Appendix – Comparison with Flawless Dataset

We also worked on a secondary dataset previously referred to as flawless, which was synthetic and generated to resemble a real medical dataset.

Key Differences:

Aspect	Flawless Dataset	UCI Cleveland Dataset
Source	AI-generated or synthetic	Real patient data (UCI repository)
Distribution	Flat/unrealistic	Realistic and naturally skewed
Class Correlation	Extremely low (0.01–0.02)	Meaningful (up to ±0.5)
Coefficients	Over 1000+ dummy features	Only 13 interpretable variables
Model Results	Misleadingly poor or inconsistent	High performance and consistent
Visual Patterns	No medical logic	Matches clinical expectations

The comparison validates the reliability and importance of using authentic datasets in medical ML.