

Defense Implemented: Input Transformation

Input Transformation Approach: Image Quilting

Metric	Clean	No Defense Attack	w/ Defense Attack
Loss	0.0021	3.91	0.00411
Accuracy	100%	23%	100%

Example Misclassifications:

No misclassified images for stage: Clean  
Attack: fgsm\_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: 49  
Attack: fgsm\_pgd\_attack  
Dataset: MNIST  
Training Epochs: 10  
Adversarial Training Images: 27105  
Test Images: 140  
Accuracy: 0.23  
Precision: 0.40  
Recall: 0.35  
F1-score: 0.37  
ROC AUC Score: 1.00

Misclassifications:

- 0 -> 8: 14
- 0 -> 6: 8
- 0 -> 5: 7
- 0 -> 2: 8
- 0 -> 4: 4
- 0 -> 9: 6
- 0 -> 7: 1
- 0 -> 3: 1

-----  
No misclassified images for stage: w/ Defense Attack  
Attack: fgsm\_pgd\_attack  
Dataset: MNIST  
Training Epochs: 10  
Retrained Clean and Adversarial Images: 81315  
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.  
-----