Real-World Bug Prediction: Paradigm Comparison

F1 Score

- Supervised ML (Stacking): 0.42
- Change-Point (Binary Seg, $\tau=10$): 0.54
- Forecasting (Naive Mean): 0.35