

Supplementary Material

Table S1 The performance of all models in the training and validation sets

model_name	Accuracy	AUC	95% CI	Group
NaiveBayes	0.700	0.762	0.721 - 0.803	training
NaiveBayes	0.603	0.811	0.745 - 0.876	validation
SVM	0.646	0.763	0.716 - 0.811	training
SVM	0.763	0.634	0.506 - 0.763	validation
KNN	0.855	0.878	0.854 - 0.901	training
KNN	0.767	0.654	0.557 - 0.751	validation
RandomForest	0.874	0.974	0.965 - 0.983	training
RandomForest	0.685	0.652	0.550 - 0.755	validation
ExtraTrees	0.917	0.983	0.977 - 0.990	training
ExtraTrees	0.685	0.608	0.506 - 0.709	validation
XGBoost	0.807	0.901	0.875 - 0.927	training
XGBoost	0.681	0.709	0.619 - 0.798	validation
LightGBM	0.739	0.854	0.824 - 0.885	training
LightGBM	0.728	0.698	0.593 - 0.804	validation
GradientBoosting	0.698	0.810	0.773 - 0.847	training
GradientBoosting	0.496	0.735	0.650 - 0.820	validation
AdaBoost	0.646	0.774	0.736 - 0.813	training
AdaBoost	0.608	0.778	0.702 - 0.856	validation
LR	0.551	0.763	0.722 - 0.804	training
LR	0.749	0.812	0.737 - 0.887	validation

AUC, Area Under the Curve; CI, Confidence Interval; SVM, Support Vector Machine; KNN, K Nearest Neighbor; LR, Logistic Regression.

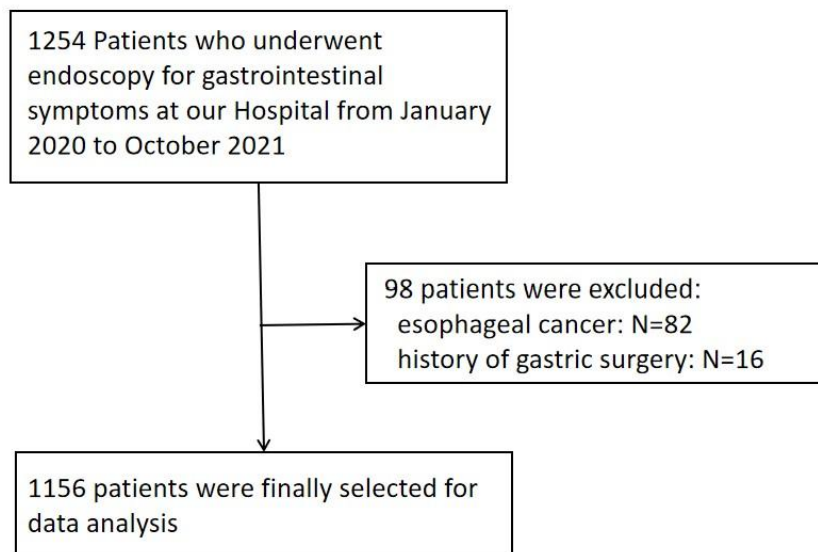


Figure S1. The flowchart for patients selection