# ThreatSentry - Threat Assessment Report

**Model: microsoft/resnet-18**

Attack Type: FGSM

*Generated: 2025-11-02 21:21:02*

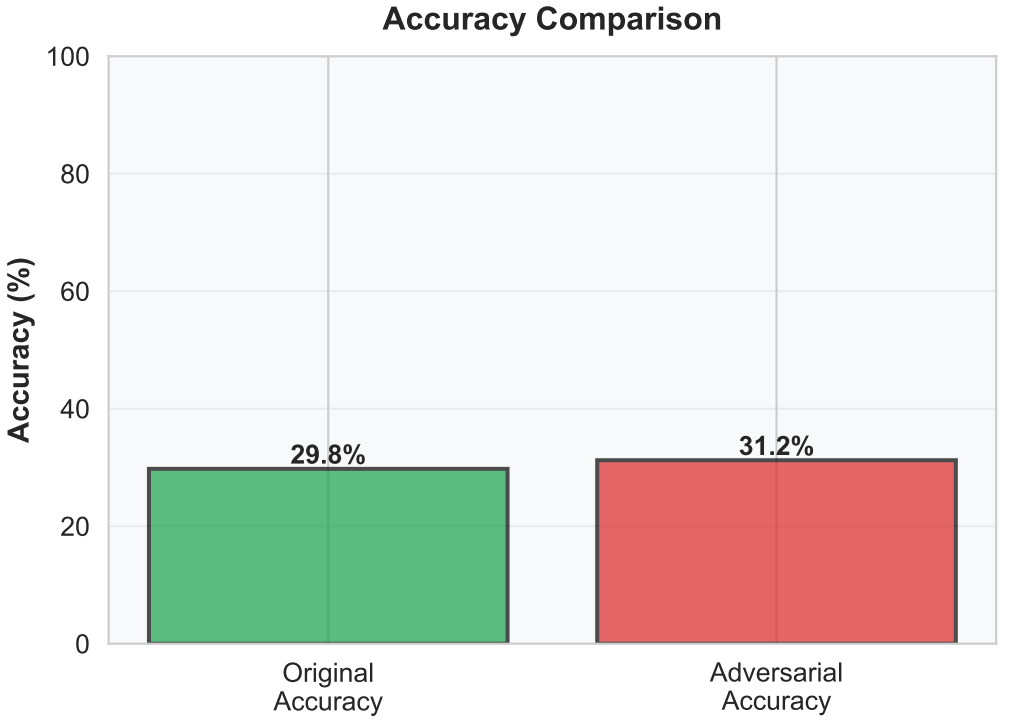
## Key Metrics

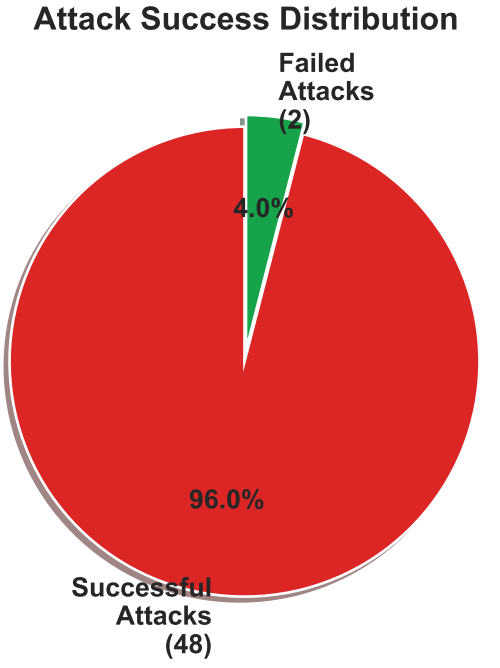
| | |
|---|---|
| **Attack Success Rate:** | 96.0% |
| **Original Accuracy:** | 29.77% |
| **Adversarial Accuracy:** | 31.23% |
| **Accuracy Drop:** | -1.47% |
| **Execution Time:** | 63.62s |
| **Images Tested:** | 50 |

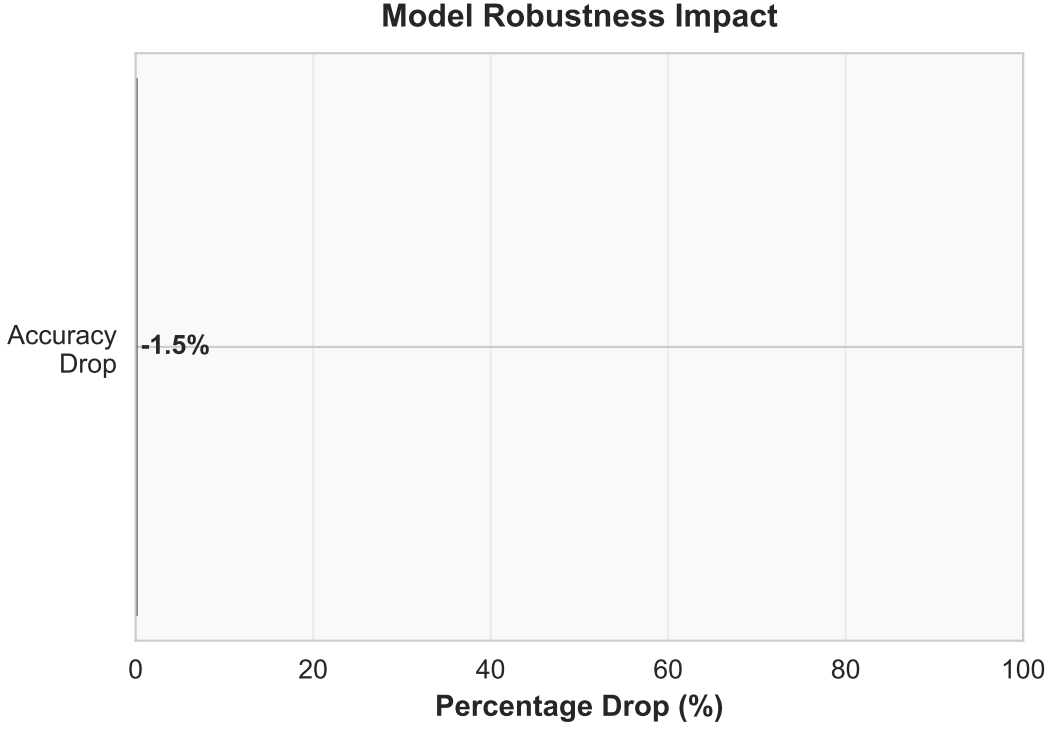
## Threat Level Assessment

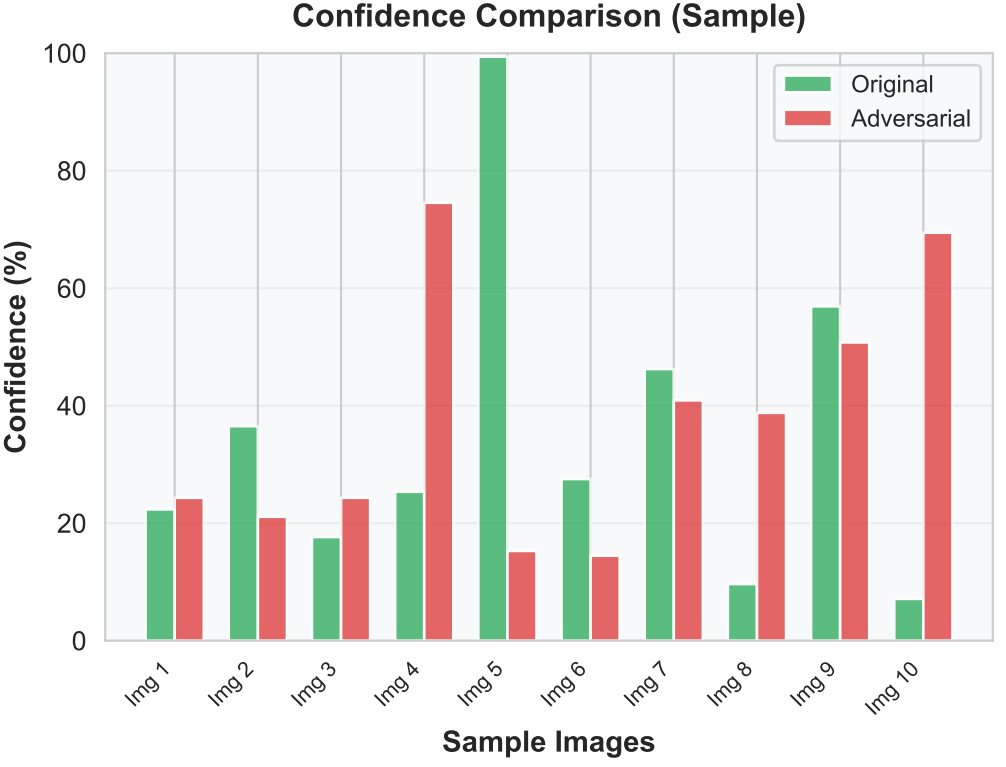
**HIGH RISK**

# Detailed Analysis



# Detailed Analysis & Recommendations

## Assessment Summary

Successfully executed FGSM attack on model microsoft/resnet-18 using 50 test images. Attac
k success rate: 96.0%. Average original accuracy: 29.77%, Average adversarial accuracy: 31
.23%. The attack successfully fooled the model in 48 out of 50 cases.

## Security Recommendations

**1. Implement Adversarial Training**

- Retrain your model with adversarial examples to improve robustness
- Use techniques like FGSM, PGD during training phase

**2. Add Input Validation & Preprocessing**

- Implement input sanitization and anomaly detection
- Use defensive distillation or feature squeezing

**3. Deploy Ensemble Methods**

- Use multiple models with different architectures
- Implement voting mechanisms for predictions

**4. Continuous Monitoring**

- Set up real-time performance monitoring
- Detect and alert on unusual prediction patterns



**5. Regular Security Audits**

- Conduct periodic threat assessments
- Stay updated with latest attack techniques