```
Defense Implemented: Input Transformation

Input Transformation Approach: Differential Privacy

+----------+---------+--------------------+---------------------+
| Metric   | Clean   | No Defense Attack  | w/ Defense Attack   |
+==========+=========+====================+=====================+
| Loss     | 0.0058  | 4.72               | 0.0391              |
+----------+---------+--------------------+---------------------+
| Accuracy | 100%    | 11%                | 98%                 |
+----------+---------+--------------------+---------------------+

Example Misclassifications:

------------------------------------------------------
No misclassified images for stage: Clean
Attack: fgsm_cw_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: 57
Attack: fgsm_cw_attack
Dataset: MNIST
Training Epochs: 10
Adversarial Training Images: 27105
Test Images: 140
Accuracy: 0.11
Precision: 0.21
Recall: 0.18
F1-score: 0.19
ROC AUC Score: 1.00
------------------------------------------------------

Misclassifications:
0 -> 7: 3
0 -> 2: 27
0 -> 9: 7
0 -> 6: 2
0 -> 5: 9
0 -> 3: 1
0 -> 4: 7
0 -> 8: 1
```

```
---------------------------------------------------------
Number of misclassified images for w/ Defense Attack: 1
Attack: fgsm_cw_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
```