Defense Implemented: Adversarial Training

Metric	Clean	No Defense Attack	w/ Defense Attack
Loss	0.0233	0.234	0.0015
Accuracy	98%	89%	100%

Example Misclassifications:

Number of misclassified images for Clean: 1
Attack: pgd_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 -> 6: 1

Number of misclassified images for No Defense Attack: 7
Attack: pgd_bim_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 9213
Test Images: 140
Accuracy: 0.89
Precision: 0.95
Recall: 0.93
F1-score: 0.94
ROC AUC Score: 1.00

Misclassifications:
0 -> 5: 1
0 -> 9: 1
0 -> 6: 3
0 -> 2: 1
0 -> 7: 1

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

Dataset: MNIST
Training Epochs: 10
Retrained Clean and Adversarial Images: 63423
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.
