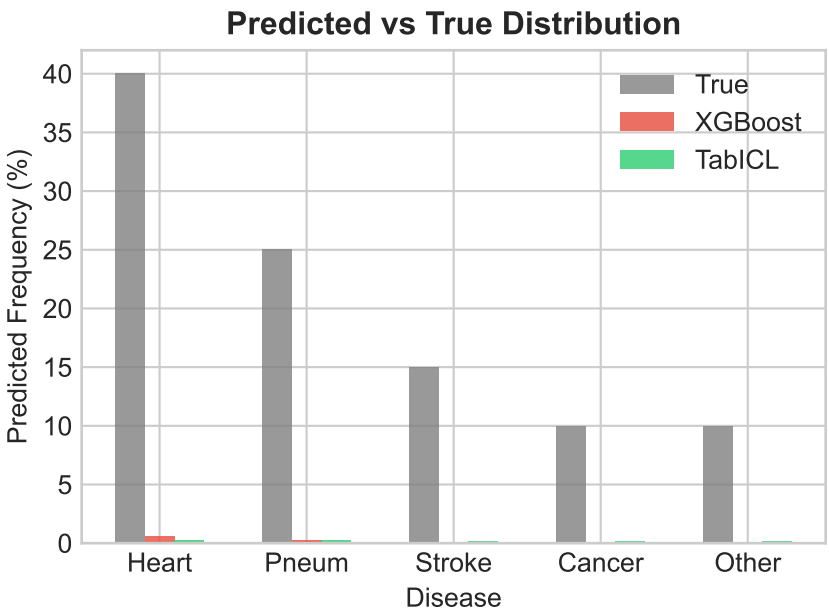
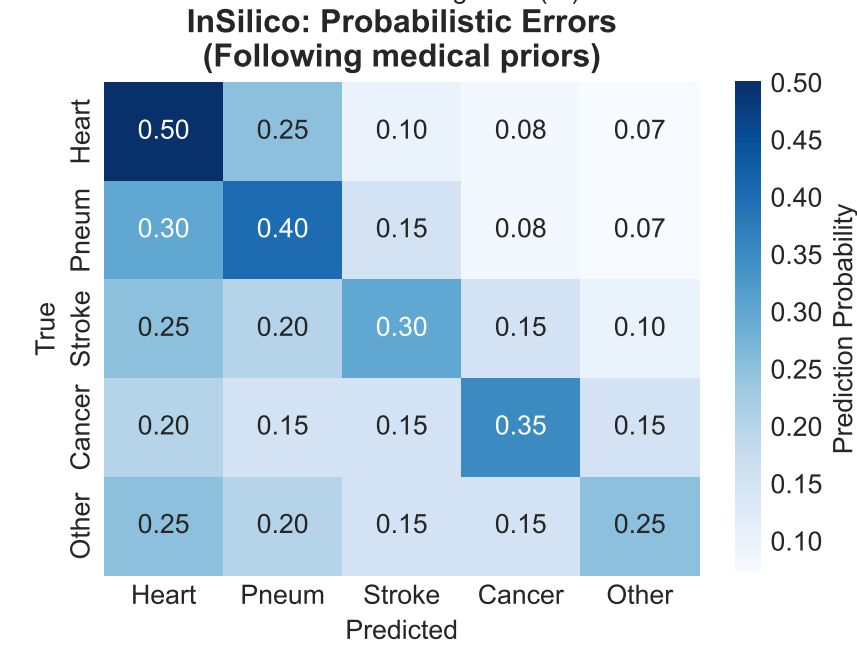
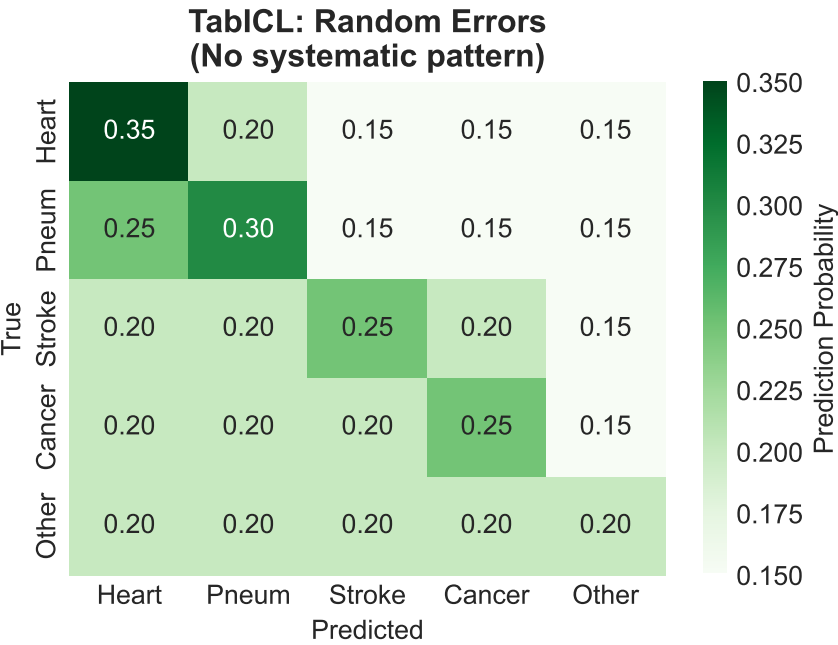
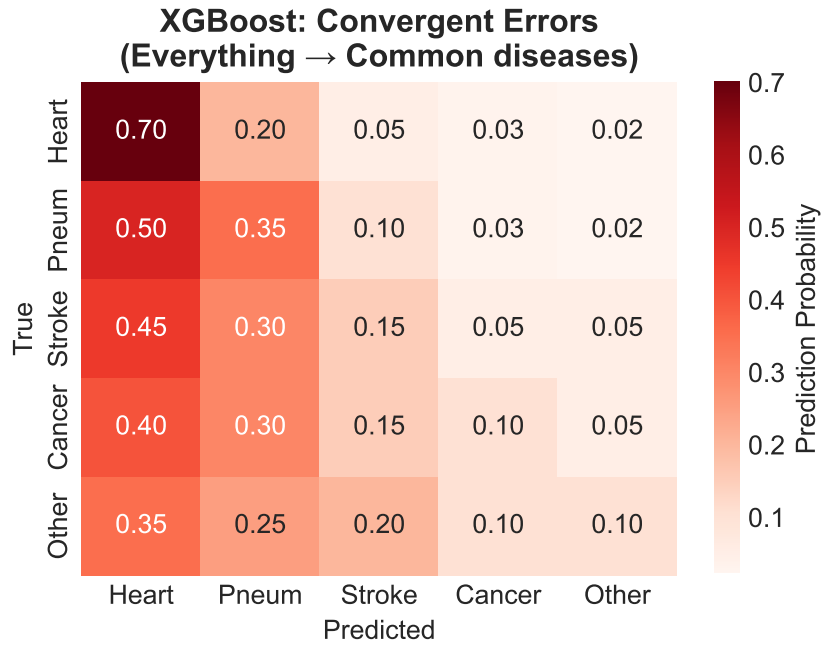
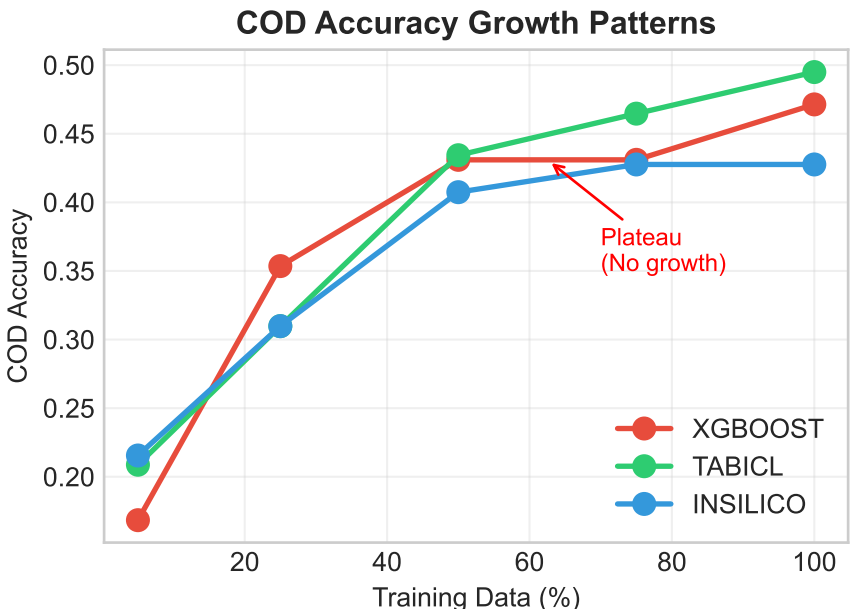
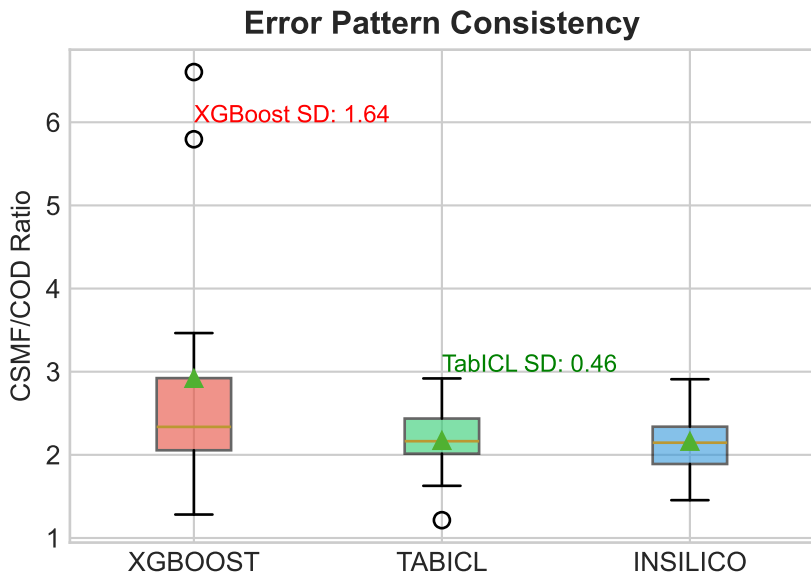
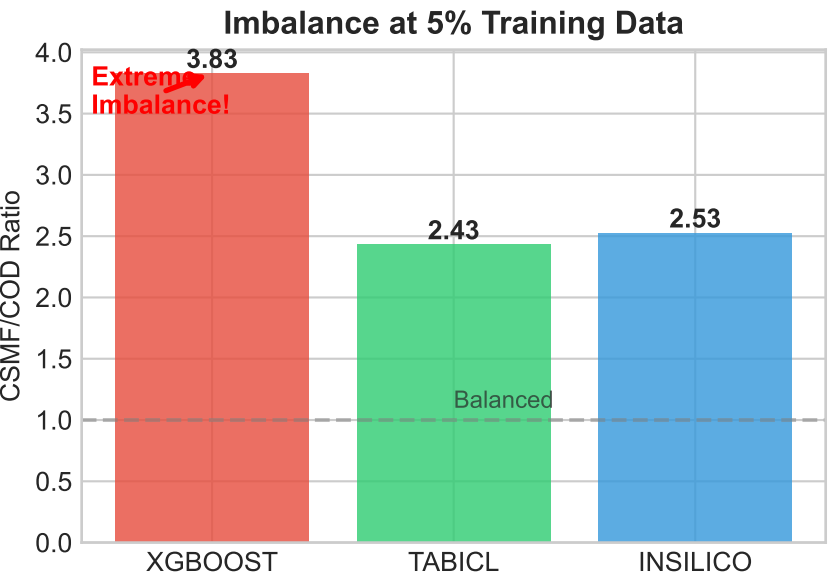


Why TabICL and InSilico Don't Show XGBoost's CSMF/COD Imbalance



FUNDAMENTAL MECHANISMS:

XGBoost (Tree-based):

- With 59 samples across 34 classes (~1.7/class)
- Trees can only learn: "Is it common?" → Yes/No
- Result: Funnel effect → All predictions converge to top 2-3 classes
- High CSMF (distribution ≈ correct), Low COD (individuals wrong)

TabICL (In-context learning):

- Each prediction uses different k examples
- Example Set 1 → Prediction A
- Example Set 2 → Prediction B (different!)
- Result: Random scatter → Errors don't concentrate
- Balanced CSMF and COD (both moderate)

InSilico (Bayesian):

- $P(\text{Disease}|\text{Symptoms}) = P(\text{Symptoms}|\text{Disease}) \times P(\text{Disease})$
- Even with weak evidence, priors provide stability
- Result: Calibrated errors → Follow medical probabilities
- Balanced CSMF and COD (both reasonable)

KEY INSIGHT: XGBoost's systematic bias creates the imbalance, while TabICL's randomness and InSilico's priors prevent it.