	100%

Example Misclassifications:

No misclassified images for stage: Clean
Attack: cw_bim_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: 5
Attack: cw_bim_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 2086
Test Images: 140
Accuracy: 0.92
Precision: 0.96
Recall: 0.96
F1-score: 0.96
ROC AUC Score: 1.00

Misclassifications:
0 -> 7: 1
0 -> 2: 1
0 -> 5: 2
0 -> 6: 1

No misclassified images for stage: w/ Defense Attack
Attack: cw_bim_attack
Dataset: MNIST
Training Epochs: 10
Retrained Clean and Adversarial Images: 56296

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.
