

Class Imbalance

⚠ The Accuracy Paradox

99% accuracy is useless if 99% of samples are negative!

Model predicting all negatives achieves high accuracy but zero clinical utility

Sampling Strategies

- Random oversampling
- Random undersampling
- SMOTE (Synthetic Minority)
- ADASYN (Adaptive Synthetic)

Cost-sensitive Learning

- Weighted loss functions
- Class weights in sklearn
- Focal loss for deep learning
- Penalize misclassifications differently

✓ Proper Evaluation Metrics

Precision

Recall

F1-score

PR-AUC

Balanced Accuracy

Matthews CC

Cohen's Kappa

G-mean