ThreatSentry - Threat Assessment Report

Model: google/mobilenet_v2_1.0_224

Attack Type: DEEPFOOL

Generated: 2025-11-02 13:49:48

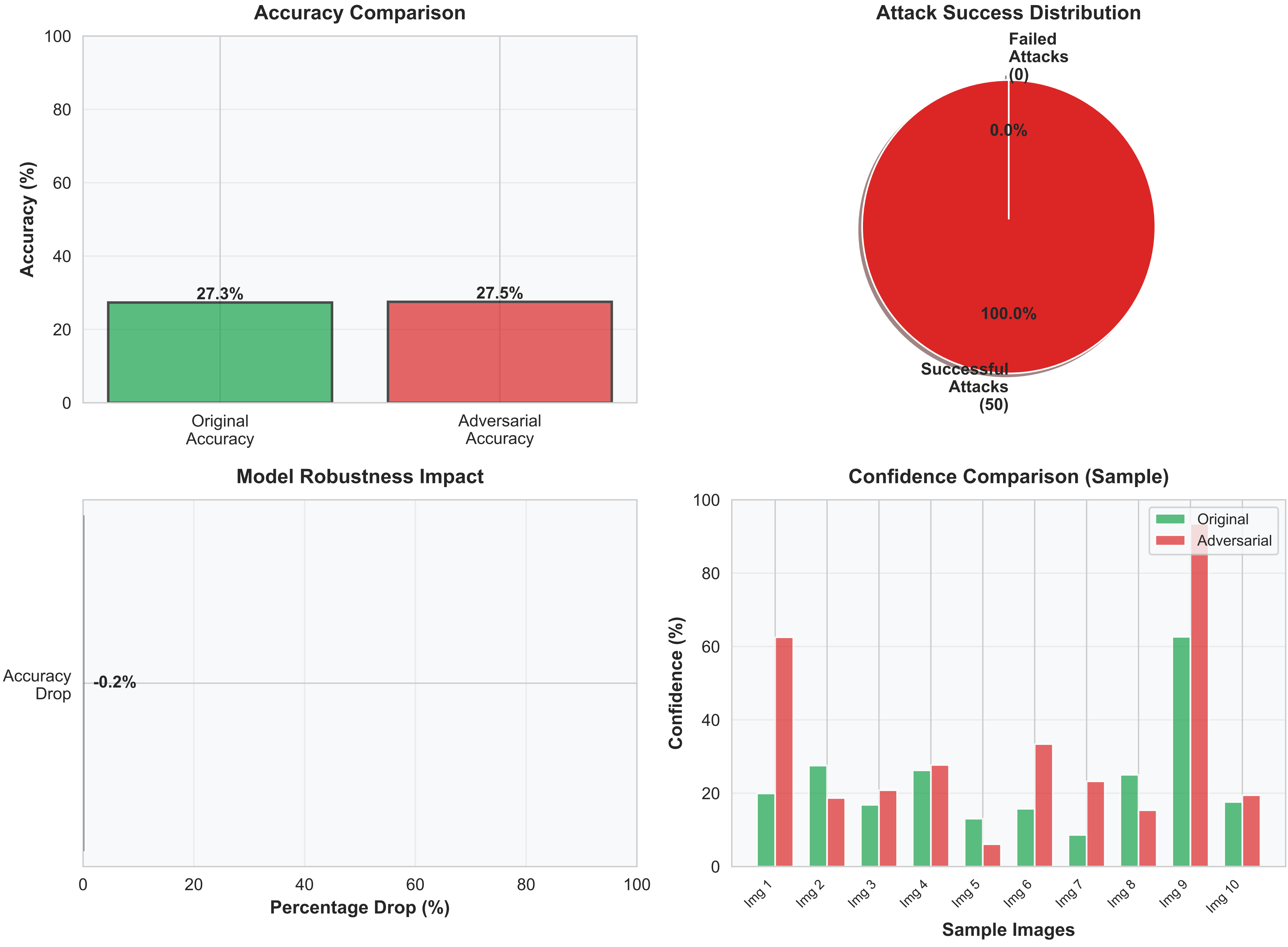
Key Metrics

Attack Success Rate:	100.0%
Original Accuracy:	27.34%
Adversarial Accuracy:	27.53%
Accuracy Drop:	-0.19%
Execution Time:	214.94s
Images Tested:	50

Threat Level Assessment

HIGH RISK

Detailed Analysis



Detailed Analysis & Recommendations

Assessment Summary

Successfully executed DEEPFOOL attack on model google/mobilenet_v2_1.0_224 using 50 test images. Attack success rate: 100.0%. Average original accuracy: 27.34%, Average adversarial accuracy: 27.53%. The attack successfully fooled the model in 50 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