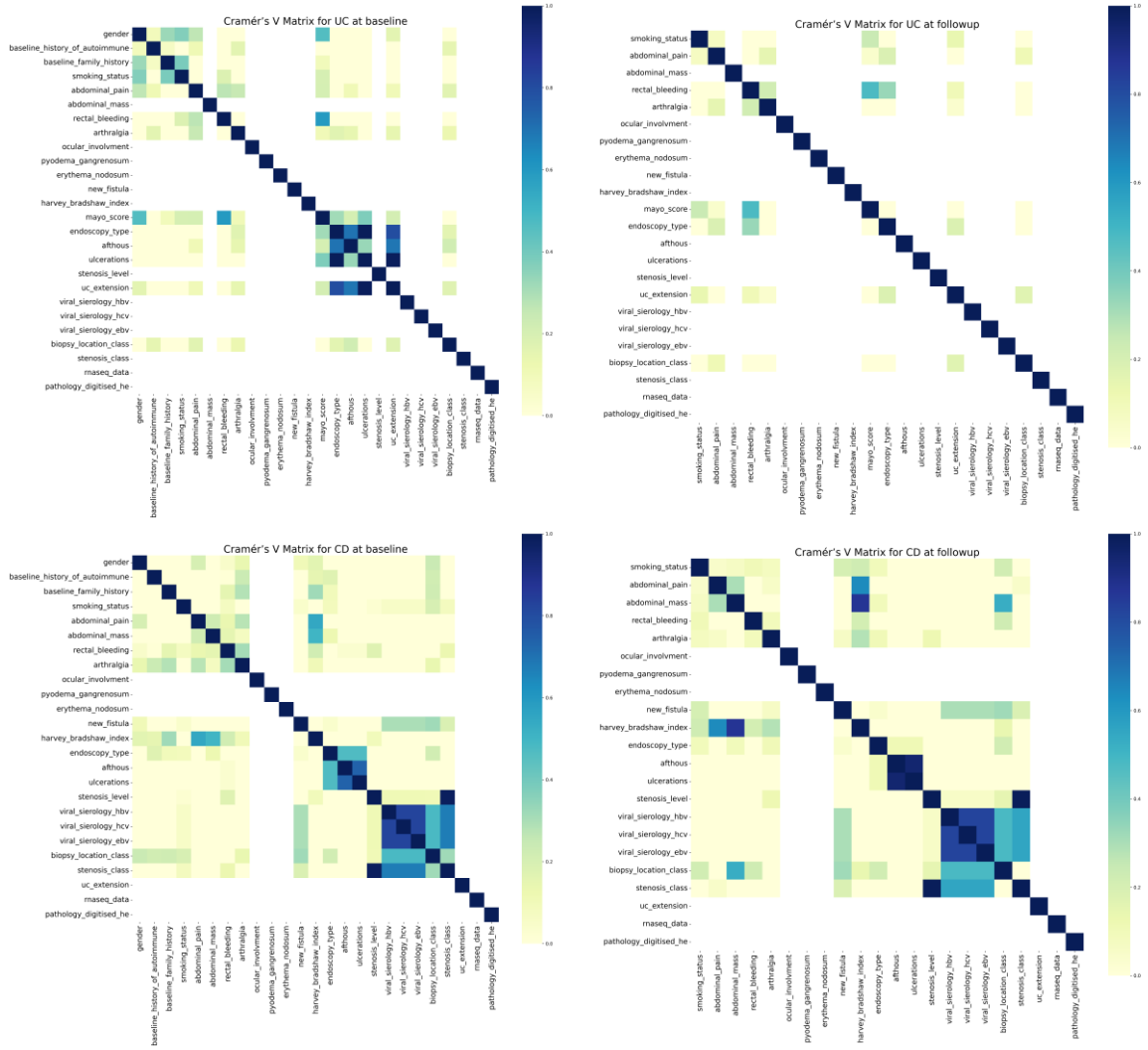
Supplementary materials

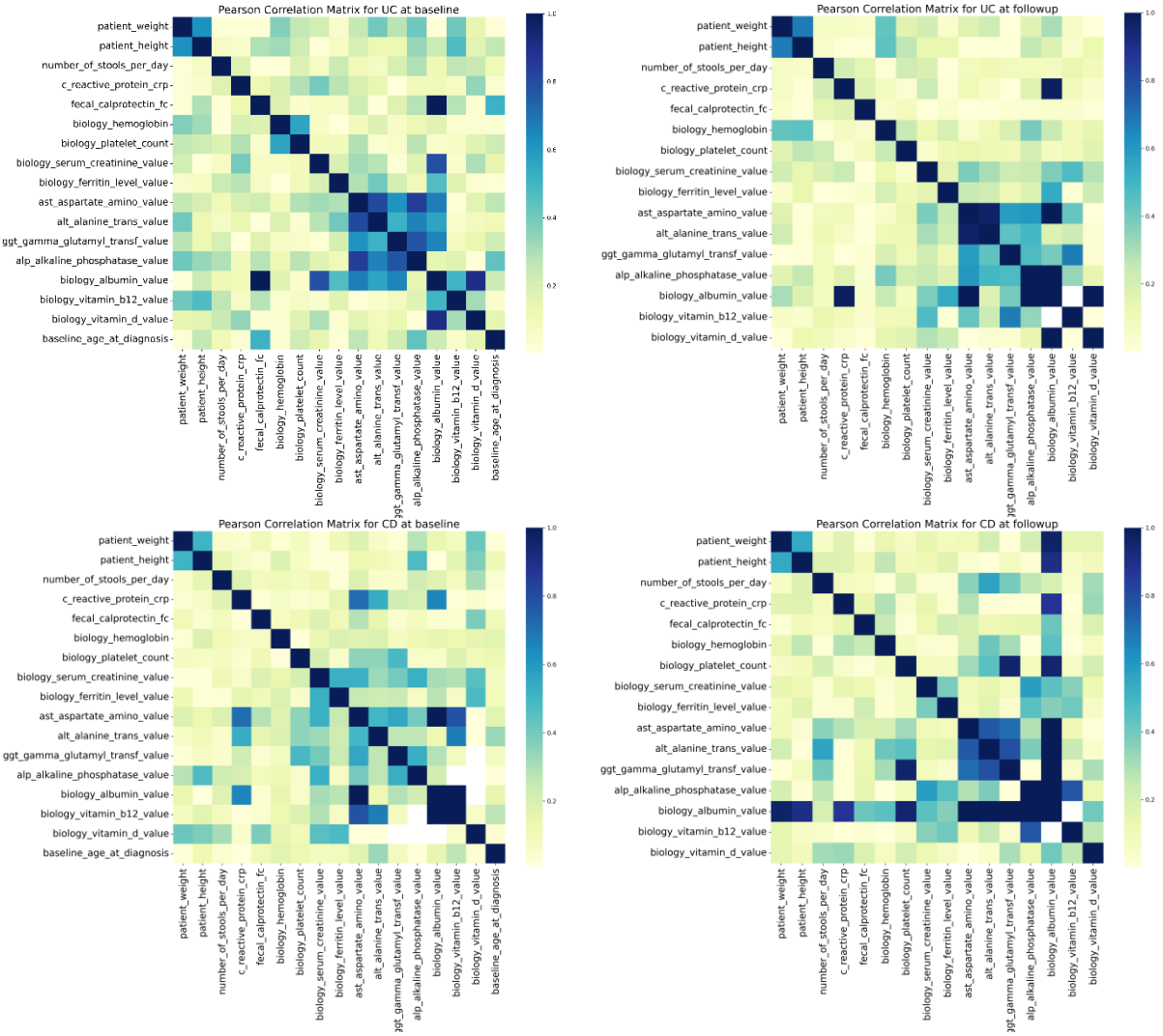
Collecting and Analyzing IBD Clinical Data for Machine Learning: Insights from an Italian Cohort

1. *Statistical Analysis:*

The dataset was analyzed for multicollinearity using Pearson's r for continuous variables and Cramer's V for categorical ones in UC and CD cohorts. Most correlations were low to moderate, with few strong associations, indicating limited redundancy in features. At follow-up, expected recovery-related correlations emerged (e.g., weight and albumin), but overall independence was maintained. Categorical features also showed low dependency, supporting the dataset's suitability for multivariable and machine learning models. Correlation matrices are illustrated in Supplementary Figure 1 and 2.



Supplementary Figure S1: Cramer's V correlation matrices (categorical variables) for Ulcerative Colitis (UC) and Chron's Disease (CD) at baseline and followup.



Supplementary Figure S2: Pearson correlation matrices (continuous variables) for Ulcerative Colitis (UC) and Chron's Disease (CD) at baseline and followup.