

# ThreatSentry - Threat Assessment Report

**Model:** google/efficientnet-b0

**Attack Type:** FGSM

*Generated: 2025-11-02 21:27:05*

## Key Metrics

**Attack Success Rate:** 86.0%

**Original Accuracy:** 31.64%

**Adversarial Accuracy:** 24.62%

**Accuracy Drop:** 7.02%

**Execution Time:** 60.53s

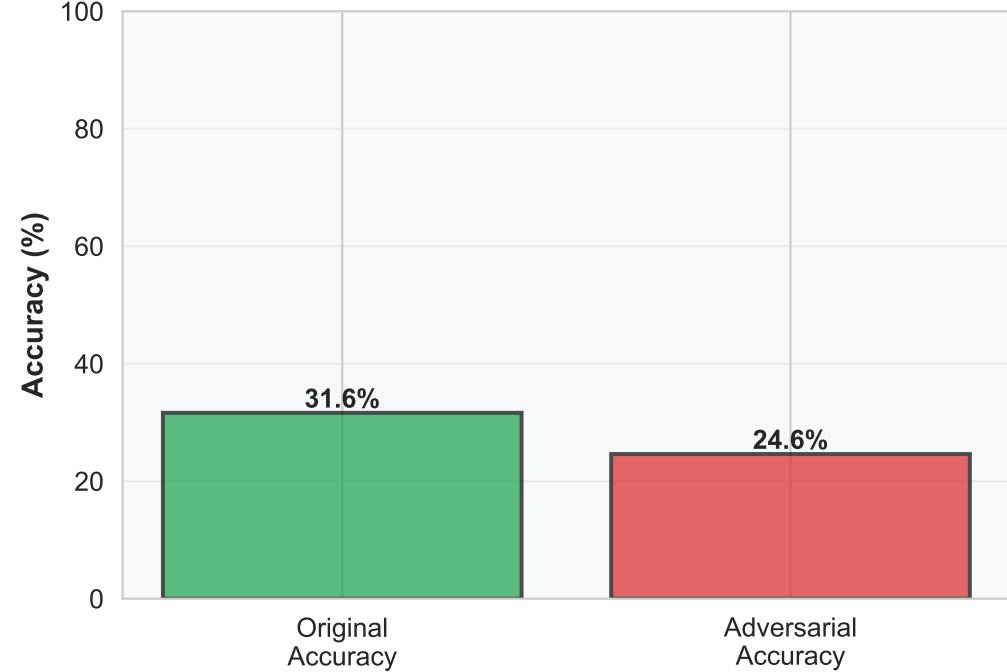
**Images Tested:** 50

## Threat Level Assessment

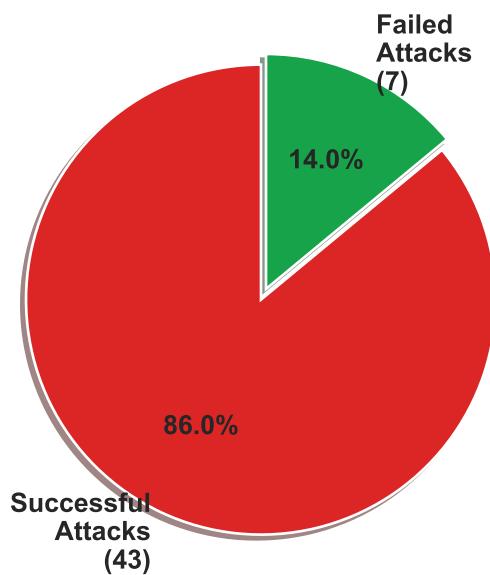
**HIGH RISK**

# Detailed Analysis

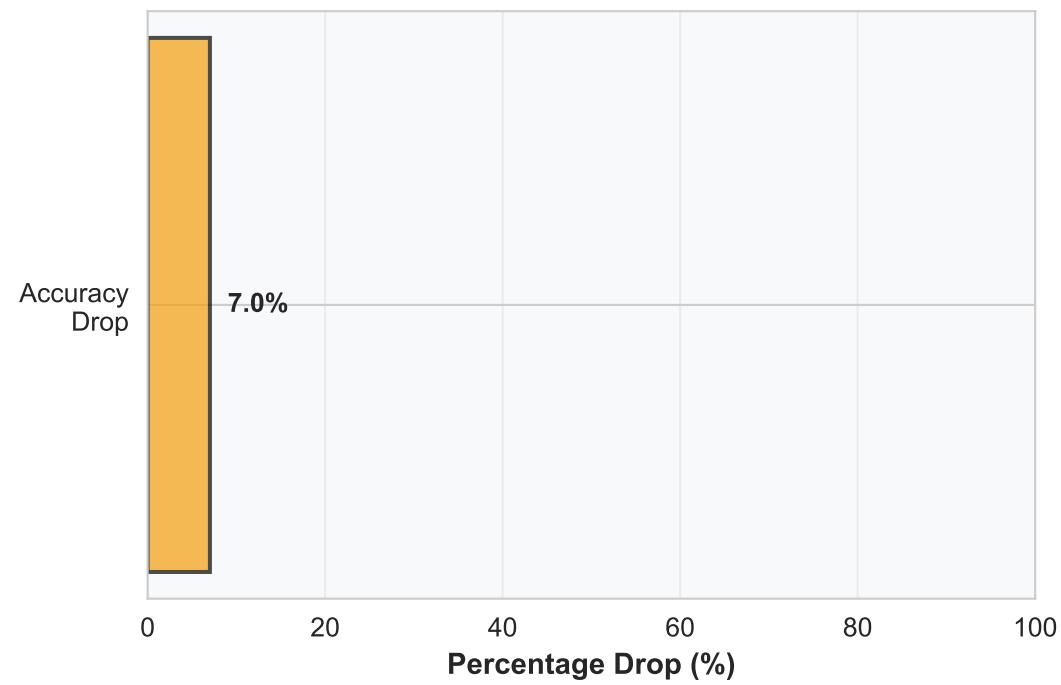
## Accuracy Comparison



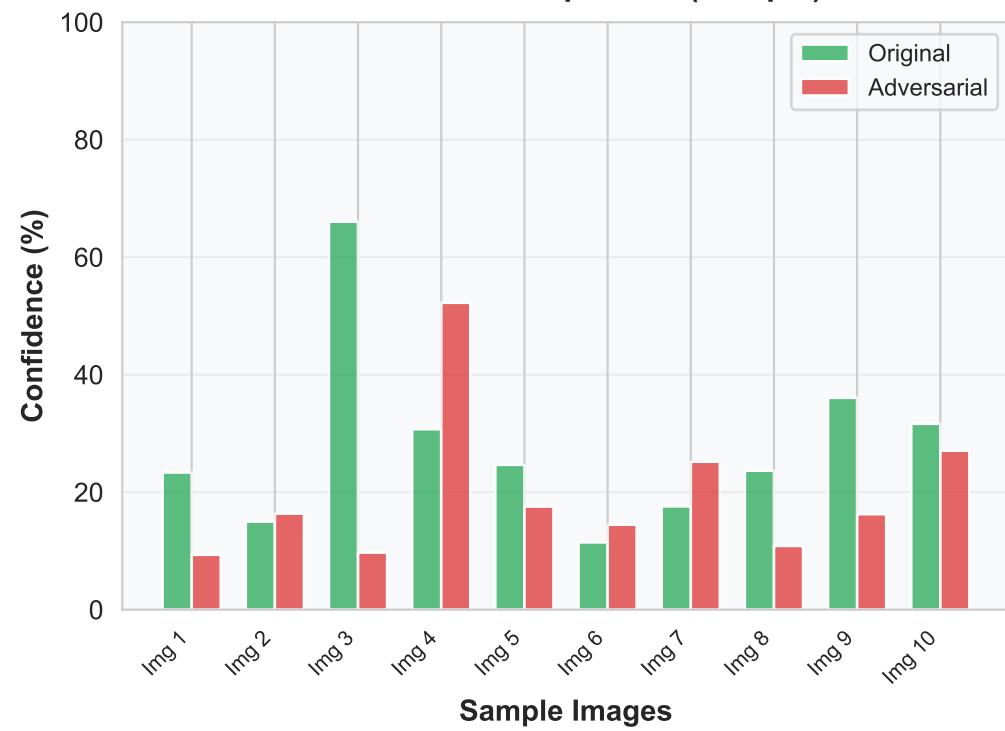
## Attack Success Distribution



## Model Robustness Impact



## Confidence Comparison (Sample)



# Detailed Analysis & Recommendations

## Assessment Summary

Successfully executed FGSM attack on model google/efficientnet-b0 using 50 test images. Attack success rate: 86.0%. Average original accuracy: 31.64%, Average adversarial accuracy: 24.62%. The attack successfully fooled the model in 43 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