Performance of explainable ensemble learning for mortality risk stratification and multimodal biomarker prediction in colorectal cancer: a retrospective, database cohert study

Zhicheng Du,[a,*] Huiyan Luo,[b,c] Xuefei Li,[d,e] Yang Liu,[a] Xi Yuan,[f] Zhenglin Chen,[a,g] Jiansong Ji,[g] and Peiwu Qin[a,g]

[a]Institute of Biopharmaceutical and Health Engineering, Shenzhen International Graduate School, Tsinghua University, China
[b]Department of Medical Oncology, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, China
[c]Research Unit of Precision Diagnosis and Treatment for Gastrointestinal Cancer, Chinese Academy of Medical Sciences, China
[d]CAS Key Laboratory for Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[e]University of Chinese Academy of Sciences, China
[f]School of Communication Engineering, Hangzhou Dianzi University, China
[g]Zhejiang Key Laboratory of Imaging and Interventional Medicine, Zhejiang Engineering Research Center of Interventional Medicine Engineering and Biotechnology, The Fifth Affiliated Hospital of Wenzhou Medical University, China

The Lancet Regional Health - Western Pacific 2025;55S1: 101459
https://doi.org/10.1016/j.lanwpc.2024.101459

Abstract

Background Colorectal cancer (CRC) is the third most common cancer globally and the second leading cause of cancer-related deaths. In 2020, an estimated 1.93 million new CRC cases and 940,000 CRC-related deaths were reported worldwide. In China, 560,000 new cases accounted for 28.8% of the global total, with 290,000 deaths representing 30.6% of CRC-related fatalities. Precision oncology for CRC necessitates the assessment of multimodal biomarkers, including microsatellite instability (MSI), tumor mutational burden (TMB), and mutations in the BRAF and RAS genes. However, the high costs, prolonged processing times, complex procedures, and the requirement for advanced laboratory infrastructure hinder timely and accurate clinical decision-making. This study aims to develop and assess interpretable ensemble deep learning for predicting prognosis and four biomarkers (MSI, TMB, RAS and BRAF mutations) using electronic health records (EHR).

Method We utilized a large genomic and clinical database from the ChangKang (Healthy Bowel) project, comprising 1,015 CRC patients from the Chinese population between 2017 and 2021. The database includes information relevant to early screening, diagnosis, prognostic evaluation, postoperative recurrence monitoring, individualized treatment, efficacy evaluation, and drug resistance analysis. Patients were categorized into "high-risk" and "low-risk" groups based on sequencing data and overall survivor (OS). We integrated state-of-the-art foundation language models for medical language to analyze the semantic features of EHR and constructed an ensemble deep learning model using multi-scale feature fusion. A 10-fold cross-validation strategy was implemented to split the cohort for model training (914 patients) and evaluation (101 patients).

Findings Our study demonstrated that patients can be stratified into high-risk and low-risk groups using ensemble deep learning. The model achieved an average accuracy of 96.83% (95.33%-98.71%) and an average F1-score of 96.83% (95.23%-98.75%) in predicting MSI, an accuracy of 97.16% (94.52%-98.67%) and an F1-score of 97.16% (94.55%-98.67%) in predicting TMB, an accuracy of 97.10% (94.44%-100.00%) and an F1-score of 97.12% (94.51%-100.00%) in predicting BRAF mutation, a lower accuracy of 73.49% (71.18%-76.28%) and an F1-score of 73.91% (71.73%-76.82%) in predicting RAS mutation. The OS prediction showed an accuracy of 87.78% (83.33%-88.89%) and an F1-score of 87.98% (85.71%-88.89%). Notably, the model dynamically adjusted the contribution of input data according to the prediction task, with pathological grade being the most significant factor across all five tasks.

Interpretation Our findings indicate that interpretable ensemble deep learning can robustly predict multimodal biomarkers and mortality risk of CRC patients, which is beneficial to improve diagnostic accuracy, treatment planning, and patient stratification. Notably, our model provides interpretable insights into the associations between

*Correspondence to duzc21@mails.tsinghua.edu.cn (Z. Du).

the input data and the prediction outcomes, enabling clinicians to intuitively understand the rationale behind the model's predictions and assess their reliability.

Funding The National Natural Science Foundation of China 32350410397; Science, Technology, Innovation Commission of Shenzhen Municipality, JCYJ20220530143014032, JCYJ20230807113017035, KCXFZ20211020163813019, Shenzhen Medical Research Funds, D2301002; Department of Chemical Engineering-iBHE special cooperation joint fund project, DCEiBHE-2022-3; Tsinghua Shenzhen International Graduate School Cross-disciplinary Research and Innovation Fund Research Plan, JC2022009; and Bureau of Planning, 669 Land and Resources of Shenzhen Municipality (2022) 207.

Keywords: Deep learning; Colorectal cancer; Computer-aided Diagnosis; AI for biomedicine

Declaration of interests
We declare no competing interests.