

Cross-validation Strategies

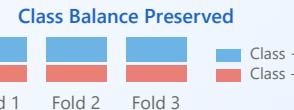
K-Fold CV



Split data into K folds, train on K-1, test on 1, repeat K times

Standard: K=5 or 10

Stratified CV



Preserve class distribution in each fold - crucial for imbalanced data

Use for classification

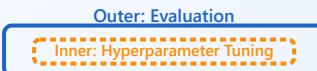
Leave-One-Out



Train on N-1 samples, test on 1, repeat N times

For very small N

Nested CV

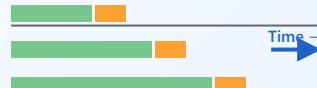


Unbiased performance estimate

Outer loop for evaluation, inner loop for hyperparameter tuning

Unbiased estimates

Time Series Split



Train on past, test on future - respects temporal order

For longitudinal data

Group CV



Keep all samples from same patient/site together

Prevent data leakage

⚠ Biomedical Pitfall

Multiple samples from same patient must stay in same fold to avoid optimistic bias!