

Defense Implemented: Randomization

Randomization Approach: Combined Randomization

Metric	Clean	No Defense Attack	w/ Defense Attack
Loss	0.000589	0.0482	3.72e-05
Accuracy	100%	97%	100%

Example Misclassifications:

No misclassified images for stage: Clean
Attack: cw_pgd_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: 2
Attack: cw_pgd_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 11120
Test Images: 140
Accuracy: 0.97
Precision: 0.99
Recall: 0.98
F1-score: 0.98
ROC AUC Score: 1.00

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

No misclassified images for stage: w/ Defense Attack
Attack: cw_pgd_attack
Dataset: MNIST
Training Epochs: 10
Retrained Clean and Adversarial Images: 65330

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.
