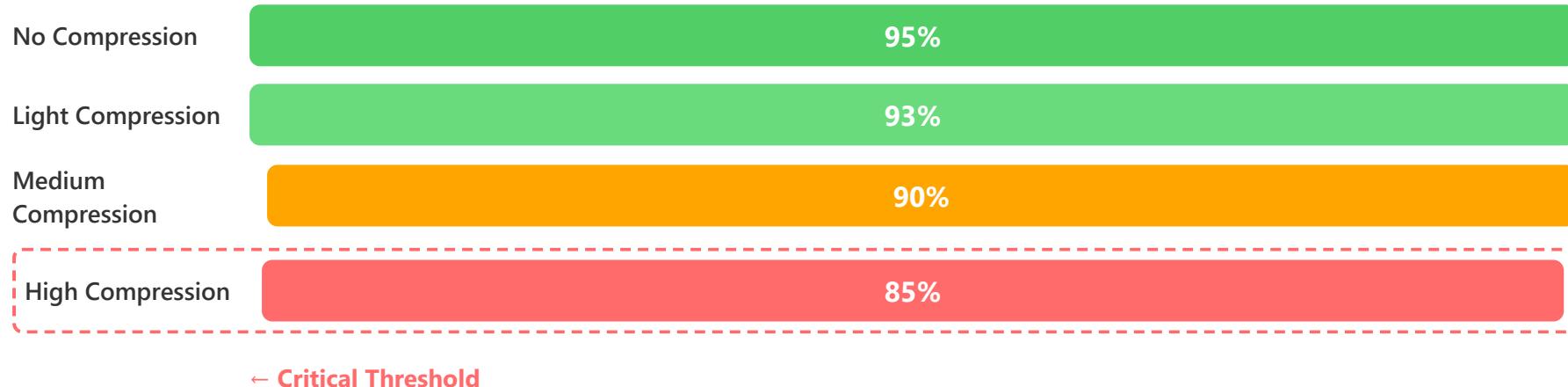


Accuracy Preservation

Accuracy Preservation

Methods to maximally preserve model performance during compression

Accuracy Changes by Compression Rate



Accuracy Preservation Strategies

1

Quantization-Aware Training (QAT)

Simulating quantization during training → Minimizing accuracy loss

2

Fine-tuning after Compression

Performance recovery through retraining after compression

3

Sensitivity Analysis

Selective compression after layer-wise sensitivity analysis

4

Knowledge Distillation

Accuracy preservation by leveraging teacher knowledge

Special Considerations for Medical AI:

Accuracy degradation directly affects patient safety → Strict threshold setting required



Verify compliance with FDA/CE approval standards