Clinically Interpretable Machine Learning Models for Predicting the Stages of Gastric Cancer

Guanmo Liu[1], Jie Li[1], Xiaoqian Liu[2], Weiming Kang[1]
1Department of General Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.
2Department of Automation, Tsinghua University
E-mail: Guanmo Liu jaisy_princeton@sina.com

Aims: Gastric cancer (GC) constitutes the third leading cause of cancer-related mortality worldwide[1]. Early GC has a 5-year survival rate over 90%, compared to only less than 30% in progressive GC [2]. Existing clinical methods for determining the stages of GC are still facing some challenges. This study aimed to integrate preoperative blood test parameters and their novel ratio features to develop an interpretable model to precisely differentiate stage I from stage II-III GC using machine learning (ML) models to facilitate diagnostic decision-making further.

Materials and methods: We systematically evaluated eleven supervised ML models. Input features comprised both original preoperative blood parameters and biologically meaningful ratio features, to capture inflammatory, metabolic, and tumor-associated interactions. Feature standardization and SMOTEENN were applied to address imbalance. Model performance was assessed using AUC, accuracy, precision, recall, and F1-score. CatBoost yielded the highest predictive performance among them and was selected for further optimization. SHapley Additive exPlanations (SHAP) were used to interpret feature contributions. A forward selection strategy guided by SHAP rankings was performed across feature subsets (top 5–30), identifying an optimal nine-feature panel. The final model was validated using nested 10-fold cross-validation and 1,000-iteration bootstrapping, confirming high stability and robustness with a 95% CI for AUC.

Results: Among 434 GC patients included, 57.8% were diagnosed with pathological stage I. After incorporating ratio-based features, the CatBoost achieved the best performance among eleven ML models, with AUC improving markedly from 0.802 (Fig. 1A) to 0.981 (Fig. 1B). SHAP analysis ranked the top contributing features as UA, APTT, Eos%, RDW/HCT, CEA, CA242, ChE, MCH, and ApoA1. UA was identified as a protective factor for stage I, whereas elevated APTT and ChE levels were associated with stage II–III. Model robustness was confirmed via 10-fold cross-validation, yielding an average AUC of 0.9477 with minimal performance variance, and bootstrap resampling produced a stable 95% CI of [0.9379,

0.9569].

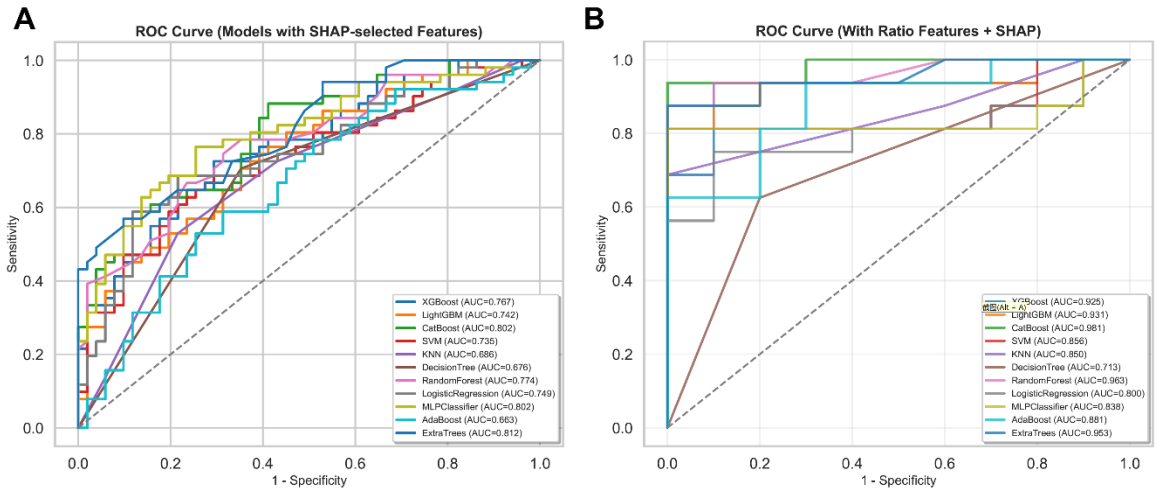

Fig. 1 (A) The receiver operating characteristic curve among eleven ML models for patients with GC. (B) The receiver operating characteristic curve among eleven ML models with ratio features for patients with GC.

Conclusions: This interpretable CatBoost model with high performance was innovatively integrated with meaningful blood ratio features enabling accurate, noninvasive staging of GC. The model may optimize preoperative decision-making, reducing overtreatment in early-stage GC and ensuring timely neoadjuvant therapy for advanced cases.

References:

1. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.* CA Cancer J Clin, 2021. **71**(3): p. 209-249.
2. Huang, Y., et al., *Global progress and future prospects of early gastric cancer screening.* J Cancer, 2024. **15**(10): p. 3045-3064.