SUPPLEMENTARY TABLE

Table S1: Dataset-Wise Preprocessing Checklist (Leakage-Free, Fold-Wise Fitting)

Dataset (UCI)	n, p	Continuous missing values	Categorical missing values	Scaling	Outlier handling	Encoding of categorical variables	Leakage prevention
Chronic Kidney Disease (CKD)	400, 24	KNN imputation (k=5), applied as needed	Mode imputation, applied as needed	z-score standardization (all numeric features)	IQR-based detection + boundary clipping	Integer label encoding (where categorical predictors are present)	Fit on training fold; apply to validation fold only
PIMA Diabetes	768, 8	KNN imputation (k=5), applied as needed	Not applicable (no categorical predictors in this dataset version)	z-score standardization (all numeric features)	IQR-based detection + boundary clipping	Not applicable	Fit on training fold; apply to validation fold only
Heart Disease (Cleveland)	303, 13	KNN imputation (k=5), applied as needed	Mode imputation, applied as needed	z-score standardization (all numeric features)	IQR-based detection + boundary clipping	Integer label encoding (where categorical predictors are present)	Fit on training fold; apply to validation fold only
Breast Cancer (Wisconsin)	569, 30	KNN imputation (k=5), applied as needed	Not applicable (no categorical predictors in this dataset version)	z-score standardization (all numeric features)	IQR-based detection + boundary clipping	Not applicable	Fit on training fold; apply to validation fold only

Note: KNN imputation used k=5 for continuous variables; categorical imputation used the mode where categorical predictors were present. All preprocessing operations (imputation, scaling, encoding, outlier clipping) were performed in a leakage-free manner by fitting on training folds and applying to validation folds only.