Robustness Test Results for Neural Network

1. Overview

This report summarizes the robustness test results of a neural network model. The testing was conducted using adversarial attacks, noise injection, and distributional shifts. Key metrics evaluated include robustness score, error rate under perturbations, and performance degradation.

2. Key Metrics

Metric	Value	Remarks
Robustness Score	87.4%	Stable under minor perturbations
Error Rate (Noise)	5.8%	Minimal performance loss
Error Rate (Adversarial)	12.3%	Moderate vulnerability
Performance Degradation	9.1%	Acceptable under distribution shifts

3. Test Data Summary

The robustness test dataset consists of original samples modified with various perturbations, including Gaussian noise, adversarial attacks, and random occlusions. The test cases were designed to evaluate the network's resilience under real-world conditions.

4. Performance Under Different Perturbations

Perturbation Type	Accuracy	Notes
Gaussian Noise	91.2%	Minor degradation observed
Adversarial Attack (FGSM)	87.5%	Moderate sensitivity
Occlusion (Random Patches)	89.0%	Resilient to occlusions
Brightness Variation	94.3%	Minimal impact
Rotation & Scaling	90.5%	Stable performance

5. Conclusion

The neural network demonstrates good robustness against common perturbations, with only moderate sensitivity to adversarial attacks. Further improvements could involve adversarial training and additional data augmentation techniques.