```
Defense Implemented: Input Transformation

Input Transformation Approach: Adversarial Logit Pairing

+----------+---------+--------------------+---------------------+
| Metric   | Clean   | No Defense Attack  | w/ Defense Attack   |
+==========+=========+====================+=====================+
| Loss     | 0.105   | 0.372              | 0.00282             |
+----------+---------+--------------------+---------------------+
| Accuracy | 97%     | 84%                | 100%                |
+----------+---------+--------------------+---------------------+

Example Misclassifications:

-------------------------------------------------------
Number of misclassified images for Clean: 2
Attack: cw_pgd_attack
Dataset: MNIST
Training Epochs: 10
Trained Clean Images: 54210
Test Images: 140
Accuracy: 0.97
Precision: 0.99
Recall: 0.98
F1-score: 0.98
ROC AUC Score: 1.00
-------------------------------------------------------

Misclassifications:
0 -> 9: 1
0 -> 7: 1

-------------------------------------------------------
Number of misclassified images for No Defense Attack: 10
Attack: cw_pgd_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 10609
Test Images: 140
Accuracy: 0.84
Precision: 0.92
Recall: 0.91
F1-score: 0.91
ROC AUC Score: 1.00
-------------------------------------------------------

Misclassifications:
0 -> 9: 5
0 -> 7: 1
0 -> 3: 1
0 -> 5: 2
0 -> 4: 1
```

```
--------------------------------------------------------
No misclassified images for stage: w/ Defense Attack
Attack: cw_pgd_attack
Dataset: MNIST
Training Epochs: 10
Retrained Clean and Adversarial Images: 64819
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.
--------------------------------------------------------
```