```
Defense Implemented: Input Transformation

Input Transformation Approach: Adversarial Logit Pairing

+----------+---------+--------------------+---------------------+
| Metric   | Clean   | No Defense Attack   | w/ Defense Attack   |
+==========+=========+=====================+=====================+
| Loss     | 0.00639 | 4.23                | 0.128               |
+----------+---------+--------------------+---------------------+
| Accuracy | 100%    | 5%                  | 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: 61
Attack: fgsm_bim_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 27105
Test Images: 140
Accuracy: 0.05
Precision: 0.09
Recall: 0.08
F1-score: 0.08
ROC AUC Score: 1.00
------------------------------------------------------

Misclassifications:
0 -> 9: 15
0 -> 5: 5
0 -> 6: 8
0 -> 8: 5
0 -> 7: 6
0 -> 2: 19
0 -> 4: 1
0 -> 1: 2
```

```
--------------------------------------------------------
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 -> 7: 1
```