

Defense Implemented: Randomization

Randomization Approach: Random Cropping

Metric	Clean	No Defense Attack	w/ Defense Attack
Loss	0.0144	0.238	0.000854
Accuracy	100%	91%	100%

Example Misclassifications:

No misclassified images for stage: Clean
Attack: pgd_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: 6
Attack: pgd_bim_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 11506
Test Images: 140
Accuracy: 0.91
Precision: 0.95
Recall: 0.94
F1-score: 0.95
ROC AUC Score: 1.00

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

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

Training Epochs: 10

Retrained Clean and Adversarial Images: 65716

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.
