Application of Artificial Intelligence in a Real-World Research for Predicting the Risk of Liver Metastasis in T1 Colorectal Cancer

Tenghui Han  
Air Force Medical University Xijing Hospital: Xijing Hospital

Jun Zhu  
Air Force Medical University Xijing Hospital: Xijing Hospital

Dong Xu  
Xi'an Medical University

Rujie Chen  
Air Force Medical University Xijing Hospital: Xijing Hospital

Shuai Wang  
Ming gang station hospital, Xi'an Institute of flight of the air force

Xiaoping Chen  
Southern Theater Air force Hospital

Jianyong Zheng  
Air Force Medical University Xijing Hospital: Xijing Hospital

Chunsheng Xu (✉ hjb2015@stu.xjtu.edu.cn)  
Air Force Medical University Xijing Hospital: Xijing Hospital  https://orcid.org/0000-0001-5755-9794

Research Article

Keywords: artificial intelligence, machine learning, T1 colorectal cancer, real-world research, liver metastasis

DOI: https://doi.org/10.21203/rs.3.rs-746689/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

**Background:** The liver is the most common metastatic site of colorectal cancer (CRC) and liver metastasis (LM) determines subsequent treatment as well as prognosis of patients, especially in T1 patients. T1 CRC patients with LM are recommended to adopt surgery and systematic treatments rather than endoscopic therapy alone. However, there is still no effective model to predict the risk of LM in T1 CRC patients and we aim to develop a novel and accurate predictive model.

**Methods:** We integrated two independent CRC cohorts from Surveillance Epidemiology and End Results database (SEER) and Xijing hospital. Artificial intelligence (AI) and machine learning methods were adopted to establish the predictive model.

**Results:** A total of 16785 and 326 T1 CRC patients from SEER database and our hospital were incorporated respectively in the study. We found that age, gender, married status, primary site, tumor size, carcinoembryonic antigen (CEA), tumor type, grade, N stage and perineural invasion were significant independent factors for predicting the presence of LM, among which tumor size is the most important. The stacking bagging model showed the best predictive capability, achieving a sensitivity of 0.8452, a specificity of 0.9566, and an area under the curve of 0.9631. In addition, the stacking model had an excellent discriminative ability and accurately screened out eight LM cases from 326 T1 patients in the outer validation cohort. Ultimately, we authenticated the prognostic value of the stacking model, which is consistent with the predictive result of LM.

**Conclusion:** We successfully established an innovative and convenient AI model for predicting LM in T1 CRC patients, which was further verified in our dataset.

Introduction

Colorectal cancer (CRC) is one of the most prevalent gastrointestinal tract malignancies with considerably high morbidity and mortality, attention of which is universally acknowledged to increase annually [1–3]. With 2/3 of CRC patients, metastasis is commonly deemed as a crucial clinical feature and a risk factor of high mortality for intractable CRC [4]. During the progression of CRC, over 50% of patients tend to develop liver metastasis (LM) which is the predominant contributor to poor prognosis of CRC [4, 5].

Endoscopic therapy is a widespread acceptance and adoption treatment approach for T1 CRC patients. Whereas, for early CRC patients with LM, traditional surgery section and chemoradiotherapy are the most effective and recommended treatments, which significantly prolong the overall survival (OS) rate [6, 7]. However, considering the inferior early screening methods, approximately 90% of CRC patients with LM failed to be diagnosed precisely in the early stage and thus undergo incomplete endoscopic resection, which finally cause adverse clinical outcomes [8, 9]. Although scholars have completed abundant researches on metastasis-related signatures in vivo and vitro, a satisfactory predictive model of LM for CRC in early stages is still lacking [10–12]. Thus, it is necessary and urgent to develop an easily applicable model to accurately predict the risk of LM for patients in early course of CRC.

Currently, there is an increasing and irreversible trend of discipline integration between medical science and artificial intelligence (AI) [13–15]. Besides, the depth and breadth of the discipline integration have significantly enhanced [14, 15]. Researchers have employed machine learning (ML) as the breakpoint to better solve the complicated problem for clinical prediction and acquired several significant breakthroughs in CRC [16–18]. Given that most of the present studies merely focused on the public database when studying the apparent discrepancy among different populations, limitations ineluctably appeared. Consequently, clinical data involving the real outer validation is vital to construct a superior prediction model.

In the study, we established a comprehensive recognition model by adopting AI and ML algorithms for the first time, which could remarkably promote the identification of T1 CRC with LM and improve the prognosis of these patients in clinical practice. In addition, the predictive model was constructed via using clinical common and accessible parameters, and further validated in an independent CRC cohort.

Materials And Methods
Clinical Sample Collection

An open-access and publicly available CRC cohort was retrieved from Surveillance, Epidemiology, and End Results (SEER) Program database in the U.S. National Cancer Institute. The CRC cohort functions as a powerful resource for investigators to comprehensively understand the natural history of CRC and significantly ameliorate the healthcare quality for CRC patients [19, 20]. An additional outer validation cohort of CRC patients who underwent surgeries from 2010 to 2021 was obtained from Xijing hospital. The CRC cohort's inclusive criteria are shown as follows: 1) The primary diagnosis is CRC; 2) patients were diagnosed with T1 CRC; 3) patients with sufficient clinical data. In addition, CRC patients who have undergone neoadjuvant radiotherapy were excluded. Written and informed consent was obtained from all participants. All aspects of the clinical cohort study were evaluated by and included in the Institutional Ethics Committee of Xijing Hospital.

Study Population

T1 CRC is defined as a category of tumor that invades only the submucosa, regardless of the presence or absence of lymph node metastasis (LNM). Utilizing the SEER database which employed the 7th cancer TNM stages of the American Joint Committee, we analyzed the data of all patients diagnosed with T1 CRC from 2010 to 2016. Primary demographic data, tumor information and laboratory indexes were extracted by utilizing SEER disease codes and then employed for model construction. Basic demographic data include age at diagnosis, gender, race, and marital status. Tumor information contain primary site, size, grade, histologic category and TNM stage. Laboratory indexes involve carcinoembryonic antigen (CEA) prior to surgery, tumor deposits, and perineural invasion (PNI). Survival time and status were collected for further estimation of the predictive model. Additionally, the information of our validation cohort was normalized via following the criteria of the SEER database.

Construction of the Predictive Model

In our research, seven ML models were employed to predict LM in patients with T1 stage CRC. For tree decision models, we adopted Light Gradient Boosting Decision (LGBM), Random Forest (RF), and Classification and Regression Trees (CART). LGBM is a gradient boosting framework that utilizes the tree-based learning algorithm, which has been applied in the construction of medical models in recent years [21, 22]. RF is a widely employed ML algorithm to deal with classification and regression issues via the multiple decision trees approach [23]. CART is a classical decision tree algorithm to handle classification or regression predictive models [24]. For the basic prediction technique, the K-Nearest Neighbor (KNN) algorithm was applied. KNN is a vital classification algorithm in the supervised ML domain and is broadly applied in pattern recognition, data mining and intrusion detection [25]. For the kernel-based model, the Support Vector Machine (SVM) was selected. SVM is a supervised ML model that employs classification algorithms for two-group categorization [26]. Gaussian Naive Bayesian (GNB) algorithm was included in the linear model and is specifically used when the features manifest continuous values [27]. Multilayer Perceptron (MLP) is a feed-forward neural network supplement and has been widely applied in various prediction models [28]. After employing the Bootstrap aggregating (Bagging) algorithm to optimize the performance of established models, stacked regression was utilized to obtain a stacking model via integrating7 models to output a superior outcome [29, 30].

To facilitate the model performance and retain the maximum authenticity of data, we strictly employed the Synthetic Minority Over-sampling technique in the inner training dataset [31]. To begin with, patients in the SEER database were randomly assigned to the training set (80%) and testing set (20%) while the proportion of LM (+) (patients with LM) and LM (-) (patients without LM) groups was nearly identical. In the training set, k-fold cross validation (k = 10) was performed, and grid search was adopted to find out the best combination of parameters. For each set of parameters, the model was in turn fitted and validated with 8/10 and 2/10 of data, respectively. Subsequently, our T1 CRC cohort in the Chinese population was utilized as an extra outer validation set further to examine both the applicability and efficiency of the model. The overall workflow is exhibited in Fig. 1.

Assessment of Model Performance

To ensure rational comparison of the models, confusion matrix, the area under the curve (AUC), sensitivity, specificity, precision, negative predictive value (NPV), false discovery rate (FDR), accuracy, and average precision (AP) were applied as indicators for
assessing model performance. In addition, the area under receiver operating characteristic curves (AU-ROC) was utilized as a performance index while the AP value was employed as the criterion for the precision-recall (PR) curve [32]. The average value of parameters was ultimately executed on the testing set and additional outer validation one. Survival analysis was further adopted in the model to evaluate its capability of predicting outcomes of CRC patients.

Statistical Analysis

SEER*Stat software (8.3.6 version) was adopted to acquire targeted CRC patients from the SEER database. Python (version 3.6.9) and R software (version 4.0.5) were utilized to perform statistical analyzes. Demographic differences between the two groups were tested using either Student’s t-test or Pearson chi-square test. Results were considered statistically significant when \( P \leq 0.05 \).

Results

Case Structures and Clinical Baselines

The initial LM data was included in 2010 and the latest one was updated in 2016 in the SEER database. In the current study, 262,285 CRC patients from 2010 to 2016 were included. According to the above inclusive and exclusive criteria, a total of 16785 patients were ultimately enrolled in the inner dataset while 326 out of 8,226 CRC patients in Xijing hospital were recruited. The data of these 326 patients was further normalized via SEER database standard. Baselines of the inner training set, inner testing set, and outer validating set were exhibited in Table 1.

Eleven independent clinical factors were included in the model, consisting of age at diagnosis, gender, marital status at diagnosis, primary site, tumor size, tumor grade, tumor type, N stage, CEA level, tumor deposits, and PNI (Table 2). Patients from SEER database were categorized into LM (-) group (16,023 patients without LM, 95.5%) and LM (+) (762 patients with LM, 4.5%) group respectively. In LM (+) patients, the age at diagnosis is mostly ranged from 40 to 90 (721/762, 94.6%). Besides, the proportion of diagnosed age less than 60 years in LM (+) group (333/762; 43.7%) is significantly surpassed the LM (-) group (6553/16,023; 40.9%; \( P< 0.001 \)). The proportion of male with T1 CRC is significantly higher in LM (+) group compared with LM (-) one (\( P = 0.001 \)), while race demonstrated no statistical difference between the two groups. Intriguingly, a higher occurrence rate was observed in the single (167/2611, 6.4%) than the married (376/8918, 4.2%; \( P<0.001 \)). The rectum is the most common primary site in both groups, and its proportion is comparatively higher in the T1 stage than other T stages in all CRC patients (\( P< 0.001 \)). Average tumor size of LM (+) group (mean = 52.1mm) was considerably larger than that of LM (-) one (mean = 17.5mm; \( P< 0.001 \)). LM (+) group portended a dramatically higher proportion of Grade II-IV than LM (-) group (92.8% vs 68%; \( P<0.001 \)). Similarly, T1 CRC patients with LM tend to have advanced N stage (\( P<0.001 \)). Adenocarcinoma (Adenocarcinoma, NOS, Adenocarcinoma in tubulovillous adenoma, and Adenocarcinoma in adenomatous polyp; 12714/16785, 75.7%) is the most common neoplastic category among all patients. Furthermore, we observed a significantly higher level of positive CEA, more tumor deposits and more PNI in LM (+) group than LM (-) one (\( P< 0.001 \)). Additionally, the baselines of SEER training, SEER testing and our outer validating sets were exhibited in Table 2.

Parameters tuning in our models

We trained the LGBM with a depth of five, a learning rate of 0.01, basic learners of 240, leaves of 16, and max bins of 128. For RF and CART, we also elected 5 as the max depth of the basic trees. The number of neighbors 200 for KNN is the best. In MLP, we ultimately selected a learning rate of 0.01, epochs of 300, hidden layer of 1, and employed the Adam Optimizer and ReLU activation function. For SVM, a combination of a C value of 0.01 and kernel smoothing parameters of 0.0001 was determined. Lastly, every Bagging model, which owns 10 basic models, was trained with identical algorithms but different data. The ultimate stacking model consists of seven bagging models, which outputs probability and a GNB as meta classifier.

Evaluation of Models

To better evaluate the performance of our constructed models, ROC curves and PR curves during the model training were plotted. Via internal verifying, all models were observed to have superior predictive abilities (AUC values > 0.94). And, by
incorporating seven other single models, the stacking model demonstrated an ultimate AUC of up to 0.9631 (Figure 2A). Except for GNB models, AP values of nearly all models attain relatively preferable levels. Noticeably, the ultimate AP of the stacking mode reached 0.693 (Figure 2B). Intriguingly, the external validation set demonstrated more desirable performance. All models have exhibited dramatically high predictive value except the MLP model, and the stacking model contains a final AUC value of 0.992 and an ultimate AP value of 0.811 (Figure 2C, D).

Additionally, via employing the confusion matrix to evaluate the value of models, predictive outcomes of both the inner testing set and outer validation set were shown in Table 3. LGBM produced fewer quantities of FN (False Negative) and FP (False Positive) than other models in both testing sets. The stacking model was capable of screening approximately all LM (+) patients in both sets. Detailed values of AUC, sensitivity, specificity, precision, NPV, FDR, accuracy, AP, F1-values, and Matthews correlation coefficient of each model in inner and outer validation sets were listed respectively in Table 4 and Table 5. The accuracy of 5 single models reached 0.95, among which LGBM displayed the highest accuracy (0.9657). The specificity of MLP and sensitivity of GNB were the highest among seven single models. Generally speaking, the stacking model demonstrated the most satisfying AUC and sensitivity, indicating that this model has clinical value for early screening of LM, excellent precision, NPV, FDR, accuracy, AP score, F1 score, and Matthews correlation coefficient value in CRC patients.

Furthermore, employing survival status and time from the SEER database, we plotted the Kaplan Meier (K-M) curves of the testing set. It is universally acknowledged that LM is an unfavorable prognostic indicator for CRC patients (Figure 3A). Likewise, we found that the stacking model resembled LM in predicting T1 CRC patients’ outcomes (Figure 3B).

Comparison of Significance of Each Factor

In all single models, tumor size, preoperative CEA levels, tumor deposits, N stage, histology, and PNI played a vital role in predicting for LM in T1 CRC. Even though the AI model manifested desirable performance, the individualized influence of each factor on the result and underlying relationships between these factors remain unknown. Hence, we calculated and digitized the significance of each factor used in the built-up AI models (Figure 4). We found that tumor size, CEA level prior to surgery, tumor deposits, and N stage were the top four crucial predictors among all models. Noticeably, tumor size was the most critical one in nearly all models.

Discussion

The liver is identified as one of the most common metastasis sites and LM is recognized as the most lethal factor of CRC patients [33, 34]. Early diagnosis of LM could assist clinicians in taking active intervention timely to improve the prognosis of patients, especially for CRC T1 patients [35, 36]. CRC patients in T1 stage could either choose surgical or endoscopic treatment, partly depending on the status of distant metastasis. Therefore, a convenient and accurate predictive model of LM is urgently demanded to offer guidance on personalized therapeutic strategies and evaluation of 5-year OS.

In the study, we established a new and convenient model to predict early LM by incorporating 11 clinicopathologic parameters in T1 CRC using seven AI methods. Our findings indicated that age, gender, married status, primary site, tumor size, CEA, tumor type, grade, N stage, and PNI were critical factors in the prediction of LM in the AI models. We firstly combined our real-world researches with public data online on a large scale to comprehensively construct and assess LM predictive models in T1 CRC. Given that the AUC of these models was more extensive than 0.94 and model accuracy was approximate as 100 % as possible, we concluded that the above-established models are ideal and robust in yielding clinical benefit, which might aid clinicians to select potential LM CRC patients efficiently.

Our real-world research incorporated 326 cases of T1 CRC, among which LM occurred in merely eight patients (8/326), significantly lower than that of the SEER database (762/16785, P < 0.001). The discrepancy in the LM ratio might be attributed to low diagnostic efficacy in developing countries [37, 38]. Interestingly, compared with more advanced T stage CRC patients (169/326), PNI was more commonly seen in T1 CRC patients of our hospital (1266/8226), consistent with the SEER database (11350/16785, P < 0.001). Abundant evidence has demonstrated that the percentage of PNI occurring in all T stages is approximately 10–15%. Moreover, PNI is an independent biomarker that could indicate aggressive behavior and unfavorable
prognosis of CRC [39–42]. Nonetheless, little literature has explained the underlying reasons behind the high ratio of PNI in T1 CRC, which deserves further investigation. In addition, serum CEA was confirmed to have a positive relationship with LM. Accumulating evidence has suggested that the expression level of CEA could function as an independent indicator for the prognosis of CRC patients [43]. Therefore, it was not surprising that the concentration of preoperative plasma CEA was significantly higher in CRC patients with LM compared with those with primary CRC [44–46]. Besides, among all indicators, tumor size is regarded as one of the most important in predicting LM status. It has been reported that tumor size is intimately associated with both lymph and hepatic metastases of CRC [47]. Furthermore, scientists have verified that age might play a nonnegligible role in the advancement and prognosis of CRC [48]. Despite increment in young CRC patients, it has been reported that the young tend to have more favorable outcomes than the old [48]. Contradictorily, we indicated that CRC patients younger than 60-year-old were more apt to experience risk of LM than counterparts, which is consistent with several researchers [49–51]. Potential reasons might be relevant to commonly frequent occurrence of mismatch repair gene mutation and more aggressive tumor biology in younger patients [52].

To date, multitudes of researchers have constructed practical models to predict the metastatic capability of CRC. For instance, MS Tang [12] et al. have built up a novel nomogram to predict LM in all T stages CRC patients by using multivariable Cox regression. They also found that synchronous LM was an independent prognostic factor for CRC patients [12]. Likewise, Ji Hyun Ahn [17] et al. have developed an innovative model to predict LNM in the early stage of CRC patients via utilizing the SEER database and adopting seven AI methods. Nevertheless, these studies were retrospective, single-center, and with small quantities of patients. Besides, acquired data are limited due to the low incidence of LM in early CRC. With the recent technical advancement of AI, the application of ML model in neoplastic diagnosis and prognostic assessment has become increasingly prevalent [53, 54]. Ichimasa et al. [55] have demonstrated that AI could reduce unnecessary surgery after endoscopic resection of LNM (-) T1 CRC compared with current guidelines. Nonetheless, few models for predicting the incidence of LM in T1 CRC patients were developed and assessed utilizing AI methods. In the current study, we established nine models and validated them in our own dataset. Besides, their efficacy of predicting LM in early CRC was also compared via using easily available clinical and histopathological features. Furthermore, we found that our constructed AI models could not only assist clinicians in selecting patients with a high risk of LM, but also resemble LM in the accurate prediction of T1 CRC patients’ outcomes.

This study still has several limitations and weaknesses. Firstly, in light that the SEER database is an open and available national program of America, these newly established models might not be perfectly applied in other countries. Secondly, quantities of enrolled patients in our hospital were far from sufficient, and merely eight patients manifested LM status. These shortcomings might lead to a limited verification outcome. In the future, more in-depth and extensive studies are urgently needed.

**Conclusions**

In the present study, we established an innovative and stacking bagging model which incorporates 11 clinicopathologic features to predict the incidence of LM in T1 CRC. Our findings indicated that age, gender, married status, primary site, tumor size, CEA, tumor type, grade, N stage and PNI were crucial factors for predicting LM, among which tumor size matters most. As expected, the stacking bagging model, which integrated strengths of seven single models, demonstrated the strongest predictive power in both databases of SEER and our hospital. In addition, we recruited 326 T1 CRC patients from Xijing hospital to verify both predictive power and identification capability of the stacking model. Corresponding results indicated that the Bagging model could successfully identify 8 LM patients in our hospital. Moreover, we suggest that the stacking model resembles LM in the accurate prediction of T1 CRC patients’ outcomes.

**Abbreviations**

CRC: colorectal cancer; LM: liver metastasis; OS: overall survival; AI: artificial intelligence; ML: machine learning; SEER: Surveillance, Epidemiology, and End Results; LNM: lymph node metastasis; CEA: carcinoembryonic antigen; PNI: perineural invasion; LGBM: Light Gradient Boosting Decision; RF: Random Forest; CART: Classification and Regression Trees; KNN: K-Nearest Neighbor; SVM: Support Vector Machine, GNB: Gaussian Naive Bayesian; MLP: Multilayer Perceptron; Bagging:
Declarations

Availability of data and materials
The datasets used and/or analyzed during the current study are included in this published article and its additional files.

Acknowledgements
Not applicable.

Funding
Not applicable.

Author information

Affiliations
Xijing Hospital, Airforce Medical University, 710032, Xi’an, China
Tenghui Han
State Key Laboratory of Cancer Biology, Institute of Digestive Diseases, Xijing Hospital, Airforce Medical University, 710032, Xi’an, China
Jun Zhu, Rujie Chen
Department of general surgery, The Southern Theater Air Force Hospital, 510062, Guangzhou, China
Xiaoping Chen, Jun Zhu
School of Clinical Medicine, Xi’an Medical University, 710032, Xi’an, China
Dong Xu
Ming gang station hospital, Xi’an Institute of flight of the air force, 464094, Minggang, China
Shuai Wang
Division of Digestive Surgery, Xijing Hospital of Digestive Diseases, Airforce Medical University, 710032, Xi’an, China
Jianyong Zheng, Chunsheng Xu

Contributions
C Xu, J Zheng & X Chen designed the study; T Han, J Zhu & D Xu contributed to the conception of the study and completed the manuscript together; R Chen contributed significantly to statistical analysis and manuscript preparation; S Wang helped perform the analysis with constructive discussions. All authors read and approved the final manuscript.

Corresponding author
Correspondence to Chunsheng Xu, Jianyong Zheng and Xiaoping Chen.
Ethics declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Xijing Hospital of Airforce Medical University. All participants gave written, informed consent.

Consent for publication

Not applicable

Conflicts of interest

The authors have no conflicts of interest to declare.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a cancer journal for clinicians 2021, 71(3):209-249.

2. Bray F, Soerjomataram I: The Changing Global Burden of Cancer: Transitions in Human Development and Implications for Cancer Prevention and Control. In: Cancer: Disease Control Priorities, Third Edition (Volume 3). edn. Edited by Gelband H, Jha P, Sankaranarayanan R, Horton S. Washington (DC): The International Bank for Reconstruction and Development / The World Bank © 2015 International Bank for Reconstruction and Development / The World Bank.; 2015.

3. Arnold M, Abnet CC, Neale RE, Vignat J, Giovannucci EL, McGlynn KA, Bray F: Global Burden of 5 Major Types of Gastrointestinal Cancer. Gastroenterology 2020, 159(1):335-349.e315.

4. Kow AWC: Hepatic metastasis from colorectal cancer. Journal of gastrointestinal oncology 2019, 10(6):1274-1298.

5. Helling TS, Martin M: Cause of death from liver metastases in colorectal cancer. Annals of surgical oncology 2014, 21(2):501-506.

6. Kopetz S, Chang GJ, Overman MJ, Eng C, Sargent DJ, Larson DW, Grothey A, Vauthey JN, Nagorney DM, McWilliams RR: Improved survival in metastatic colorectal cancer is associated with adoption of hepatic resection and improved chemotherapy. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2009, 27(22):3677-3683.

7. Chakedis J, Schmidt CR: Surgical Treatment of Metastatic Colorectal Cancer. Surgical oncology clinics of North America 2018, 27(2):377-399.

8. Giannis D, Sideris G, Kakos CD, Katsaros I, Zilogas IA: The role of liver transplantation for colorectal liver metastases: A systematic review and pooled analysis. Transplantation reviews (Orlando, Fla) 2020, 34(4):100570.

9. Arru M, Aldrighetti L, Castoldi R, Di Palo S, Orsenigo E, Stella M, Pulitano C, Gavazzi F, Ferla G, Di Carlo V et al: Analysis of prognostic factors influencing long-term survival after hepatic resection for metastatic colorectal cancer. World journal of surgery 2008, 32(1):93-103.

10. Xu H, Wang C, Song H, Xu Y, Ji G: RNA-Seq profiling of circular RNAs in human colorectal Cancer liver metastasis and the potential biomarkers. Molecular cancer 2019, 18(1):8.

11. Li H, Dai W, Xia X, Wang R, Zhao J, Han L, Mo S, Xiang W, Du L, Zhu G et al: Modeling tumor development and metastasis using paired organoids derived from patients with colorectal cancer liver metastases. Journal of hematology & oncology 2020, 13(1):119.

12. Tang M, Wang H, Cao Y, Zeng Z, Shan X, Wang L: Nomogram for predicting occurrence and prognosis of liver metastasis in colorectal cancer: a population-based study. International journal of colorectal disease 2021, 36(2):271-282.

13. Topol EJ: High-performance medicine: the convergence of human and artificial intelligence. Nature medicine 2019, 25(1):44-56.
14. Hamet P, Tremblay J: Artificial intelligence in medicine. *Metabolism: clinical and experimental* 2017, 69:S36-s40.

15. Iqbal MJ, Javed Z, Sadia H, Qureshi IA, Irshad A, Ahmed R, Malik K, Raza S, Abbas A, Pezzani R et al. Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. *Cancer cell international* 2021, 21(1):270.

16. Wang Y, He X, Nie H, Zhou J, Cao P, Ou C: Application of artificial intelligence to the diagnosis and therapy of colorectal cancer. *American journal of cancer research* 2020, 10(11):3575-3598.

17. Ahn JH, Kwak MS, Lee HH, Cha JM, Shin HP, Jeon JW, Yoon JY: Development of a Novel Prognostic Model for Predicting Lymph Node Metastasis in Early Colorectal Cancer: Analysis Based on the Surveillance, Epidemiology, and End Results Database. *Frontiers in oncology* 2021, 11:614398.

18. Kudo SE, Ichimasa K, Villard B, Mori Y, Misawa M, Saito S, Hotta K, Saito Y, Matsuda T, Yamada K et al. Artificial Intelligence System to Determine Risk of T1 Colorectal Cancer Metastasis to Lymph Node. *Gastroenterology* 2021, 160(4):1075-1084.e1072.

19. Surveillance Epidemiology and End Results (SEER) Program, Research Data (National Cancer Institute, DCCPS Surveillance Research Program, Surveillance Systems Branch) [www.seer.cancer.gov]

20. Daly MC, Paquette IM: Surveillance, Epidemiology, and End Results (SEER) and SEER-Medicare Databases: Use In Clinical Research for Improving Colorectal Cancer Outcomes. *Clinics in colon and rectal surgery* 2019, 32(1):61-68.

21. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y: Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 2017, 30:3146-3154.

22. Létinier L, Jouganous J, Benkebil M, Bel-Létoile A, Goehrs C, Singier A, Rouby F, Lacroix C, Miremont G, Micallef J et al. Artificial Intelligence for Unstructured Healthcare Data: Application to Coding of Patient Reporting of Adverse Drug Reactions. *Clinical pharmacology and therapeutics* 2021.

23. Breiman L: *RANDOM FORESTS–RANDOM FEATURES*. *machine learning* 1999.

24. Fearn, Tom: *Classification and regression trees (CART)*. *Journal of Near Infrared Spectroscopy* 2006, 17(1):13.

25. Keller JM, Gray MR, Givens JA: A fuzzy K-nearest neighbor algorithm. *IEEE Transactions on Systems Man & Cybernetics* 2012, SMC-15(4).

26. Joachims T: Text categorization with Support Vector Machines: Learning with many relevant features. In: *Proc Conference on Machine Learning*: 1998; 1998.

27. Chickering DM, Heckerman D: Efficient Approximations for the Marginal Likelihood of Bayesian Networks with Hidden Variables. *Machine Learning* 1997, 29(2):181-212.

28. Ruck DW: Feature Selection Using a Multilayer Perceptron. *Neural Network Comput* 1990, 2:40–48.

29. Leo, Breiman: Stacked Regressions. *Machine Learning* 1996.

30. Breiman L: Bagging prediction. *Machine Learning* 1996, 24.

31. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP: SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 2002, 16(1):321-357.

32. Davis JJ, Goadrich MH: The relationship between Precision-Recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*: 2006; 2006.

33. Engstrand J, Nilsson H, Strömberg C, Jonas E, Freedman J: Colorectal cancer liver metastases - a population-based study on incidence, management and survival. *BMC Cancer* 2018, 18(1):78.

34. van der Geest LG, Lam-Boer J, Koopman M, Verhoef C, Elferink MA, de Wilt JH: Nationwide trends in incidence, treatment and survival of colorectal cancer patients with synchronous metastases. *Clin Exp Metastasis* 2015, 32(5):457-465.

35. Yin J, Bai Z, Song J, Yang Y, Wang J, Han W, Zhang J, Meng H, Ma X, Yang Y et al: Differential expression of serum miR-126, miR-141 and miR-21 as novel biomarkers for early detection of liver metastasis in colorectal cancer. *Chin J Cancer Res* 2014, 26(1):95-103.

36. Lv Y, Feng QY, Wei Y, Ren L, Ye Q, Wang X, Cui Y, Liu T, Zhou B, Wang M et al: Benefits of multi-disciplinary treatment strategy on survival of patients with colorectal cancer liver metastasis. *Clin Transl Med* 2020, 10(3):e121.
37. Yao T, Shiono S: Differences in the pathological diagnosis of colorectal neoplasia between the East and the West: Present status and future perspectives from Japan. Dig Endosc 2016, 28(3):306-311.
38. Schlemper RJ, Itabashi M, Kato Y, Lewin KJ, Riddell RH, Shimoda T, Sipponen P, Stolte M, Watanabe H: Differences in the diagnostic criteria used by Japanese and Western pathologists to diagnose colorectal carcinoma. Cancer 1998, 82(1):60-69.
39. Alotaibi AM, Lee JL, Kim J, Lim SB, Yu CS, Kim TW, Kim JH, Kim JC: Prognostic and Oncologic Significance of Perineural Invasion in Sporadic Colorectal Cancer. Ann Surg Oncol 2017, 24(6):1626-1634.
40. Al-Sukhni E, Attwood K, Gabriel EM, LeVea CM, Kanehira K, Nurkin SJ: Lymphovascular and perineural invasion are associated with poor prognostic features and outcomes in colorectal cancer: A retrospective cohort study. Int J Surg 2017, 40:37-49.
41. Yang Y, Huang X, Sun J, Gao P, Song Y, Chen X, Zhao J, Wang Z: Prognostic value of perineural invasion in colorectal cancer: a meta-analysis. J Gastrointest Surg 2015, 19(6):1113-1122.
42. Knijn N, Mogk SC, Teerenstra S, Simmer F, Nagtegaal ID: Perineural Invasion is a Strong Prognostic Factor in Colorectal Cancer: A Systematic Review. Am J Surg Pathol 2016, 40(1):103-112.
43. Zhu J, Hao J, Ma Q, Shi T, Wang S, Yan J, Chen R, Xu D, Jiang Y, Zhang J et al: A Novel Prognostic Model and Practical Nomogram for Predicting the Outcomes of Colorectal Cancer: Based on Tumor Biomarkers and Log Odds of Positive Lymph Node Scheme. Front Oncol 2021, 11:661040.
44. Pakdel A, Malekzadeh M, Naghibalhossaini F: The association between preoperative serum CEA concentrations and synchronous liver metastasis in colorectal cancer patients. Cancer Biomark 2016, 16(2):245-252.
45. Polivka J, Windrichova J, Pesta M, Houfkova K, Rezackova H, Macanova T, Vycital O, Kucera R, Slouka D, Topolcan O: The Level of Preoperative Plasma KRAS Mutations and CEA Predict Survival of Patients Undergoing Surgery for Colorectal Cancer Liver Metastases. Cancers (Basel) 2020, 12.
46. Lou Z, Meng RG, Zhang W, Yu ED, Fu CG: Preoperative carcinoembryonic antibody is predictive of distant metastasis in pathologically T1 colorectal cancer after radical surgery. World J Gastroenterol 2013, 19(3):389-393.
47. Guo K, Feng Y, Yuan L, Wasan HS, Sun L, Shen M, Ruan S: Risk factors and predictors of lymph nodes metastasis and distant metastasis in newly diagnosed T1 colorectal cancer. Cancer Med 2020, 9(14):5095-5113.
48. Abasse Kassim S, Tang W, Abbas M, Wu S, Meng Q, Zhang C, Li X, Chen R: Clinicopathologic and epidemiological characteristics of prognostic factors in post-surgical survival of colorectal cancer patients in Jiangsu Province, China. Cancer Epidemiol 2019, 62:101565.
49. Mo S, Cai X, Zhou Z, Li Y, Hu X, Ma X, Zhang L, Cai S, Peng J: Nomograms for predicting specific distant metastatic sites and overall survival of colorectal cancer patients: A large population-based real-world study. Clin Transl Med 2020, 10(1):169-181.
50. Luo D, Liu Q, Yu W, Ma Y, Zhu J, Lian P, Cai S, Li Q, Li X: Prognostic value of distant metastasis sites and surgery in stage IV colorectal cancer: a population-based study. Int J Colorectal Dis 2018, 33(9):1241-1249.
51. Tohmé C, Labaki M, Hajj G, Abboud B, Noun R, Sarkis R: [Colorectal cancer in young patients: presentation, clinicopathological characteristics and outcome]. Le Journal medical libanais The Lebanese medical journal 2008, 56(4):208-214.
52. Law JH, Koh FH, Tan KK: Young colorectal cancer patients often present too late. Int J Colorectal Dis 2017, 32(8):1165-1169.
53. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI: Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015, 13:8-17.
54. Xiao Y, Wu J, Lin Z, Zhao X: A deep learning-based multi-model ensemble method for cancer prediction. Comput Methods Programs Biomed 2018, 153:1-9.
55. Ichimasa K, Kudo SE, Mori Y, Misawa M, Matsudaira S, Kouyama Y, Baba T, Hidaka E, Wakamura K, Hayashi T et al: Artificial intelligence may help in predicting the need for additional surgery after endoscopic resection of T1 colorectal...
cancer. *Endoscopy* 2018, 50(3):230-240.

**Tables**

Table 1 Clinical baseline features of SEER and Xijing hospital database.
| Variables                                      | SEER database | Xijing CRC cohort |
|-----------------------------------------------|---------------|-------------------|
|                                               | Training set  | Testing set       | Outer validation set |
| Age at diagnosis, n (%)                        |               |                   |                      |
| 0-9                                           | 14 (0.1)      | 0(0)              | 0(0)                 |
| 10-19                                         | 128 (1.0)     | 31 (0.9)          | 0(0)                 |
| 20-29                                         | 265 (2.0)     | 67 (2.0)          | 4 (1.2)              |
| 30-39                                         | 390 (2.9)     | 79 (2.4)          | 5 (1.5)              |
| 40-49                                         | 1084 (8.1)    | 251 (7.5)         | 35 (10.7)            |
| 50-59                                         | 3632 (27.0)   | 945 (28.2)        | 104 (31.9)           |
| 60-69                                         | 3649 (27.2)   | 911 (27.1)        | 92 (28.2)            |
| 70-79                                         | 2659 (19.8)   | 670 (20.0)        | 65 (19.9)            |
| 80-89                                         | 1403 (10.4)   | 354 (10.5)        | 21 (6.4)             |
| 90-99                                         | 204 (1.5)     | 49 (1.5)          | 0(0)                 |
| Gender, n (%)                                 |               |                   |                      |
| Female                                        | 6982 (52.0)   | 1695 (50.5)       | 189 (58.0)           |
| Male                                          | 6446 (48.0)   | 1662 (49.5)       | 137 (42.0)           |
| Race, n (%)                                   |               |                   |                      |
| White                                         | 10226 (76.2)  | 2552 (76.0)       | 0(0)                 |
| Black                                         | 1754 (13.1)   | 466 (13.9)        | 0(0)                 |
| Asian or Pacific Islander                     | 1354 (10.1)   | 319 (9.5)         | 326 (100.0)          |
| American Indian/Alaska Native                 | 94 (0.7)      | 20 (0.6)          | 0(0)                 |
| Marital status at diagnosis, n (%)            |               |                   |                      |
| Married and separated                         | 7615 (56.7)   | 1855 (55.2)       | 322 (98.8)           |
| Divorced                                      | 1207 (9.0)    | 293 (8.7)         | 2 (0.6)              |
| Unmarried                                     | 2219 (16.5)   | 559 (16.7)        | 2 (0.6)              |
| Other                                         | 2387 (17.8)   | 650 (19.3)        | 0(0)                 |
| LM, n (%)                                     |               |                   |                      |
| Yes                                           | 12821 (95.5)  | 3202 (95.4)       | 318 (97.5)           |
| No                                            | 607 (4.5)     | 155 (4.6)         | 8 (2.5)              |
| Primary site, n (%)                           |               |                   |                      |
| Rectum, NOS                                   | 3786 (28.2)   | 955 (28.4)        | 228 (69.9)           |
| Sigmoid colon                                 | 2925 (21.8)   | 777 (23.1)        | 35 (10.7)            |
| Ascending colon                               | 1646 (12.3)   | 413 (12.3)        | 24 (7.4)             |
| Cecum                                         | 1586 (11.8)   | 393 (11.7)        | 6 (1.8)              |
| Appendix                                      | 868 (6.5)     | 216 (6.4)         | 0(0)                 |
| Location                          | Tumor size, mm, mean (SD) | Tumor grade, n (%) | Tumor type, n (%) |
|----------------------------------|---------------------------|--------------------|-------------------|
| Rectosigmoid junction            | 846 (6.3) 215 (6.4) 7 (2.1) | Well differentiated; Grade I 4171 (31.1) 1015 (30.2) 69 (21.2) | Adenocarcinoma, NOS 4368 (32.5) 1099 (32.7) 91 (27.9) |
| Transverse colon                 | 723 (5.4) 166 (4.9) 9 (2.8) | Moderately differentiated; Grade II 8306 (61.9) 2114 (63.0) 240 (73.6) | Adenocarcinoma in tubulovillous adenoma 2969 (22.1) 743 (22.1) 76 (23.3) |
| Descending colon                 | 481 (3.6) 106 (3.2) 1 (0.3) | Poorly differentiated; Grade III 827 (6.2) 191 (5.7) 15 (4.6) | Adenocarcinoma in adenomatous polyp 2827 (21.1) 708 (21.1) 125 (38.3) |
| Hepatic flexure of colon         | 303 (2.3) 70 (2.1) 7 (2.1) | Undifferentiated; anaplastic; Grade IV 124 (0.9) 37 (1.1) 2 (0.6) | Carcinoid tumor, NOS 1837 (13.7) 454 (13.5) 0 (0) |
| Splenic flexure of colon         | 172 (1.3) 31 (0.9) 3 (0.9) | | Adenocarcinoma in villous adenoma 483 (3.6) 126 (3.8) 9 (2.8) |
| Colon, NOS                       | 50 (0.4) 9 (0.3) 4 (1.2) | | Neuroendocrine carcinoma, NOS 409 (3.0) 93 (2.8) 0 (0) |
| Overlapping lesion of colon      | 42 (0.3) 6 (0.2) 2 (0.6) | | Mucinous adenocarcinoma 238 (1.8) 61 (1.8) 7 (2.1) |
| Tumor size, mm, mean (SD)        | 19.16 (25.1) 18.82 (22.3) 24.6 (14.0) | | Squamous cell carcinoma, NOS 52 (0.4) 8 (0.2) 0 (0) |
| Tumor grade, n (%)               |                           | | Atypical carcinoid tumor 38 (0.3) 11 (0.3) 0 (0) |
| Well differentiated; Grade I     |                           | | Signet ring cell carcinoma 28 (0.2) 6 (0.2) 0 (0) |
| Moderately differentiated; Grade II |                      | | Mucin-producing adenocarcinoma 26 (0.2) 6 (0.2) 0 (0) |
| Poorly differentiated; Grade III |                           | | Tubular adenocarcinoma 22 (0.2) 8 (0.2) 18 (5.5) |
| Undifferentiated; anaplastic; Grade IV |                  | | Gastrointestinal stromal sarcoma 17 (0.1) 0 (0) 0 (0) |
| Tumor type, n (%)                |                           | | Carcinoma, NOS 14 (0.1) 5 (0.1) 0 (0) |
| Adenocarcinoma, NOS              | 4368 (32.5) 1099 (32.7) 91 (27.9) | | Villous adenocarcinoma 10 (0.1) 2 (0.1) 0 (0) |
| Adenocarcinoma in tubulovillous adenoma | 2969 (22.1) 743 (22.1) 76 (23.3) | | Other 90 (0.7) 27 (0.8) 0 (0) |
| Adenocarcinoma in adenomatous polyp | 2827 (21.1) 708 (21.1) 125 (38.3) | | | |
| Carcinoid tumor, NOS             | 1837 (13.7) 454 (13.5) 0 (0) | | | |
| Adenocarcinoma in villous adenoma | 483 (3.6) 126 (3.8) 9 (2.8) | | | |
| Neuroendocrine carcinoma, NOS    | 409 (3.0) 93 (2.8) 0 (0) | | | |
| Mucinous adenocarcinoma          | 238 (1.8) 61 (1.8) 7 (2.1) | | | |
| Squamous cell carcinoma, NOS     | 52 (0.4) 8 (0.2) 0 (0) | | | |
| Atypical carcinoid tumor         | 38 (0.3) 11 (0.3) 0 (0) | | | |
| Signet ring cell carcinoma       | 28 (0.2) 6 (0.2) 0 (0) | | | |
| Mucin-producing adenocarcinoma   | 26 (0.2) 6 (0.2) 0 (0) | | | |
| Tubular adenocarcinoma           | 22 (0.2) 8 (0.2) 18 (5.5) | | | |
| Gastrointestinal stromal sarcoma | 17 (0.1) 0 (0) 0 (0) | | | |
| Carcinoma, NOS                   | 14 (0.1) 5 (0.1) 0 (0) | | | |
| Villous adenocarcinoma           | 10 (0.1) 2 (0.1) 0 (0) | | | |
| Other                            | 90 (0.7) 27 (0.8) 0 (0) | | | |
| N, n (%)                         |                           | | | |
| N0                               | 12142 (90.4) 3031 (90.3) 295 (90.5) | | | |
| N1                               | 1150 (8.6) 296 (8.82) 30 (9.2) | | | |
| N2                               | 136 (1.0) 30 (0.9) 1 (0.3) | | | |
| CEA, n (%)                       |                           | | | |
| Positive                         | 1223 (9.1) 300 (8.9) 110 (33.7) | | | |
|                        | Borderline | Tumor deposits, n (%) | Perineural invasion, n (%) |
|------------------------|------------|-----------------------|----------------------------|
|                        |            | No tumor deposits     | Yes                        |
|                        |            | Tumor Deposits identified | 95 (0.7) 27 (0.8) 1 (0.3) |
|                        |            | Unknown               | 4556 (33.9) 1117 (33.3) 0 (0) |
|                        |            | No                    | 9104 (67.8) 2246 (66.9) 169 (51.8) |
|                        |            | No                    | 105 (0.8) 48 (1.4) 157 (48.2) |
|                        |            | Unknown               | 4219 (31.4) 1063 (31.7) 0 (0) |

SEER, Surveillance, Epidemiology, and End Results; CRC, colorectal cancer; LM, liver metastasis; SD, standard deviation; CEA, carcinoembryonic antigen.

Table 2 Distributions of clinicopathological characteristics in two groups.
| Variables                      | LM (-) | LM (+) | P value |
|-------------------------------|--------|--------|---------|
| N                             | 16023  | 762    |         |
| Age at diagnosis, n (%)       |        |        | <0.001  |
| 0-9                           | 14 (0.1)| 0 (0.0)|         |
| 10-19                         | 158 (1.0)| 1 (0.1)|         |
| 20-29                         | 324 (2.0)| 8 (1.0)|         |
| 30-39                         | 447 (2.8)| 22 (2.9)|     |
| 40-49                         | 1238 (7.7)| 97 (12.7)|   |
| 50-59                         | 4372 (27.3)| 205 (26.9)|   |
| 60-69                         | 4363 (27.2)| 197 (25.9)|   |
| 70-79                         | 3185 (19.9)| 144 (18.9)|   |
| 80-89                         | 1679 (10.5)| 78 (10.2)|     |
| 90-99                         | 243 (1.5)| 10 (1.3)|      |
| Gender, n (%)                 |        |        |         |
| Female                        | 7784 (48.6)| 324 (42.5)| 0.001  |
| Male                          | 8239 (51.4)| 438 (57.5)|       |
| Race, n (%)                   |        |        | 0.215   |
| White                         | 12213 (76.2)| 565 (74.1)|   |
| Black                         | 2100 (13.1)| 120 (15.7)|     |
| Asian or Pacific Islander     | 1601 (10.0)| 72 (9.4)|        |
| American Indian/Alaska Native| 109 (0.7)| 5 (0.7)|         |
| Marital status at diagnosis, n (%) |        |        | <0.001  |
| Married                       | 8918 (55.7)| 376 (49.3)|   |
| Single                        | 2611 (16.3)| 167 (21.9)|     |
| Widowed                       | 1740 (10.9)| 90 (11.8)|       |
| Divorced                      | 1417 (8.8)| 83 (10.9)|        |
| Unknown                       | 1131 (7.1)| 36 (4.7)|         |
| Separated                     | 166 (1.0)| 10 (1.3)|          |
| Unmarried or Domestic Partner | 40 (0.2)| 0 (0.0)|        |
| Primary site, n (%)           |        |        | <0.001  |
| Rectum, NOS                   | 4502 (28.1)| 239 (31.4)|   |
| Sigmoid colon                 | 3540 (22.1)| 162 (21.3)|     |
| Ascending colon               | 1969 (12.3)| 90 (11.8)|         |
| Cecum                         | 1884 (11.8)| 95 (12.5)|       |
| Appendix                      | 1081 (6.7)| 3 (0.4)|          |
| Location                        | N  | SD |
|--------------------------------|----|----|
| Rectosigmoid junction          | 967| 6.0|
| Transverse colon               | 863| 5.4|
| Descending colon               | 569| 3.6|
| Hepatic flexure of colon       | 356| 2.2|
| Splenic flexure of colon       | 194| 1.2|
| Colon, NOS                     | 53 | 0.3|
| Overlapping lesion of colon    | 45 | 0.3|
| Tumor size, mm, mean (SD)      | 17.5| 22.5|
| Tumor grade, n (%)             |    |    |
| Well differentiated; Grade I   | 5131| 32.0|
| Moderately differentiated; Grade II | 9853| 61.5|
| Poorly differentiated; Grade III | 891 | 5.6|
| Undifferentiated; anaplastic; Grade IV | 148| 0.9|
| Tumor type, n (%)              |    |    |
| Adenocarcinoma, NOS            | 4859| 30.3|
| Adenocarcinoma in tubulovillous adenoma | 3669| 22.9|
| Adenocarcinoma in adenomatous polyp | 3495| 21.8|
| Carcinoid tumor, NOS           | 2287| 14.3|
| Adenocarcinoma in villous adenoma | 596 | 3.7|
| Neuroendocrine carcinoma, NOS  | 495 | 3.1|
| Mucinous adenocarcinoma        | 281 | 1.8|
| Squamous cell carcinoma, NOS   | 59  | 0.4|
| Atypical carcinoid tumor       | 49  | 0.3|
| Signet ring cell carcinoma     | 32  | 0.2|
| Mucin-producing adenocarcinoma | 30  | 0.2|
| Tubular adenocarcinoma         | 30  | 0.2|
| Gastrointestinal stromal sarcoma | 17  | 0.1|
| Villous adenocarcinoma         | 12  | 0.1|
| Carcinoma, NOS                 | 11  | 0.1|
| Other                          | 101 | 0.6|
| N, n (%)                       |    |    |
| N0                             | 14711| 91.8|
| N1                             | 1179 | 7.4|
| N2                             | 133  | 0.8|
| CEA, n (%)                     |    |    |
| Positive                       | 999 | 6.2|

| Comparison          | N  | SD |
|---------------------|----|----|
| Rectosigmoid junction | 94 | 12.3|
| Transverse colon    | 26 | 3.4|
| Descending colon    | 18 | 2.4|
| Hepatic flexure of colon | 17 | 2.2|
| Splenic flexure of colon | 9  | 1.2|
| Colon, NOS          | 6  | 0.8|
| Overlapping lesion of colon | 3  | 0.4|

<0.001
|                  | Inner Validation | Outer Validation |
|------------------|------------------|------------------|
|                  | Actual | Prediction | Actual | Prediction |
|                  | LM (-) | LM (+)     | LM (-) | LM (+)     |
| LGBM             | LM (+) | 42        | 113    | 4          | 4          |
|                  | LM (-) | 3123      | 79     | 317        | 1          |
| RF               | LM (+) | 46        | 109    | 3          | 5          |
|                  | LM (-) | 3136      | 66     | 318        | 0          |
| GNB              | LM (+) | 32        | 123    | 0          | 8          |
|                  | LM (-) | 3051      | 151    | 313        | 5          |
| KNN              | LM (+) | 49        | 106    | 4          | 4          |
|                  | LM (-) | 3111      | 91     | 316        | 2          |
| MLP              | LM (+) | 64        | 91     | 5          | 3          |
|                  | LM (-) | 3131      | 71     | 303        | 15         |
| CART             | LM (+) | 41        | 114    | 3          | 5          |
|                  | LM (-) | 3100      | 102    | 313        | 5          |
| SVM              | LM (+) | 35        | 120    | 0          | 8          |
|                  | LM (-) | 3059      | 143    | 293        | 25         |
| Stacking         | LM (+) | 26        | 129    | 0          | 8          |
|                  | LM (-) | 3062      | 140    | 303        | 15         |

LM, liver metastasis; LGBM, Light Gradient Boosting Decision; RF, Random Forest; GNB, Gaussian Naive Bayesian; KNN, K-Nearest Neighbor; MLP, Multilayer Perceptron; CART, Classification and Regression Trees; SVM, Support Vector Machine.
Table 4 Performance of developed models in inner datasets.

| Models | AUC    | Sensitivity | Specificity | Precision | NPV     | FDR     | Accuracy | AP     | F1     | MCC    |
|--------|--------|-------------|-------------|-----------|---------|---------|----------|--------|--------|--------|
| LGBM   | 0.9608 | 0.7677      | 0.9753      | 0.6010    | 0.9886  | 0.3990  | 0.9657   | 0.7150 | 0.6742 | 0.6619 |
| RF     | 0.9589 | 0.7226      | 0.9744      | 0.5773    | 0.9864  | 0.4227  | 0.9628   | 0.7051 | 0.6418 | 0.6268 |
| GNB    | 0.9504 | 0.7935      | 0.9535      | 0.4522    | 0.9896  | 0.5478  | 0.9461   | 0.4981 | 0.5761 | 0.5745 |
| KNN    | 0.9520 | 0.6839      | 0.9719      | 0.5408    | 0.9845  | 0.4592  | 0.9586   | 0.6244 | 0.6040 | 0.5869 |
| MLP    | 0.9443 | 0.6000      | 0.9778      | 0.5671    | 0.9806  | 0.4329  | 0.9604   | 0.5788 | 0.5831 | 0.5625 |
| CART   | 0.9558 | 0.7806      | 0.9685      | 0.5450    | 0.9892  | 0.4550  | 0.9598   | 0.6819 | 0.6419 | 0.6326 |
| SVM    | 0.9422 | 0.7677      | 0.9544      | 0.4491    | 0.9884  | 0.5509  | 0.9458   | 0.5524 | 0.5667 | 0.5620 |
| Stacking | 0.9631 | 0.8452      | 0.9566      | 0.4852    | 0.9922  | 0.5148  | 0.9514   | 0.6927 | 0.6165 | 0.6187 |

AUC, area under curve; NPV, negative predictive value; FDR, false discovery rate; AP, average precision; MCC, matthews correlation coefficient; LGBM, Light Gradient Boosting Decision; RF, Random Forest; GNB, Gaussian Naive Bayesian; KNN, K-Nearest Neighbor; MLP, Multilayer Perceptron; CART, Classification and Regression Trees; SVM, Support Vector Machine.

Table 5 Performance of developed models in our real world dataset.

| Models | AUC    | Sensitivity | Specificity | Precision | NPV     | FDR     | Accuracy | AP     | F1-Score | MCC    |
|--------|--------|-------------|-------------|-----------|---------|---------|----------|--------|----------|--------|
| LGBM   | 0.9882 | 0.7500      | 1.0000      | 1.0000    | 0.9938  | 0.0000  | 0.9939   | 0.8864 | 0.8571   | 0.8633 |
| RF     | 0.9917 | 0.6250      | 1.0000      | 1.0000    | 0.9907  | 0.0000  | 0.9908   | 0.8596 | 0.7692   | 0.7869 |
| GNB    | 0.9906 | 0.8750      | 0.9843      | 0.5833    | 0.9968  | 0.4167  | 0.9816   | 0.6407 | 0.7000   | 0.7061 |
| KNN    | 0.9336 | 0.6250      | 0.9937      | 0.7143    | 0.9906  | 0.2857  | 0.9847   | 0.6281 | 0.6667   | 0.6604 |
| MLP    | 0.8247 | 0.2500      | 0.9560      | 0.1250    | 0.9806  | 0.8750  | 0.9387   | 0.0873 | 0.1667   | 0.1475 |
| CART   | 0.9969 | 0.8750      | 0.9937      | 0.7778    | 0.9968  | 0.2222  | 0.9908   | 0.9152 | 0.8235   | 0.8203 |
| SVM    | 0.9894 | 0.8750      | 0.9465      | 0.2917    | 0.9967  | 0.7083  | 0.9448   | 0.8147 | 0.4375   | 0.4867 |
| Stacking | 0.9917 | 0.8750      | 0.9623      | 0.3684    | 0.9967  | 0.6316  | 0.9601   | 0.8113 | 0.5185   | 0.5529 |

AUC, area under curve; NPV, negative predictive value; FDR, false discovery rate; AP, average precision; MCC, matthews correlation coefficient; LM, liver metastasis; LGBM, Light Gradient Boosting Decision; RF, Random Forest; GNB, Gaussian Naive Bayesian; KNN, K-Nearest Neighbor; MLP, Multilayer Perceptron; CART, Classification and Regression Trees; SVM, Support Vector Machine.

Figures
Figure 1
The workflow of selection procedure for colorectal cancer patients
Figure 2

Predictive value of models after optimization Inner validation in SEER database: (a) ROC curves of seven individual models and stacking model. (b) PR curves, indicating the tradeoff between precision and recall. Outer validation in our Chinese cohort: (c) ROC curves of seven individual models and stacking model. (d) PR curves, indicating the tradeoff between precision and recall. SEER: Surveillance, Epidemiology, and End Results; ROC: receiver operating characteristic; and PR: precision-recall.
Figure 3

Evaluation of the prognostic value for stacking-bagging model (a) The survival curve based on real data. (b) The survival curve based on predictive outcomes.
Figure 4

Factor importance of the developed models. Bar graphs describe the proportion of importance of the different predictors in models. The top ten factor importance were exhibited in models: (a) Average of factor importance in seven models, (b) LGBM, (c) RF, (d) GNB, (e) KNN, (f) MLP, (g) CART, and (h) SVM. LGBM: Light Gradient Boosting Decision; RF: Random Forest; GNB: Gaussian Naive Bayesian; KNN: k-nearest neighbor algorithm; MLP: Multilayer Perceptron; CART: Classification and Regression Trees; and SVM: Support Vector Machine.