Defense Implemented: Randomization

Randomization Approach: Combined Randomization

| Metric | Clean | No Defense Attack | w/ Defense Attack |
|---|---|---|---|
| Loss | 0.0037 | 3.99 | 0.0567 |
| Accuracy | 100% | 16% | 98% |

Example Misclassifications:

---
No misclassified images for stage: Clean
Attack: fgsm_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: 54
Attack: fgsm_bim_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 27105
Test Images: 140
Accuracy: 0.16
Precision: 0.28
Recall: 0.24
F1-score: 0.26
ROC AUC Score: 1.00
---

Misclassifications:
0 -> 9: 14
0 -> 8: 6
0 -> 5: 11
0 -> 2: 10
0 -> 3: 8
0 -> 6: 3
0 -> 1: 1
0 -> 4: 1

---
Number of misclassified images for w/ Defense Attack: 1
Attack: fgsm_bim_attack
Dataset: MNIST
Training Epochs: 10
Retrained Clean and Adversarial Images: 81315
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